This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Tirronen, Maria; Kuparinen, Anna

Title: Parameter estimation for allometric trophic network models : A variational Bayesian inverse problem approach

Year: 2024

Version: Published version

Please cite the original version:

Tirronen, M., & Kuparinen, A. (2024). Parameter estimation for allometric trophic network models : A variational Bayesian inverse problem approach. Methods in Ecology and Evolution, Early online. https://doi.org/10.1111/2041-210x.14447

RESEARCH ARTICLE

Parameter estimation for allometric trophic network models: A variational Bayesian inverse problem approach

Maria Tirronen[1,2] | Anna Kuparinen[1]

[1]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

[2]Natural Resources Institute Finland (Luke), Jyväskylä, Finland

Correspondence
Maria Tirronen
Email: maria.tirronen@gmail.com

Abstract

1. Differential equation models are powerful tools for predicting biological systems, capable of projecting far into the future and incorporating data recorded at arbitrary times. However, estimating these models' parameters from observations can be challenging because numerical methods are often required to approximate their solution. An example of such a model is the allometric trophic network model, for which studies considering its inverse problem are limited, particularly in the Bayesian framework. Here we develop a variational Bayesian method for parameter inference of the allometric trophic network model and explore how accurately we can recover its parameter values.

2. We represent the model as a Bayesian neural network, which combines an artificial neural network with Bayesian inference, using a surrogate for the posterior distribution of model parameters, and train this model by evolutionary optimization to avoid potentially costly computation of the gradient with respect to the model parameters. Using synthetic data, we compare the accuracy of this variational inference to ordinary least squares estimation. To reduce the number of estimated parameters, we focus on the inference of functional response parameters.

3. Our variational Bayesian method yields parameter estimates that are comparable to the ordinary least squares results in terms of accuracy. The method provides a promising approach for including uncertainty quantification in parameter estimation, which the simple ordinary least squares approach as it is does not address. Regardless of the method, potential multimodality of the inference problem is nonetheless important to keep in mind.

4. The present study suggests a technique for parameter inference of ordinary differential equation models in the Bayesian context. We propose the method especially for validation of the allometric trophic network model against empirical data.

KEYWORDS
allometric trophic network, Bayesian inference, Bayesian inverse problem, Bayesian neural network, evolutionary optimization, food web, ODE system

1 | INTRODUCTION

Predictive modelling has been widely used in ecology for decades and its utility continues to increase as the need for ecological forecasts that can inform ecosystem management and conservation is expanding (Barros et al., 2023; Bodner et al., 2021; Dietze, 2017). The task for forecasting is imperative today, as biodiversity losses, unexpected fires, and extreme weathers are testing the resilience and recovery ability of natural systems. To these ends, potential tools for making future predictions about biological systems are continuous-time models that describe the dynamics of the system over time by differential equations. In principle, such models can be used for making predictions far to the future and are able to incorporate data recorded at arbitrary times (Chen et al., 2018). Nonetheless, estimation of these models' parameters from observations can be challenging because they are often complex and do not permit an analytical solution, therefore requiring numerical methods to approximate their solution (e.g. Chen et al., 2018; Ghosh et al., 2021; Valderrama-Bahamóndez & Fröhlich, 2019). This may limit validation of the predictive ability of these models against empirical data.

An example of such a differential equation model is the bioenergetic (Yodzis & Innes, 1992), