# Disentangling the sources of variation in functional responses: between-individuals variability, measurements errors and inherent stochasticity of the prey-predator interaction process

July 9, 2025

Charlotte Baey[1,*], Sylvain Billiard[2,*,†], Maud Delattre[3,*]
[1]Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000 Lille, France (charlotte.baey@univ-lille.fr)
[2]Univ. Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France (sylvain.billiard@univ-lille.fr)
[3]Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France (maud.delattre@inrae.fr)
[*]These authors contributed equally to this work.
[†]Corresponding author

Keywords: foraging | population ecology | community ecology | stochastic process | non-linear mixed models | stochastic gradient descent

**Abstract**

The consumption rate of prey by predators, or functional responses, are known to be highly variable even within a single population. Identifying and estimating the different sources of variation of functional responses is a long-standing challenge. We develop here a statistical framework derived from a mechanistic stochastic process model that explicitly accounts for inter-individual variability, intrinsic stochasticity due to the foraging and interactions processes themselves, and measurement error. We show that it is possible to estimate all sources of variation under realistic experimental conditions. Our results also show that model fitting can compensate by overestimating residual source of variation, leading to biased parameter estimates when the mechanistic noise is misspecified. Applied to empirical data, the model reveals that standard assumptions, such as prey renewal and lack of spatial structure, fail to capture observed variability. We also show how experimental design affects parameter identifiability, highlighting the trade-off between the number of individuals and repeated observations.

# 1 Introduction

Functional responses are functions describing the rate of interactions between two or more species in population or community ecology models. For example, functional responses relate the consumption rate of prey by a predator and the prey density in the environment, such as captured by the classical Holling types I, II or III models or many others [16, 17]. As prey and predator dynamics are coupled through such a functional response, the chosen functional form introduced into models generally has a dramatic impact on the conditions for the coexistence between species, the stability of a community and the speed of the flux of matter and energy in an ecosystem. As a consequence, it is a major challenge in ecology to properly estimate functional responses and their parameters from data, as shown by the large number of experiments devoted to this goal that have been conducted in the last decades [17, 25, 38].

It has long been recognized that prey consumption rates measured in experiments typically show unexpected large variance as well as heteroscedasticity [35]. For instance, under the assumption that the number of preys eaten by a predator follows a binomial distribution, data typically show overdispersion [35, 22, 7, 14]. A beta-binomial distribution was then proposed as an alternative for inference as it allows for overdispersion, albeit at the cost of a higher number of parameters to be estimated [7, 14]. However, no biological justification was given for such a choice. More recently, consumption rates have been shown to commonly have a standard deviation of the same order of magnitude than the mean [4]. Such a large variance in data raises the challenge to correctly identify and quantify the possible sources of such a large variability. There are four possible sources of variation of the functional responses as observed in experiments or natural populations: i) measurements errors [18, 3]; ii) between-individuals variability in the traits involved in the prey-predator interaction such as individuals size, personality, age or satiety (*e.g.* [22, 24, 28, 31]); iii) variability of the environment, *e.g.* dimension, temperature, vegetative cover (*e.g.* [24, 36]); iv) *mechanistic* (or *intrinsic*) source of variation, *i.e.* variability due to the foraging and the interaction processes themselves [6, 4, 12].

Such different sources of variability of functional responses are at most partially taken into account, when they are. Functional responses are generally estimated by fitting non-linear models that describe the relationship between prey density and consumption rate to data collected in a controlled environment where the number of preys consumed per unit of time per predator are measured at different prey densities (*e.g.* [3, 29], see also [25, 38] for extended reviews), rarely in natural populations (*e.g.* [24, 30]). Several methodological approaches are applied to select which functional response best fits the data among possible alternatives, or to estimate their parameters. Depending on the chosen approach, different sources of variations of the functional response could, or could not, be evaluated. A first approach consists in estimating parameters using independent behavioral measurements and direct observations, for example for the time taken to handle a prey, or the movement speed, and plug these estimates in candidate functions, in order to predict consumption rates, and sometimes to qualitatively compare predictions to data (e.g. [22, 34, 3, 21, 5]).

A second approach is based on directly fitting deterministic equations to observed prey consumption rates, or the number of preys consumed after a given fixed time, using least squares or likelihood optimization procedures [15, 35, 26, 8, 36, 33, 34]. In the first and second approaches, variability of the functional responses and its possible sources are totally neglected. Third, goodness of fit or regression procedures are applied to estimate the parameter of a deterministic equation specifying a particular error structure with prey density as explanatory variable [35, 7, 34, 30, 14, 29], sometimes associated with other independent variables such as environmental treatments [23]. In this case, the only quantified source of variability is the measurements error estimated as the residual variance. Fourth, by using a maximum-likelihood procedure and a likelihood ratio test, [6] fitted a Holling II-type model including both residual variance and intrinsic mechanistic variance, neglecting between-individual variability, with data lacking repeated measures on identical predator individuals, precluding a correct estimate of the various sources of variation. Finally, A last approach consists in directly including inter-individual variability into non-linear regression models by adding individuals as a random effect, or considering covariates such as social status [24]. In such an approach, measurements errors and inter-individual variability are thus evaluated as parameters of the supposed error structure (residual variance) or the distribution of the random effect (inter-individual variance). Inter-individual variability is sometimes indirectly evaluated in two steps: the parameters of the functional response are first estimated from direct observations data (as in the first approach, see before). Correlations are then searched between these estimated individual parameters with covariates (*e.g.* individuals size or behavior), the environment (*e.g.* temperature), either for different individuals of a single species [32], or for different individuals for different species [37, 28], or through comparison between the variance observed in data and the one predicted in a partial differential equation model accounting for inter individual variance only [22].

In a few words, to our knowledge there is no statistical framework accounting for all possible sources of variation adapted to functional responses inference. It is yet important as model selection and parameter estimations of functional responses are necessarily biased or lack precision because of such a large uncontrolled and unidentified variance [25], for three main reasons. First because it is common that several different candidate non-linear models fit close enough to the data so that it is impossible to select one among others, because of the nature of data and non-linear models themselves (*e.g.* [3, 33, 25]). Second, if the inter-individual variability or the mechanistic variability are neglected, they are necessarily conflated with residual noise, leading to poor parameter estimates. Third, because heteroscedasticity —unequal variation in errors across prey densities— is frequently observed. At low prey densities, errors may be relatively small, while at higher densities, variability in consumption rates often increases. Failing to account for these differences in error can result in biased estimates, particularly regarding the effect of prey density and other mechanistic parameters. Therefore, it is essential to adopt a rigorous approach that accounts for these factors, not only to improve the accuracy of parameter estimates but also to ensure that the results can be interpreted and generalized to broader ecological contexts.

This paper aims to demonstrate how to conduct a robust statistical analysis of experi-

mental data on functional responses, incorporating all sources of variability. It will illustrate how appropriate statistical modeling, combined with robust inference methods, can correct these biases. In particular, the use of mixed-effects models allows for inter-individual variability to be accounted for, thereby improving the quality of parameter estimates. Our statistical framework will also be based on the elaboration of functional responses derived from stochastic processes modeling of the interaction between prey and predators, which is necessary to take into account the mechanistic (or intrinsic) source of variation which is due to the foraging and interactions process themselves. We will describe the algorithmic procedure that allows estimation of such functional responses based on Stochastic Gradient Descent (SGD). We will then show how to account for the different possible sources of variation. We will then compare models with or without the mechanistic source of variability. Such a model comparison will allow to evaluate the relative importance of the mechanistic variation comparatively to inter-individual variability and to residual variance. Finally, we will also compare the reliability of the estimations of different experimental designs.

# 2 Methods

The fundamental idea of the statistical approach presented hereafter is that data observed in an experiment, here the number of preys consumed by several isolated predators at different prey densities after a given time $\Delta$, denoted by $R$, are the realizations of a stochastic process. In other words, we suppose that $R$ follows a distribution which is not *a priori* chosen but that emerges from the process itself. This distribution constitutes the core of the statistical model used for inference. For that, we suppose that the statistical model can be decomposed into two main components: a deterministic (or mean) part, and a stochastic (or variance) part.

In the following, we first present the statistical model including all sources of variation (measurement errors, inter-individual variability, mechanistic noise). Second, we show that the two components of the statistical model naturally emerge from the convergence of the mechanistic model to an approximation of the distribution of $R$ in general (Eq. 3), and then in a particular context and parameterization (Eqs. 4 and 5). Even though including a variance part to a statistical model is classical, either constant or dependent on the parameters of the model or the data structure, the form given to the variance is often an *a priori* or *adhoc* choice. Here on the contrary, the form given to the variance, and especially its dependency on the parameters, directly comes from the mechanistic and probabilistic model itself. The part of the variability emerging from the process originating the data (the mechanistic noise) is thus a source of information for parameter estimation.

## 2.1 A mechanistic stochastic model for addressing variation sources in functional responses

**Statistical model.** The statistical model that accounts for all sources of variation in the observed number of preys consumed, denoted $Y_{ij}$, by predator $i$ at prey density $d_{ij}$ (with

$i = 1, \ldots, N$ and $j = 1, \ldots, n_i$) in a given experimental environment $e$, is defined as the following nonlinear heteroscedastic mixed-effects model:

$$\begin{aligned} Y_{ij} &= R(\varphi_i, d_{ij}, e) + \epsilon_{ij} \ , \ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_0^2), \\ R(\varphi_i, d_{ij}, e) &= r(\varphi_i, d_{ij}, e) + \delta_{ij} \ , \ \delta_{ij} \sim \mathcal{N}(0, v(\varphi_i, d_{ij}, e)), \\ \varphi_i &= \varphi + \eta_i \ , \ \eta_i \sim \mathcal{N}(0, \Omega). \end{aligned} \tag{1}$$

Here, the latent variable $R$, whose definition is detailed below, represents the interaction process between predators and preys. It is modeled as a deterministic function $r$, capturing the mean response, combined with a stochastic deviation $\delta_{ij}$, reflecting intrinsic mechanistic variability. Both $r$ and the associated variance function $v$ depend on the individual-specific traits $\varphi_i$, which are modeled as random effects with population mean $\varphi$ and variance $\Omega$. Model (1) thus captures inter-individual variability in predation behavior through the distribution of $\varphi_i$, and incorporates both process stochasticity (via $\delta_{ij}$) and measurement error (via $\epsilon_{ij}$). Importantly, since the process variance $v(\varphi_i, d_{ij}, e)$ results from a mechanistic approximation of the stochastic interaction model, it has a closed-form expression that depends explicitly on the model parameters and the form of the chosen functional response (see justification below). Assuming independence between these two sources of noise, the marginal distribution of $Y_{ij}$ is Gaussian with mean $r(\varphi_i, d_{ij}, e)$ and variance

$$\sigma_{ij}^2 = \sigma_0^2 + v\left(\varphi_i, d_{ij}, e\right). \tag{2}$$

Since the variance of $Y_{ij}$ depends on the individual traits $\varphi_i$, prey density $d_{ij}$, and environment $e$, the model is heteroscedastic by construction.

**Theoretical justification.** Our inferential model (Eq. 1) can be derived from a mechanistic, microscopic and stochastic model describing how fast a single individual predator, with its peculiar set of traits, interacts with preys in a given environment during a foraging bout. Such a general model has been developed by [6] where they showed that given i) the density of preys $d$, ii) a set of predator traits $\varphi$ (*e.g.* its foraging pattern, movement speed, handling time, probability of capture, etc.), iii) the distribution of the total times $T_k$ taken by the predator to successfully eat the $k^{th}$ prey during the foraging bout, and iv) an environment $e$ and its dimension, then the distribution of the functional response $R\left(\varphi, d, e\right)$, *i.e.* the total number of prey consumed per unit of time during a foraging bout with a duration $\Delta$, can be approximated, when $\Delta \to \infty$, by

$$R\left(\varphi, d, e\right) \approx \mathcal{N}\left(\frac{1}{\mathbb{E}\left[T(\varphi, d, e)\right]}, \frac{1}{\Delta} \frac{\mathrm{Var}\left[T(\varphi, d, e)\right]}{\mathbb{E}\left[T(\varphi, d, e)\right]^3}\right). \tag{3}$$

Approximation (3) is obtained under the assumptions that the times $T_k$ are independent and have the same distribution as variable $T(\varphi, d, e)$ in the above formula, that the foraging bout duration $\Delta$ is large enough for a large number of successful interactions to have

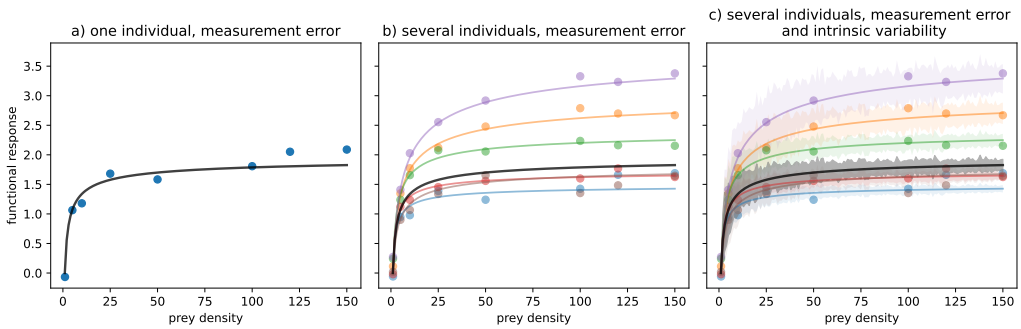


Figure 1: The different sources of variability in a functional response. In all plots, the black curve represents the mean functional response $r$. On the left, the dots corresponds to the mean functional response plus a measurement error. On the middle, each color corresponds to a different individual, with its theoretical functional response in solid line and its observations with measurement errors in dots. On the right, mechanistic variability has been added to measurement errors: the solid line represents the mean functional response and the colored area represents the range of the distribution of the functional response.

occurred, and that the variation of prey density $d$ during the foraging bout is negligible (in other words, under the assumptions of renewal theory).

The functions $r$ and $v$ used to parameterize the inferential model (Eq. 1) are precisely obtained as the mean and variance of the distribution specified in (3). Eq. (3) thus helps to disentangle different sources of variation expressed in the inferential model (Eq. 1). As Eq. (3) was derived at the microscopic (individual) level, it allows to quantify both the expectation of the functional response $r$ and its random fluctuations due to the foraging process itself $v$. Estimating $v$, the variation due to the foraging process itself, is made possible because the size of the random fluctuations of the functional response depends on the individual predator traits $\varphi$, under the condition that collecting data from the same single individual is possible.

**Example.** In order to illustrate the derivation of a statistical model from Eq. (3) and its use for performing inferences from a dataset, we will consider in the numerical illustrations of the paper a specific simplified context which is yet classically used in ecology. We suppose that a single predator forages in a constant environment $e$, with prey density $d$ fixed by the experiment design. The predator traits set is characterized by its searching rate $\lambda$ and the time it takes for handling a prey when captured $h$, *i.e.* $\varphi = (\lambda, h)$, and we suppose that its foraging pattern is simply that it goes directly to the closest available prey. [6] showed that under these assumptions (similar to the Holling type II model [16], which is deterministic) then an approximation of the stochastic functional response is given by a Gaussian distribution with mean

$$r(\lambda, h, d) = \frac{d^{1/D} - 1}{C_E\lambda + h(d^{1/D} - 1)}, \tag{4}$$

and variance

$$v(\lambda, h, d) = \frac{C_V - C_E^2}{\Delta} \frac{\lambda^2(d^{1/D} - 1)}{(C_E\lambda + h(d^{1/D} - 1))^3}, \tag{5}$$

where $C_E$ and $C_V$ are known constants that only depend on the dimension $D$ of the environment (see [6]). $\Delta$ is the time scale of the experiment which is fixed to $\Delta = 1$ in the following for simulated datasets without loss of generality, so that $r$ is to be expressed as the number of preys ingested by the predator per unit of time $\Delta$. In the case of the real dataset, it was fixed to $\Delta = 2$ min as it was the duration of an experimental sequence. In this model, $1/\lambda$ represents the speed of prey ingestion per unit area and $1/h$ represents the asymptotic number of preys a predator can ingest when the prey density tends to infinity. Assuming that all measurements are made under the same experimental conditions, the environment $e$ is omitted from the notations above for simplicity.

## 2.2 Statistical Tools for Mixed-Effects Models with Heteroscedasticity

The vector of parameters to be estimated in the model given in Eq. (1) is $\theta = (\varphi, \Omega, \sigma_0^2)$. Inference in this model requires dedicated statistical tools, due to its nonlinear structure and the presence of latent variables. In particular, the individual-specific random effects $\varphi_i$ enter the model nonlinearly, both in the mean function $r$ and in the variance function $v$. As a result, standard estimation procedures based on explicit likelihoods or simple sufficient statistics are no longer tractable. Likelihood-based inference must instead integrate over the distribution of the latent random effects, which introduces additional computational and methodological challenges. Moreover, although the theoretical framework for mixed-effects models is well established (see, for instance, [27] and [20]), no generic software currently supports mixed-effects models that combine inter-individual variability with heteroscedasticity — especially when the latter arises from a nonlinear variance structure derived from a mechanistic stochastic process. In the following, we present statistical tools specifically designed for parameter estimation and model comparison in functional response models that incorporate multiple sources of variation, with the aim of making them accessible and practical for experimenters.

### 2.2.1 Algorithms for parameter estimation.

**A brief review of the state of the art.** Several methods have been developed to tackle parameter estimation in nonlinear mixed-effects models. However, two particularly powerful and flexible likelihood-based strategies stand out for our context: the Stochastic Approximation EM (SAEM) algorithm [13, 19] and Stochastic Gradient Descent (SGD) [1]. SAEM approximates the complete-data likelihood through stochastic updates, while SGD seeks to drive unbiased estimates of the gradient of the log-likelihood toward zero,

and is often easier to implement. Importantly, an additional advantage of SGD is that, beyond parameter estimation, it naturally provides an estimate of the Fisher information matrix as a by-product of the optimization process. This output is valuable for uncertainty quantification. Detailed descriptions of both algorithms are provided in Appendix A.

SAEM is implemented in several software packages, including the `saemix` package for R [9]. A key limitation of current implementations is that they only support heteroscedastic models where the variance depends on the mean through simple parametric forms—typically the proportional error model ($\sqrt{v} = br$) or the combined error model ($\sqrt{v} = a + br$), with constants $a$ and $b$. More general forms of heteroscedasticity, such as those arising from mechanistic stochastic processes, are not covered. In these cases, the model no longer belongs to the exponential family, and the required updates at each iteration of the algorithm do not admit closed-form expressions (see [19]). This significantly increases the implementation complexity and computational cost of SAEM (see Appendix A). In contrast, to our knowledge, there are no dedicated R or Python packages for SGD in mixed-effects models. We therefore developed a tailored code to enable statistical inference in functional-response models under this framework. Appendix A details the implementation, and the full source code used for numerical experiments and data analysis is available on GitHub at `https://github.com/baeyc/functional_responses`.

**Practical considerations for SGD.** Applying Stochastic Gradient Descent in nonlinear mixed-effects models involves several practical considerations to ensure reliable and efficient convergence. As an iterative method, SGD progressively refines parameter estimates over successive updates, starting from initial values. The final estimate corresponds to the last iteration, under the assumption that convergence has been achieved. The performance of SGD depends on a careful tuning of algorithmic parameters and diagnostic monitoring throughout the estimation process. A first critical aspect is initialization: because SGD tends to converge to a local rather than a global optimum, initial values should be chosen carefully, ideally informed by domain knowledge, to guide the algorithm toward plausible and interpretable solutions. Another key factor is the choice of the learning rate (denoted by $\gamma_k$ in Appendix A), which controls the magnitude of parameter updates. A large initial learning rate can promote exploration of the parameter space, helping to avoid poor local optima, while a decreasing learning rate allows for finer adjustments as the algorithm converges. Designing an appropriate learning rate evolution therefore requires balancing between exploration and stability. In addition, convergence diagnostics play an important role in practice. Plotting the evolution of the estimated parameters or the log-likelihood across iterations can help assess whether the algorithm is stabilizing appropriately. Signs of instability—such as oscillations or stagnation—may indicate the need to revise the initialization or adjust the learning rate. Figure 2 illustrates such behavior in a representative example. In order to be confident enough in the inferences made with data from the model, the algorithm should properly converge and estimate values in accordance with data, and the robustness and consistency of the estimates should be assessed. These are necessary conditions before moving forward to comparing different models, selecting an appropriate noise

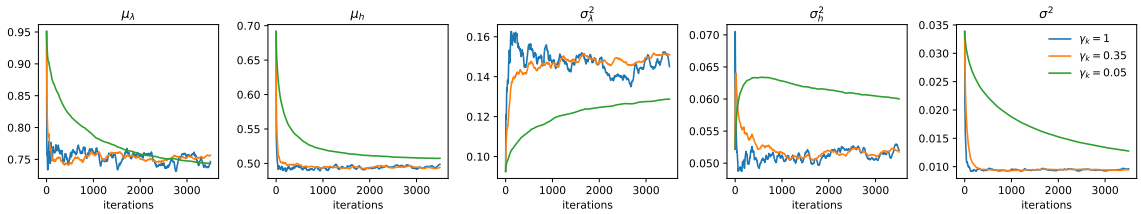


Figure 2: Effect of the learning rate on the convergence of the Stochastic Gradient Descent algorithm for estimating the parameters of the model. When the learning rate $\gamma_k$ is too small (in green), the trajectory is smooth but the convergence is too slow. As the step size increases, the algorithm achieves a faster convergence but the trajectory exhibits higher variability.

structure, or determining whether a parameter should include inter-individual variability.

### 2.2.2 Model comparisons

In the context of nonlinear mixed-effects models, information criteria such as the Bayesian Information Criterion (BIC) provide a valuable and practical framework for model selection. While alternative approaches and more classical statistical tests (e.g., likelihood ratio tests) exist, these often involve significant analytical and computational challenges due to the models' hierarchical structure, where individual-level parameters are latent random effects, and their intrinsic nonlinearity. Such complexity leads to test statistics that may follow non-standard distributions, making their evaluation both analytically intractable and numerically expensive, especially when determining appropriate covariance structures for the random effects [2].

Moreover, the BIC offers the advantage of enabling comparison between models making assumptions on any source of variability, be it in the mean structure, variance components, or random effects, whereas many classical tests or alternative tools tend to be tailored to specific hypotheses or focused on particular model components. For these reasons, we focus here on information criteria as a simpler, more intuitive, and computationally efficient tool for model comparison, without the need for costly $p$-value estimation.

An important modeling question in our context is whether the variance structure should include a mechanistic term. To address this, we compare candidate models that differ in their assumptions about the variance of the observations, by including or excluding the stochastic component derived from the mechanistic formulation

$$\sigma^2_{ij} = \sigma^2_0 \quad \text{(model 1)} \qquad \text{vs.} \qquad \sigma^2_{ij} = \sigma^2_0 + v(\varphi_i, d_{ij}) \quad \text{(model 2)}.$$

In this context, the BIC takes the form

$$BIC = -2\,\mathcal{L}(\hat{\theta}) + \dim\theta \log\left(\sum_{i=1}^{N} n_i\right),$$

where $\mathcal{L}(\hat{\theta})$ denotes the log-likelihood evaluated at the estimated parameters, $\dim \theta$ is the number of estimated parameters, and $\sum_{i=1}^{N} n_i$ is the total number of observations. Comparing models based on their BIC values by selecting the one with the smallest criterion, provides a straightforward and unified decision rule, regardless of the nature of the variability components. The main technical challenge here lies in the evaluation of the log-likelihood, which is not available in closed form due to the presence of latent random effects and the complexity of the variance structure. To overcome this, we rely on Monte Carlo approximations of the likelihood, as detailed in Appendix A.2.

### 2.3 Numerical experiments.

**Objective.** We conducted a simulation study in order to illustrate how experimental design and data-generating conditions affect the quality of parameter estimation and model selection in heteroscedastic mixed-effects models, particularly with respect to the identification and estimation of the different sources of variability, and indirectly to provide guidance for ecologists in designing their experiments. To facilitate analysis and interpretation, we considered a two-dimensional model with random parameters that do not depend on covariates. Under a fixed measurement budget, we examined how the trade-off between the number of individuals and the number of repeated measurements per individual influences estimation bias, estimator variance, and model selection based on the Bayesian Information Criterion (BIC). We also investigated how the magnitude of variability—both between and within individuals—affects these performance metrics.

**Model.** We generated data following the statistical model defined in Eqs. (1)-(2) where the observations variance is splitted into a mechanistic term $v$ and a residual term $\sigma_0^2$. $r$ is given by Eq. (4) and $v$ by Eq. (5) with environment dimension $D = 2$. A log-transformation was applied to the two behavioral parameters to account for their positiveness. More precisely, we assumed the following:

$$\begin{pmatrix} \log \lambda_i \\ \log h_i \end{pmatrix} = \mathcal{N}\left( \begin{pmatrix} \mu_\lambda \\ \mu_h \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \sigma_{\lambda,h}\sigma_\lambda\sigma_h \\ \sigma_{\lambda,h}\sigma_\lambda\sigma_h & \sigma_h^2 \end{pmatrix} \right).$$

The model is associated with typical curve shapes that vary according to the parameter values. Consequently, depending on the densities at which observations are collected, they are likely to provide information on all or part of the curve, resulting in better or poorer quality estimates. Two example datasets associated with different parameter sets are given in Figure 3, with $\theta_1 = (\mu_\lambda = 0.7, \mu_h = 0.5, \sigma_\lambda^2 = 0.07, \sigma_{\lambda,h} = 0.5, \sigma_h^2 = 0.05, \sigma_0^2 = 0.01)$ in the first case and with $\theta_2 = (\mu_\lambda = 1.75, \mu_h = 0.5, \sigma_\lambda^2 = 0.175, \sigma_{\lambda,h} = 0.5, \sigma_h^2 = 0.05, \sigma_0^2 = 0.01)$ in the second case. The asymptotic number of preys a predator can ingest is given by $1/h$ and is identical in both settings. The difference between the two parameter sets lies in the value of the behavioral parameter $\lambda$, which influences the speed of preys consumption. For a prey density that varies between 1 and 150, we can observe the whole functional response trajectory with parameter set $\theta_1$, while the asymptotic regime is not reached with the second parameter set $\theta_2$.

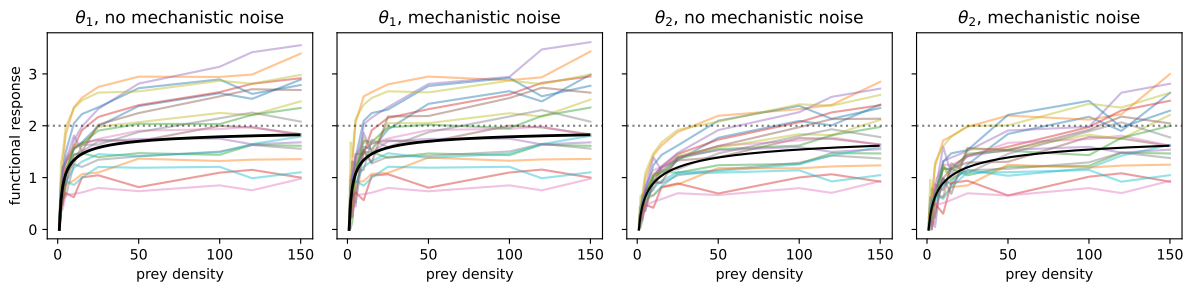


Figure 3: Example of simulated datasets using $\theta_1$ and $\theta_2$, with and without mecanistic noise. The dotted line represents $1/h$, the asymptotic number of preys a predator can eat. The black curve represents the mean functional response in the population.

The generation of simulated data thus requires two types of inputs: i) a set of parameter values from which the functional response can be derived, and ii) the observation scheme, *i.e.* the prey densities at which the data are to be observed.

**Experimental settings.** In all cases below, we evaluate the quality of the estimates and the possibility of choosing the right model when comparing the model with or without the mechanistic source of variability $v(\varphi_i, d_{ij})$ by means of the BIC (*i.e.* we compare a model with variance $\sigma_{ij}^2 = \sigma_0^2 + v(\varphi_i, d_{ij})$ to a model with variance $\sigma_{ij}^2 = \sigma_0^2$ ).

1. We first considered several configurations for the number of individuals $N$ and the number of measurements per individual $n$ (we therefore assumed the same number of measurements for each individual). For each parameter set, $\theta_1$ and $\theta_2$ above, and for a fixed total number of measurements of approximately $nN = 500$, we let the number of individuals and the number of measurements vary as $(N, n) \in \{(10, 50), (20, 25), (30, 17), (40, 12), (50, 10), (60, 8), (70, 7), (80, 6), (90, 5)\}$. Then, the prey densities at which the data are to be observed were randomly generated between 1 and 150, for each value of $n$.

2. We then investigated the influence of the signal-to-noise ratio through two complementary experiments. First, we varied the coefficient of variation of the random effects in $\{25\%, 50\%, 75\%\}$, thereby increasing the relative dispersion of individual parameters around the population mean. Second, we modulated the measurement noise by adjusting the residual standard deviation $\sigma_0$, expressed as a proportion of the asymptotic response, using values of $\sigma_0 \in \{5\%, 10\%, 25\%, 50\%, 75\%\} \times \frac{1}{h}$. These experiments are conducted for parameter set $\theta_1$ only, assuming zero covariance among the random effects.

**Evaluation criteria.** We assess the estimation accuracy by measuring the bias and the variance of the estimates obtained on the simulated datasets, which can then be combined

in the mean squared error (MSE). For that purpose, for a given parameter set $\theta$ and for each tested sample size, we generated $K = 1000$ datasets and computed the corresponding estimate $\hat{\theta}_i$ for each dataset. We then computed the aforementioned criteria as:

$$\text{bias}(\hat{\theta}) = \bar{\hat{\theta}} - \theta := \frac{1}{K}\sum_{i=1}^{K}\hat{\theta}_i - \theta, \quad \text{Var}(\hat{\theta}) = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\theta}_i - \bar{\hat{\theta}}\right)^2,$$
$$MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{bias}^2(\hat{\theta}).$$

In order to compare the estimates of parameters that can be defined on different scales, one can use the relative root MSE, which is defined as $relRMSE = \sqrt{RMSE(\hat{\theta})}/\theta$. Model selection is assessed by counting the number of times, out of $K = 1000$ repetitions, that the correct model is actually found, *i.e.* associated with a smaller BIC value.

### 2.4 Real dataset

We ran our inferential model on a real dataset from [32] (see Figure 4). The number of invertebrate preys (*Artemia salina*) ingested by an isolated predator fish (*Heterandria formosa*) in two minutes was measured by direct visual observation at 14 different initial prey densities (range: 10-2000 prey per 2L). Forty-nine different predator individuals were used for the experiment. The number of preys ingested was measured for the 14 initial prey densities for all individual predators (*i.e.* a given individual predator was measured 14 times independently, which gives information about the variation due to the stochastic interaction process itself). Two feeding trials were performed per day and a trial series lasted 7 days. The duration of one feeding trial being $\Delta = 2$ minutes, it gives the time dimension of the estimated parameters as they are scaled to $\Delta$ in Eq. (3).

## 3 Results

### 3.1 Estimation accuracy

**Experimental designs comparison.** Results are given in Figs. 5 and 6 for both parameter sets $\theta_1$ and $\theta_2$. Different behaviors can be observed for the relative MSE. In general, the relative MSE decreases as the number of individuals increases. However, for some parameters we observed a first decreasing phase followed by a rising phase. We recall that the total number of measurements is maintained approximately constant around 500, so that when the number of individuals increases, the number of measurements per individual decreases. The latter behaviour could thus be explained by the fact that in those configurations, the number of measurements per individual is too small to correctly estimate the corresponding parameter. This is the case for example for parameters $\mu_\lambda$ and $\mu_h$ when the data were simulated using $\theta_2$, starting from $N = 50, n = 10$, and to a lesser extent, for parameter $\mu_\lambda$ when the data were simulated using $\theta_1$, starting from $N = 70, n = 7$. In these settings, we observe that as $\mu_\lambda$ is underestimated, $\mu_h$ is overestimated, suggesting that compensation can occur

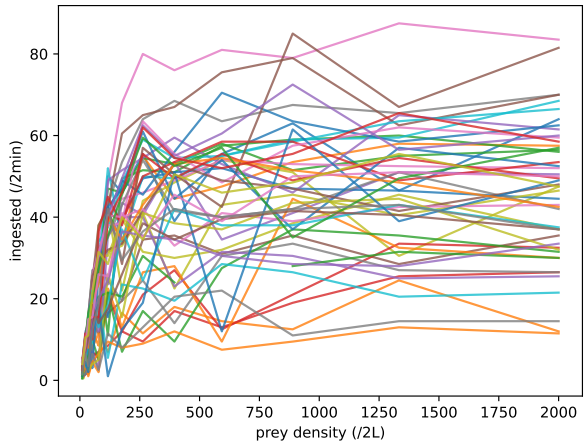


Figure 4: Number of ingested *A. salina* by a single predator *H. formosa* during a time window of 2 minutes in a 2L aquarium, at 14 different food densities (see [32]). Each color corresponds to the measurements made on the same single predator at different prey densities.

when the number of measurements per individual is too small. This is particularly true when the observation scheme does not cover the asymptotic regime, as this is the case with $\theta_2$.

**Effect of the signal-to-noise ratio.** Results are given in Figs. 7, 8, 9 and 10. When the residual variance increases (Figures 9 and 10), all model parameters tend to be estimated with larger bias. The estimation uncertainty as measured by the standard deviation of the estimators, also increases with higher residual variance. A similar effect is observed with increasing random effect variance (Figures 7, 8). This suggests that reducing the measurement errors can directly improve the quality and precision of estimates and inference, a gain to be balanced by the associated experimental costs. An alternative could also be to better control the inter-individual variability which might be difficult to achieve as inter-individual variability is intrinsic to biological systems, and which might even not be desirable if evaluating the role of inter-individual variability is one of the aim of the experiment.

**Effect of model misspecification.** Fig. 11 depict results corresponding to a misspecification of the noise structure ($\sigma_{ij}^2 = \sigma_0^2$ instead of $\sigma_{ij}^2 = \sigma_0^2 + v(\varphi_i, d_{ij})$). When the noise structure is misspecified, we observe an over-estimation of the residual variance term $\sigma_0^2$ that compensates for the extra-variability of the data compared to the case where no mechanistic noise would be present. We also observe an over-estimation of the variance component $\sigma_\lambda$, in particular when the sample size $N$ increases and the number of measures per individual $n$ decreases. Since the estimate of residual variance is decreasing at the same time, it could

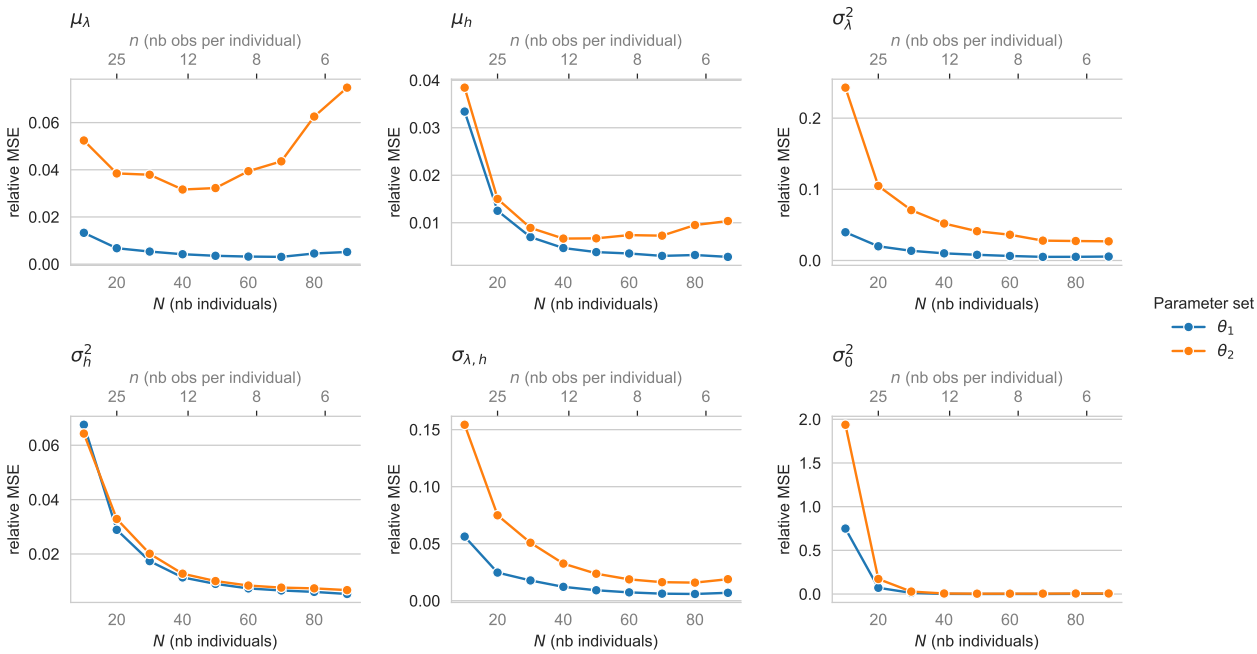


Figure 5: Relative Mean Squared Error (MSE) of each parameter as a function of the number of individuals in the population, for both parameter sets $\theta_1$ and $\theta_2$, computed over $K = 1000$ repetitions. (bottom)

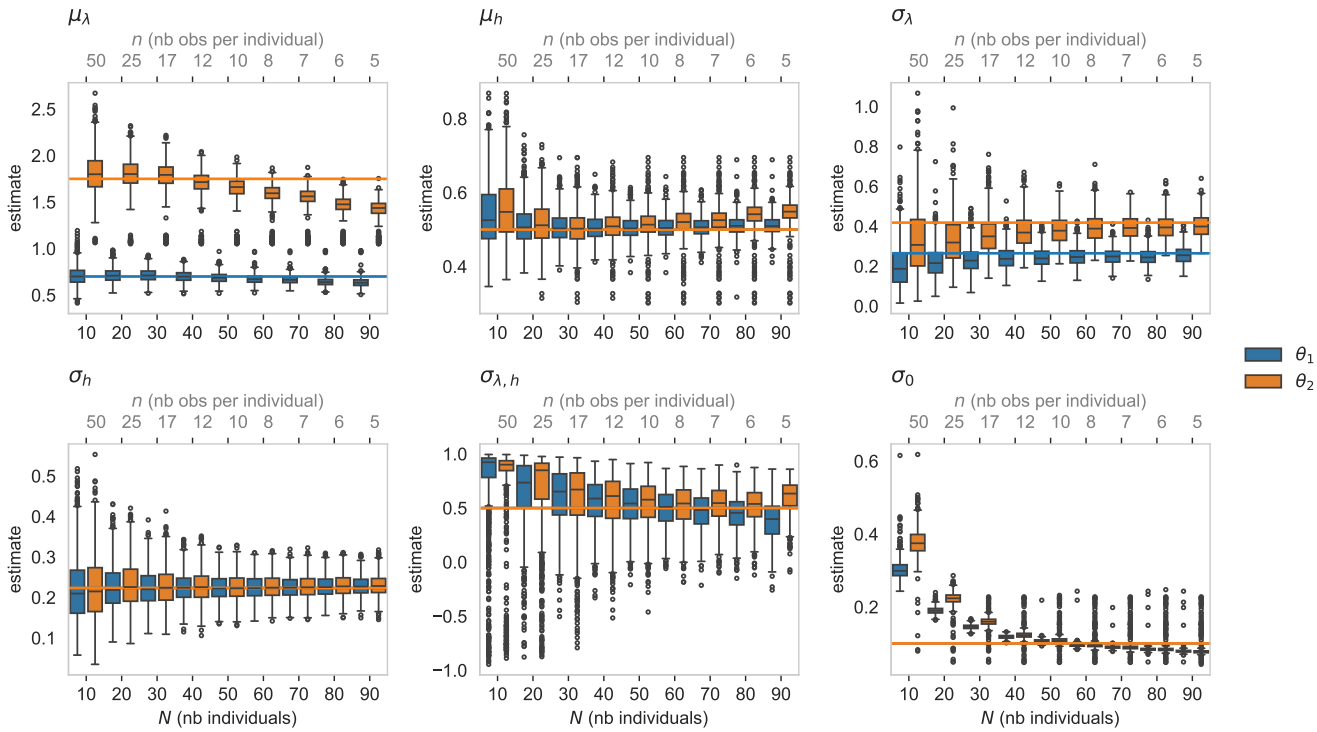


Figure 6: Boxplot of the estimates obtained on $K = 1000$ repetitions, as a function of the number of individuals in the population (principal x-axis) and of the number of observations per individual (secondary x-axis), for the two parameter sets $\theta_1$ (orange) and $\theta_2$ (blue). The horizontal line shows the true value of each parameter under each parameter set.

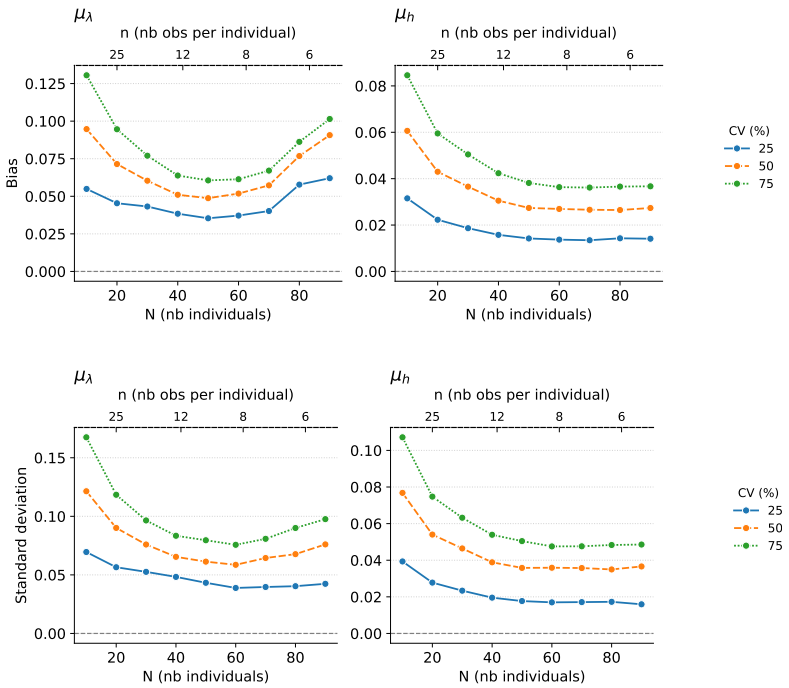


Figure 7: (top) Bias (in absolute value) of each mean parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the random effects and the parameter set $\theta_1$, computed over $K = 1000$ repetitions. (bottom) Standard deviation of each mean parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the random effects and the parameter set $\theta_1$, computed over $K = 1000$ repetitions.

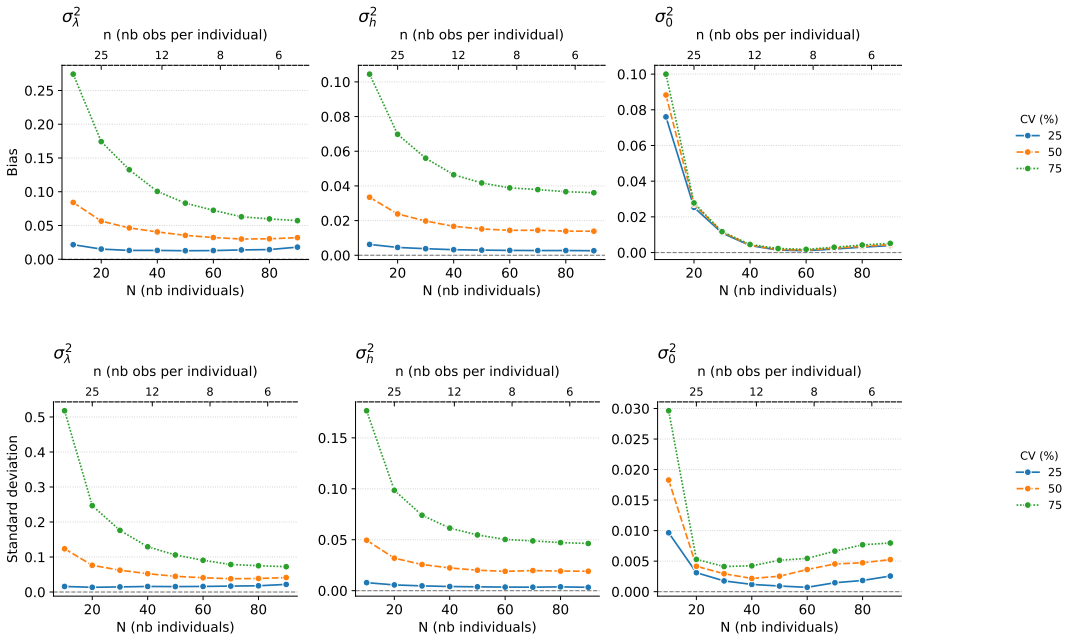


Figure 8: (top) Bias (in absolute value) of each variance parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the random effects and the parameter set $\theta_1$, computed over $K = 1000$ repetitions. (bottom) Standard deviation of each variance parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the random effects and the parameter set $\theta_1$, computed over $K = 1000$ repetitions.

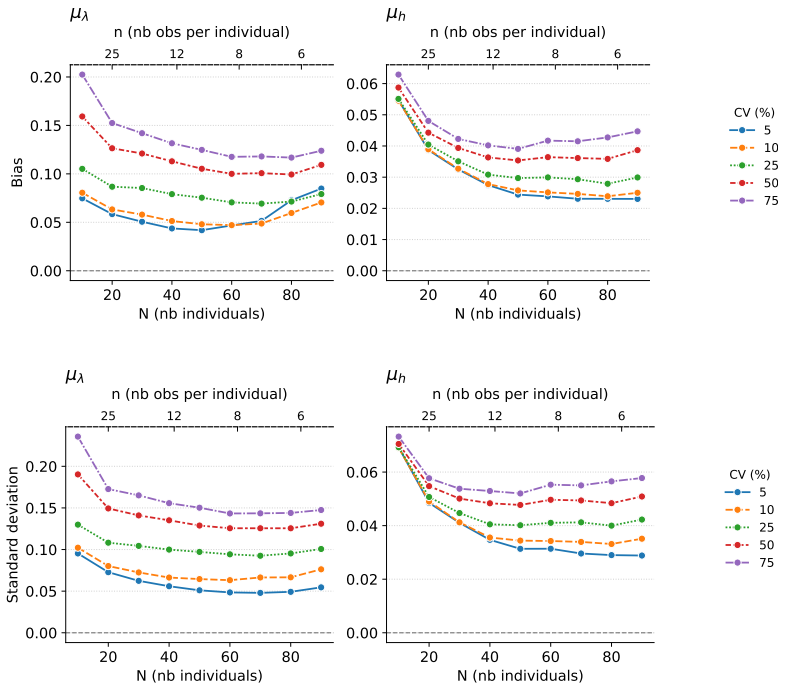


Figure 9: (top) Bias (in absolute value) of each mean parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the residual variability and the parameter set $\theta_1$, computed over $K = 1000$ repetitions. Only values with biases greater than 1 are shown (for reasons of scale). (bottom) Standard deviation of each mean parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the residual variability and the parameter set $\theta_1$, computed over $K = 1000$ repetitions. Only values with standard deviations greater than 1 are shown (for reasons of scale).

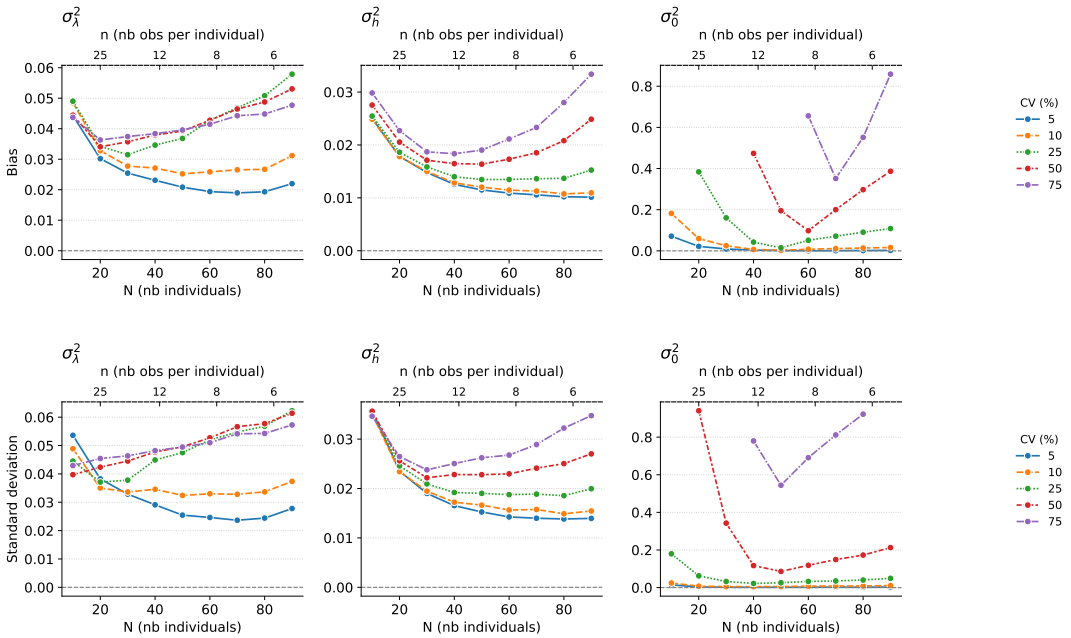


Figure 10: (top) Bias (in absolute value) of each variance parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the random effects and the parameter set $\theta_1$, computed over $K = 1000$ repetitions. (bottom) Standard deviation of each variance parameter as a function of the sample structure, *i.e.* the number of individuals in the population (principal x-axis) and the number of observations per individual (secondary x-axis), for different coefficients of variation for the residual variability and the parameter set $\theta_1$, computed over $K = 1000$ repetitions.

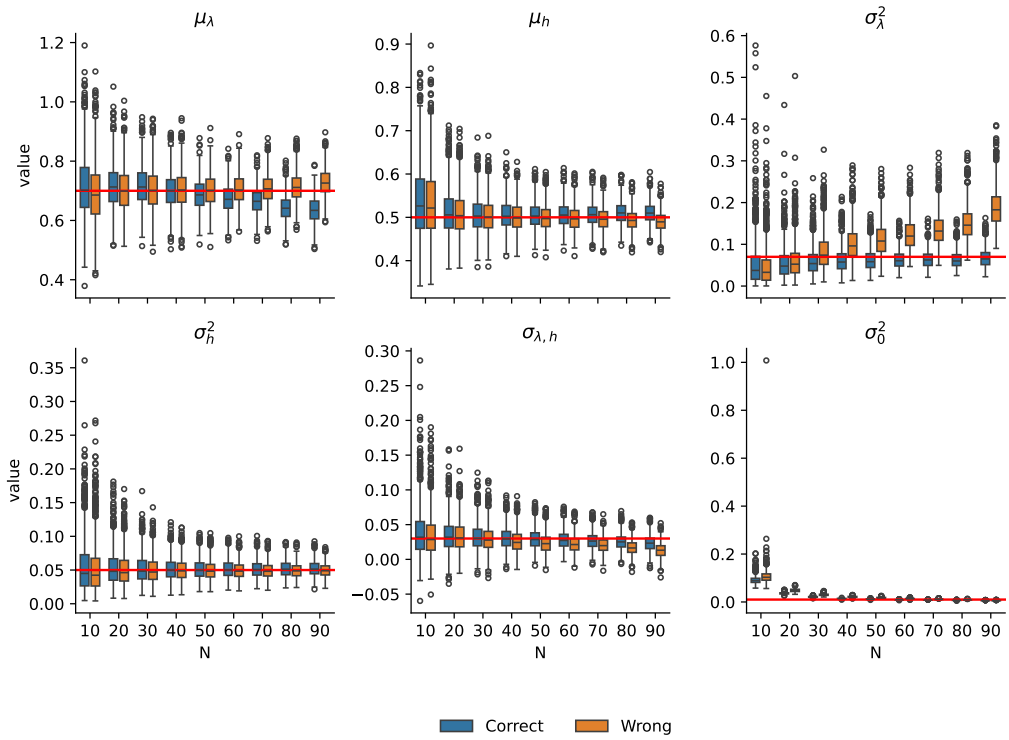


Figure 11: Effect of a misspecification of the noise structure on the parameter estimates, as a function of the sample size $N$. In blue, parameter estimates in the correctly specified model and in orange, parameter estimated in the misspecified model, over $K = 1000$ datasets

be explained by the fact that the extra-variability is absorbed by $\sigma_\lambda$.

## 3.2 Model selection

We only focus on model selection with respect to the noise structure. Recall that the true data-generating process combines both mechanistic variability $v(\cdot)$ and residual measurement error $\sigma_0^2$. We compare this full model to a simpler alternative that includes only residual error, without accounting for mechanistic variability, for different signal-to-noise ratio. Simulation results are provided in Tab. 1 for different levels of random effects variability and in Tab. 2 for different levels of residual noise. They show that, regardless of the dataset structure (values of $N$ and $n$), the ability to identify mechanistic variability decreases as residual variance increases — in other words, distinguishing between mechanistic and residual sources of variation becomes increasingly difficult. We also observe that higher inter-individual variability similarly reduces the ability to detect the presence of mechanistic noise. Similar experiments were made when the random effects are assumed to be correlated, and led to similar results.

Table 1: Proportion of datasets for which the smallest BIC is associated to the correct model for different coefficient of variation (cv) for the random effects.

| cv \ $N$ | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| 25 | 0.56 | 0.64 | 0.83 | 0.96 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 |
| 50 | 0.47 | 0.45 | 0.61 | 0.78 | 0.83 | 0.90 | 0.92 | 0.95 | 0.92 |
| 75 | 0.44 | 0.38 | 0.39 | 0.37 | 0.36 | 0.36 | 0.34 | 0.37 | 0.32 |

Table 2: Proportion of datasets for which the smallest BIC is associated to the correct model for different noise to signal ratios

| cv \ $N$ | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.51 | 0.54 | 0.72 | 0.90 | 0.95 | 0.99 | 0.99 | 0.99 | 0.98 |
| 10 | 0.50 | 0.52 | 0.61 | 0.75 | 0.83 | 0.90 | 0.92 | 0.95 | 0.94 |
| 25 | 0.47 | 0.50 | 0.52 | 0.55 | 0.56 | 0.60 | 0.62 | 0.65 | 0.66 |
| 50 | 0.50 | 0.50 | 0.50 | 0.51 | 0.50 | 0.51 | 0.50 | 0.50 | 0.49 |
| 75 | 0.51 | 0.50 | 0.50 | 0.51 | 0.50 | 0.48 | 0.49 | 0.47 | 0.47 |

### 3.3 Application on a real dataset [32]

We fitted the 3-dimensional ($D = 3$, Eqs. 4 and 5) functional response model to the *A. salina* (prey) vs. *H. formosa* (predator) data, with $\lambda$ and $h$ as correlated random effects and considering a mechanistic term in the residual variance (model 2). As the SGD algorithm can converge to a local optimum, we ran ten repetitions of the algorithm with random initialization. As results were very similar for all runs (see Figs. A.1 and A.2), we averaged the estimates from the 10 runs.

Table 3 gives the parameter estimates along with their 95% confidence intervals (CIs), with or without considering a mechanistic source for the residual variance. The CI were computed using the inverse of the Fisher Information matrix, which is computed by the SGD algorithm, and the delta-method. The introduction of mechanistic noise has an impact on the variance component estimates: $\sigma_\lambda$ and $\sigma_0$ are estimated a bit smaller while $\sigma_h$ and the correlation coefficient $\sigma_{\lambda,h}$ are estimated a bit larger in this case. Moreover, the CIs are smaller when mechanistic noise is included in the model, especially for the variance components parameters.

The residual variability $\sigma_0$ is, by far, estimated to be the most important source of variation. However, care must be taken in the interpretation of the different levels of variability since the scales of the different processes are different. For example, when related to the estimated asymptotic number of preys ingested, i.e. to $1/\mu_h = 80$, the residual

variability is approximately 12% of this asymptotic response, which is comparable to the levels of residual variabilities that were tested on the simulated datasets. We computed the BIC associated with the average parameter estimates obtained over the 10 repetitions, and results are reported in Table 3. The BICs are very comparable between the two models (a difference of 6 corresponds to a difference of 3 in the log-likelihood scale). When comparing the individual BIC values obtained on the 10 repetitions for the two models, results are even more comparable, which strengthens the fact that the two models share very similar performances with respect to data fitting.

| Parameter | No mechanistic noise | | Mechanistic noise | |
|---|---|---|---|---|
| | Estimate | 95% CI | Estimate | 95% CI |
| $\mu_\lambda$ | 0.1307 | [0.1119 ; 0.1496] | 0.1259 | [0.1058 ; 0.1460] |
| $\mu_h$ | 0.0126 | [0.0096 ; 0.0155] | 0.0124 | [0.0099 ; 0.0148] |
| $\sigma_\lambda$ | 0.0453 | [0.0280 ; 0.0626] | 0.0397 | [0.0262 ; 0.0533] |
| $\sigma_h$ | 0.0069 | [0.0058 ; 0.0079] | 0.0074 | [0.0065 ; 0.0084] |
| $\sigma_{\lambda,h}$ | 0.7751 | [0.6370 ; 0.9132] | 0.7844 | [0.7073 ; 0.8615] |
| $\sigma_0$ | 9.8380 | [9.1773 ; 10.499] | 9.7615 | [9.1238 ; 10.400] |
| BIC | 2015.10 | | 2021.39 | |

Table 3: 95% confidence intervals for the parameters of the 3D functional model on the real data set.

To assess the ability of the model to capture the variability of the population, several diagnostic plots were be performed. First, individual predictions were compared to observed data (Fig. 12), where the individual predictions were obtained using the estimated individual parameters $\hat{\varphi}_i$ and the mean functional response $r$, i.e. the individual predictions for individual $i$ at prey density $j$ is $r(\hat{\varphi}_i, d_{ij}, e)$ (see Eq. 4). Estimations of the individual parameters were obtained as a by-product of the SGD algorithm, using the samples $\varphi_i^K$ obtained at the last iteration of the algorithm. To account for the intrinsic stochasticity of the functional response model when mechanistic noise is considered, we simulated 100 realizations of the stochastic model for each individual. As illustrated in Fig. 12, the inter-individual variability is well captured by the functional response model, which is able to reproduce the specific behavior of each individual in the population. The individual predictions obtained from the model fitted with or without considering a mechanistic noise are very similar.

Second, we used the Visual Predictive Check (VPC) [39] to assess the ability of the model to correctly reproduce the observed data (Fig. 13). In a VPC, the theoretical quantiles obtained from the model are compared to the empirical quantiles obtained from the observed data. More precisely, the procedure is as follows: i) we generated 5000 simulated datasets of 49 individuals (i.e. the same sample size as the observed data) under the model with its estimated parameters; ii) for each simulated dataset and each prey density, we computed the 5%, 50% and 95% quantiles of the number of ingested preys over the 49 individuals of the population; and iii) we computed the 5%, 50% and 95% quantiles of the quantities

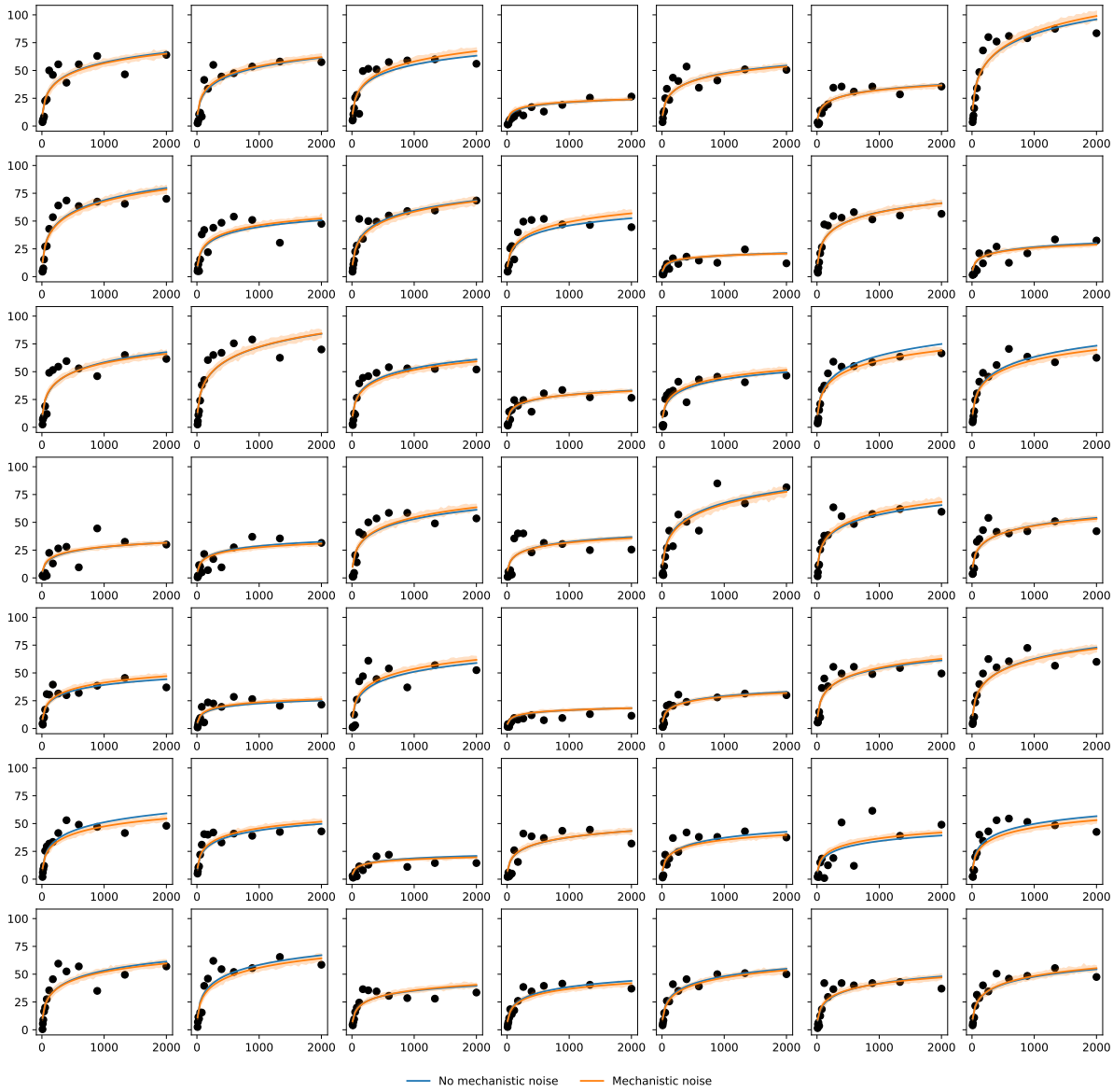


Figure 12: Individual predictions using the functional model run with the individual parameters, for each fish in the population.

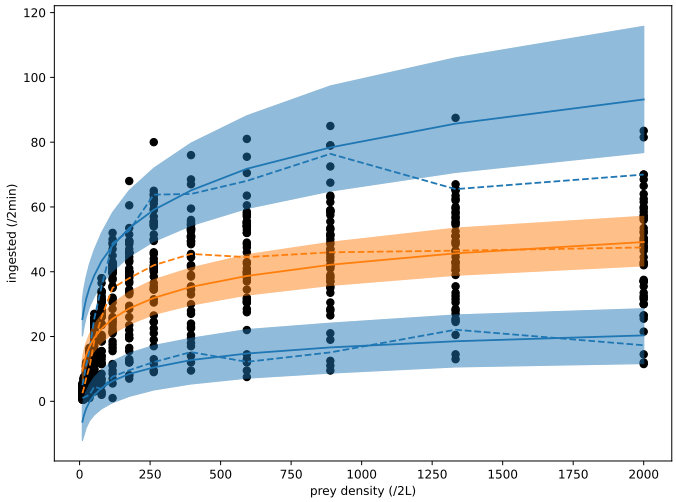


Figure 13: Visual Predictive Check (VPC) of the 3D functional model with mechanistic noise on the real data set. In blue, the distribution of the 5% and 95% theoretical quantiles obtained from the model, in orange, the distribution of the theoretical median and in black dots, the observed data.

obtained at the previous stage over the 5000 repetitions. Results are very similar in both models (i.e. with or without mechanistic noise). Overall, the model is able to capture the intra-individual variability, with a slight tendency to over-estimate the upper tail of the distribution, i.e. individuals with the highest asymptotic number of preys ingested. This over-estimation is partly due to the fact that, from a prey density of approx. 800 onward, the curve of the empirical 95% quantile starts to drop, which might be explained by the small sample size. Indeed, these empirical quantiles are computed, at each observed prey density, on a total of 49 points. With a higher number of observations, we might be able to identify whether this overestimation comes from the model.

# 4 Discussion

In this study, we aimed to disentangle the various sources of variation in functional response models which remains a major challenge in population ecology [35, 12, Chap. 10.2]. We developed a heteroscedastic mixed non-linear model, where both the deterministic component and the variance structure were derived from an underlying stochastic mechanistic model. This approach explicitly incorporated all key sources of variation: individual differences, observational errors, and intrinsic stochasticity of the process.

**Model specification and mechanistic intrinsic noise.** We showed that it is possible to estimate the different sources of variability under realistic experimental conditions: with a few hundred observations, around a dozen individuals, and a limited number of param-

eters to estimate. We also showed that neglecting mechanistic stochasticity can introduce estimation biases, particularly when the experimental design is unbalanced with respect to the number of replicates and the number of individuals. This is because the model components and the sources of variation are interdependent and can exhibit compensatory behavior. Consequently, failing to account for mechanistic stochasticity may result in an overestimation of residual variability or a misattribution of observed variation to individual differences, ultimately distorting both parameter estimates and biological interpretations.

As an illustration, applying our model to a real dataset resulted in the estimation of a large residual variance, $\sigma_0^2$. Although this variance scaled appropriately with the mean asymptotic number of prey consumed, it may also indicate that the mechanistic variance was underestimated especially given that its estimate remained largely unchanged when additional mechanistic noise was incorporated into the model. In theory, $\sigma_0$ should capture only measurement error. However, in practice, it also absorbs model bias, that is, the discrepancy between the proposed model and the true underlying biological process. While improving measurement procedures can help reduce observational error, residual variability may still remain high due to an inadequate or misspecified variance structure, or from omitting intrinsic sources of mechanistic variability. Evidence of such misspecification is suggested by the overestimation of upper quantiles at high prey densities in the VPC checks (Fig. 13). This pattern indicates that the number of prey consumed may plateau for predators with high consumption rates, potentially reflecting a satiety effect not accounted for in the current mechanistic model.

Another potential model misspecification explaining the large estimated value of the residual variance $\sigma_0$ could stem from the stochastic process model itself as it inherently limits mechanistic noise: the mean $r$ increases with $d^{1/3}$ while the variance decreases with $d^{-2/3}$. The fitting procedure may thus compensate by attributing most of the variability to measurement error, inflating $\sigma_0$. This underestimation of mechanistic variability may result from simplifying assumptions in the model, such as prey renewal and the absence of spatial structure, whereas the experimental setup involves prey depletion and space. As noted by [4], foraging in spatially structured environments with prey depletion can lead to significant inherent variability in functional responses, consistent with our findings. In a different ecological context, [12, Chap. 10.2] conducted stochastic simulations assuming that the time between prey–predator interactions followed an exponential distribution—that is, without explicitly modeling the foraging process or accounting for spatial structure. He concluded that stochastic foraging alone could not generate the observed variability in empirical data. However, we argue that such a conclusion cannot be drawn without specifying a proper statistical model, derived from a mechanistic stochastic process, that explicitly accounts for all relevant sources of variability, as we did here.

**Experimental design.** We also showed that successfully disentangling the different sources of variation requires careful experimental design—specifically, balancing the trade-off between the number of individuals ($N$) and the number of observations per individual ($n$). This consideration is particularly important for accurately estimating the mean prey search speed parameter, $\mu_\lambda$, as well as the parameters associated with inter-individual variability,

$\sigma_\lambda$ and $\sigma_h$, especially under low signal-to-noise conditions (i.e., when $\sigma_0$ is large). Given that foraging trials typically involve using each individual only once [12], our results underscore the need to reconsider conventional approaches to experimental design in this context.

**Extensions and perspectives.** We focused in this study mostly on the importance of the experiment design and on comparing inference considering or not the intrinsic and mechanistic source of variation. Our statistical framework can also be used to address other issues such as comparing different models for testing the dimensionality of foraging, including covariates such as individual weights or behaviors, or assessing the presence or importance of inter-individual variability in specific parameters. The standard BIC penalty may need to be adapted in those cases. In hierarchical models, the number of effective sample size is not always straightforward to define, especially when random effects are introduced selectively across model components. Several extensions of the BIC have been proposed to account for such structures, adjusting the penalty term to better reflect the complexity induced by individual variability. These adaptations provide more reliable guidance for model selection when testing for random effects or comparing models with different covariance structures [10, 11]. Finally, an open question concerns the scaling parameter $\Delta$, which arises naturally in the derivation of the stochastic process approximation and carries over into the statistical model. In our approach, we assigned $\Delta$ a fixed value determined externally by the experimental setup, making it the parameter that sets the measurement scale and provides quantitative meaning to the estimated parameters. However, one might ask whether $\Delta$ could instead be treated as an internal parameter to be estimated within the statistical framework itself. In that case, issues of identifiability would inevitably arise.

## Acknowledgements

This work was supported by the Chair 'Modélisation Mathématique et Biodiversité' of Veolia Environnement-Ecole Polytechnique-Museum National d'Histoire Naturelle-Fondation X, by the ANR ABIM (ANR-16-CE40-0001) and by the grant 80PRIME 2023 of the CNRS. The authors acknowledge the support of the CDP C2EMPI, as well as the French State under the France-2030 programme, the University of Lille, the Initiative of Excellence of the University of Lille, the European Metropolis of Lille for their funding and support of the R-CDP-24-004-C2EMPI project. The authors thank Vincent Bansaye, Jean-René Chazottes and Geoffroy Berthelot for helpful discussions, and Arne Schröder and Gregor Kalinkat for kindly sharing their dataset. This paper is dedicated to Elizabeta Vergu who passed away in the early stage of the current work.

## References

[1] Charlotte Baey, Maud Delattre, Estelle Kuhn, Jean-Benoist Leger, and Sarah Lemler. Efficient preconditioned stochastic gradient descent for estimation in latent variable models. In International Conference on Machine Learning, pages 1430–1453. PMLR, 2023.

[2] Charlotte Baey and Estelle Kuhn. vartestnlme: An r package for variance components testing in linear and nonlinear mixed-effects models. arXiv preprint arXiv:2007.04791, 2020.

[3] D.J. Baker, R.A. Stillman, B.M. Smith, J.M. Bullock, and K.J. Norris. Vigilance and the functional response of granivorous foragers. Functional Ecology, 24:1281–1290, 2010.

[4] Vincent Bansaye, Geoffroy Berthelot, Amina El Bachari, Jean-René Chazottes, and Sylvain Billiard. Stochastic foraging paths primarily drives within-species variations of prey consumption rates. bioRxiv, pages 2024–05, 2024.

[5] Andréanne Beardsell, Dominique Gravel, Jeanne Clermont, Dominique Berteaux, Gilles Gauthier, and Joël Bêty. A mechanistic model of functional response provides new insights into indirect interactions among arctic tundra prey. Ecology, 103(8):e3734, 2022.

[6] Sylvain Billiard, Vincent Bansaye, and J-R Chazottes. Rejuvenating functional responses with renewal theory. Journal of The Royal Society Interface, 15(146):20180239, 2018.

[7] Jérôme Casas and Beat Hulliger. Statistical analysis of functional response experiments. Biocontrol Science and Technology, 4(2):133–145, 1994.

[8] K Chan, S Boutin, TJ Hossie, CJ Krebs, M O'Donoghue, and DL Murray. Improving the assessment of predator functional responses by considering alternate prey and predator interactions. Ecology, 98(7):1787–1796, 2017.

[9] Emmanuelle Comets, Audrey Lavenu, and Marc Lavielle. Parameter estimation in nonlinear mixed effect models using saemix, an r implementation of the saem algorithm. Journal of Statistical Software, 80:1–41, 2017.

[10] Maud Delattre, Marc Lavielle, and Marie-Anne Poursat. A note on bic in mixed-effects models. 2014.

[11] Maud Delattre and Marie-Anne Poursat. An iterative algorithm for joint covariate and random effect selection in mixed effects models. The international journal of biostatistics, 16(2):20190082, 2020.

[12] John P DeLong. Predator ecology: Evolutionary ecology of the functional response. Oxford University Press, 2021.

[13] Bernard Delyon, Marc Lavielle, and Eric Moulines. Convergence of a stochastic approximation version of the em algorithm. Annals of statistics, pages 94–128, 1999.

[14] John S Fenlon and Malcolm J Faddy. Modelling predation in functional response. Ecological modelling, 198(1-2):154–162, 2006.

[15] MHJP Fernando and MP Hassell. Predator-prey responses in an acarine system. Population Ecology, 22(2):301–322, 1980.

[16] C.S. Holling. Some characteristics of simple types of predation and parasitism. Canadian Entomologist, 91:385–398, 1959.

[17] J.M. Jeschke, M. Kopp, and R. Tollrian. Predator functional responses: Discriminating between handling and digesting prey. Ecological Monographs, 72:95–112, 2002.

[18] Christian Jost and Roger Arditi. Identifying predator–prey processes from time-series. Theoretical population biology, 57(4):325–337, 2000.

[19] Estelle Kuhn and Marc Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. Computational statistics & data analysis, 49(4):1020–1038, 2005.

[20] Marc Lavielle. Mixed effects models for the population approach: models, tasks, methods and tools. CRC press, 2014.

[21] Stefan M Linzmaier and Jonathan M Jeschke. Towards a mechanistic understanding of individual-level functional responses: Invasive crayfish as model organisms. Freshwater Biology, 65(4):657–673, 2020.

[22] JAJ Metz, MW Sabelis, and JH Kuchlein. Sources of variation in predation rates at high prey densities: an analytic model and a mite example. Experimental & applied acarology, 5:187–205, 1988.

[23] Julien Mocq, Pavel R Soukup, Joacim Näslund, and David S Boukal. Disentangling the nonlinear effects of habitat complexity on functional responses. Journal of Animal Ecology, 90(6):1525–1537, 2021.

[24] Erlend B Nilsen, John DC Linnell, John Odden, and Reidar Andersen. Climate, season, and social status modulate the functional response of an efficient stalking predator: the eurasian lynx. Journal of Animal Ecology, 78(4):741–751, 2009.

[25] Mark Novak and Daniel B Stouffer. Systematic bias in studies of consumer functional responses. Ecology Letters, 24(3):580–593, 2021.

[26] Mark O'Donoghue, Stan Boutin, Charles J Krebs, Gustavo Zuleta, Dennis L Murray, and Elizabeth J Hofer. Functional responses of coyotes and lynx to the snowshoe hare cycle. Ecology, 79(4):1193–1208, 1998.

[27] José Pinheiro and Douglas Bates. Mixed-effects models in S and S-PLUS. Springer science & business media, 2006.

[28] Björn C Rall, Ulrich Brose, Martin Hartvig, Gregor Kalinkat, Florian Schwarzmüller, Olivera Vucic-Pestic, and Owen L Petchey. Universal temperature and body-mass scaling of feeding rates. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1605):2923–2934, 2012.

[29] 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.

[30] Wendy A Ruscoe, Joseph S Elkinton, David Choquenot, and Robert B Allen. Predation of beech seed by mice: effects of numerical and functional responses. Journal of Animal Ecology, pages 1005–1019, 2005.

[31] A. Schröder, G. Kalinkat, and R. Arlinghaus. Individual variation in functional response parameters is explained by body size but not by behavioural types in a poeciliid fish. Oecologia, 182:1129–1140, 2016.

[32] Arne Schröder, Gregor Kalinkat, and Robert Arlinghaus. Individual variation in functional response parameters is explained by body size but not by behavioural types in a poeciliid fish. Oecologia, 182(4):1129–1140, 2016.

[33] G.T. Skalski and J.F. Gilliam. Functional responses with predator interference: viable alternatives to the Holling Type II model. Ecology, 82:3083–3092, 2001.

[34] N Thompson Hobbs, John E Gross, Lisa A Shipley, Donald E Spalinger, and Bruce A Wunder. Herbivore functional response in heterogeneous environments: a contest among models. Ecology, 84(3):666–681, 2003.

[35] Joel C Trexler, Charles E McCulloch, and Joseph Travis. How can the functional reponse best be determined? Oecologia, 76:206–214, 1988.

[36] Stella F Uiterwaal and John P DeLong. Multiple factors, including arena size, shape the functional responses of ladybird beetles. Journal of Applied Ecology, 55(5):2429–2438, 2018.

[37] Stella F Uiterwaal and John P DeLong. Functional responses are maximized at intermediate temperatures. Ecology, 101(4):e02975, 2020.

[38] Stella F. Uiterwaal, Ian T. Lagerstrom, Shelby R. Lyon, and John P. DeLong. Forage database: A compilation of functional responses for consumers and parasitoids. Ecology, 103(7):e3706, 2022.

[39] Yoshitaka Yano, Stuart L Beal, and Lewis B Sheiner. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. Journal of pharmacokinetics and pharmacodynamics, 28:171–192, 2001.

# A Appendix: detailed implementations of algorithms

## A.1 Parameter estimation

To describe the algorithms, we introduce the following additional notations, which complement those used in the main body of the paper. Let $y_i$ denote the vector of observations for subject $i$, where $i = 1, \ldots, N$. Also, let $p_\theta(\cdot|y_i)$ represent the conditional density of $\varphi_i$ given $y_i$ for a fixed value $\theta$ of the model parameter, and $f(y_i, \varphi_i; \theta)$ the joint density of $y_i$ and $\varphi_i$ for the parameter value $\theta$. These distributions depend on model (1) specification. The two algorithms below iterate for $K$ iterations, producing a sequence of estimates of the unknown parameter $\theta$, denoted by $(\theta_k)$, using a decreasing sequence of step sizes $(\gamma_k)$. Guidelines for choosing these step sizes can be found in [1] for the stochastic gradient descent and [13] for the SAEM algorithm. An important aspect to understand about the operation of these algorithms is that $\varphi_i$ is a latent variable, meaning that we do not have an observed value for it. Consequently, it must be simulated at each iteration. It is simulated either from its exact conditional distribution $p_\theta(\cdot|y_i)$ or through a small number of iterations of an MCMC procedure with kernel $\Pi_\theta(\cdot|\cdot, y_i)$, assumed to have the conditional distribution as its stationary law.

### A.1.1 Stochastic gradient descent

**for** $k = 1, \ldots, K$ **do**
  **for** $i = 1, \ldots, N$ **do**
    $\varphi_i^k \leftarrow$ sample from $p_{\theta_{k-1}}(\cdot \mid y_i)$ or $\Pi_{\theta_{k-1}}(\cdot \mid \cdot, y_i)$
  **end for**
  $v_k \leftarrow \frac{1}{N} \sum_{i=1}^{N} \nabla_\theta \log f(y_i, \varphi_i^k; \theta_{k-1})$
  **for** $i = 1, \ldots, N$ **do**
    $\Delta_i^k \leftarrow (1 - \gamma_k)\Delta_i^{k-1} + \gamma_k \nabla_\theta \log f(y_i, \varphi_i^k; \theta_{k-1})$
  **end for**
  $I_k \leftarrow \frac{1}{N} \sum_{i=1}^{N} \Delta_i^k (\Delta_i^k)^T$
  $\theta_k \leftarrow \theta_{k-1} + \gamma_k I_k^{-1} v_k$
**end for**

The final estimates of the Fisher information matrix and the model parameters correspond to quantities $I_K$ and $\theta_K$ evaluated at the last iteration, assuming the algorithm has converged.

### A.1.2 Stochastic approximation EM (SAEM) algorithm

**for** $k = 1, \ldots, K$ **do**
  **for** $i = 1, \ldots, N$ **do**
    $\varphi_i^k \leftarrow$ sample from $p_{\theta_{k-1}}(\cdot \mid y_i)$ or $\Pi_{\theta_{k-1}}(\cdot \mid \cdot, y_i)$

**end for**

$Q_k(\theta) \leftarrow (1-\gamma_k)Q_{k-1}(\theta) + \gamma_k \sum_{i=1}^{N} \log f(y_i, \varphi_i^k; \theta_{k-1})$

$\theta_k \leftarrow \text{argmax}_\theta \, Q_k(\theta)$

**end for**

Note that for this algorithm, the implementation of an iteration is significantly simplified when the model belongs to the curved exponential family (see [19]).

The final estimate of the model parameters correspond to $\theta_K$ evaluated at the last iteration, assuming the algorithm has converged.

### A.2 Log-likelihood estimation

To estimate the log-likelihood in a nonlinear mixed-effects model, we need to evaluate integrals over the random effects. The marginal likelihood is given by:

$$L(\theta) = \prod_{i=1}^{N} L_i(\theta),$$
$$L_i(\theta) = \int f(y_i, \varphi_i; \theta) d\varphi_i, \; i = 1, \ldots, N.$$

These integrals are generally intractable. We can use Monte-Carlo methods such as Importance Sampling to approximate them. The idea is to draw many samples of the random effects, $\varphi_i^{(m)}$, $m = 1, \ldots, M$, $i = 1, \ldots, N$, from a known, easy-to-sample distribution $\tilde{\Pi}(\cdot; \theta)$ and then weight each sample according to the following formula:

$$\widehat{L_i(\theta)} = \frac{1}{M} \sum_{m=1}^{M}, \frac{f(y_i, \varphi_i^{(m)}; \theta)}{\tilde{\Pi}(\varphi_i^{(m)}; \theta)}, \; i = 1, \ldots, N.$$

The log-likelihood is then approximated by

$$\mathcal{L}(\theta) \approx \sum_{i=1}^{N} \log(\widehat{L_i(\theta)}).$$

A critical aspect of this method is the choice of the importance distribution $\tilde{\Pi}(\cdot; \theta)$, which should be close enough to the conditional distribution of the random effects given the observations to ensure estimation stability. In practice, there is no universal choice: selecting a good importance distribution often involves trial and error, along with stability checks.

## B Convergence of the SGD algorithm on the real dataset

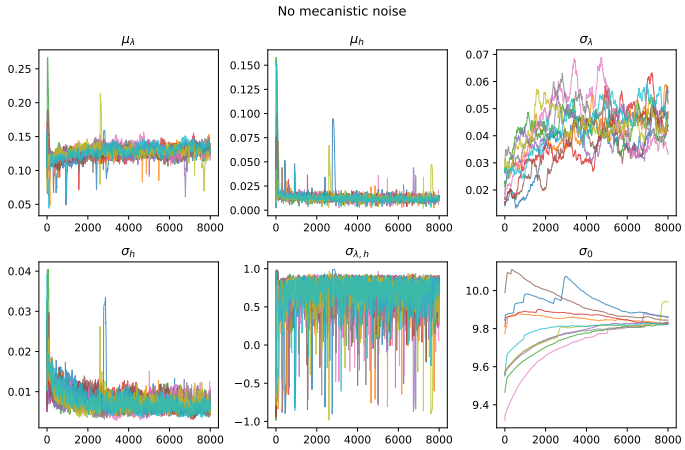


Figure A.1: Evolution of the parameter estimates along the iterations of the SGD algorithms, for each of the ten runs corresponding to ten different initializations, when no mechanistic noise is included in the model

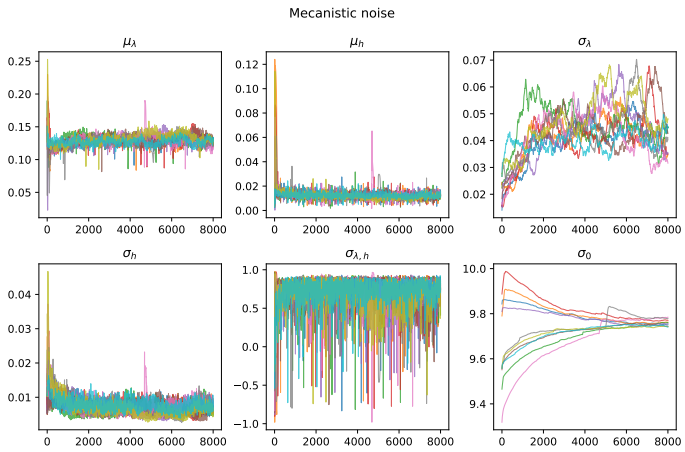


Figure A.2: Evolution of the parameter estimates along the iterations of the SGD algorithms, for each of the ten runs corresponding to ten different initializations, when mechanistic noise is included in the model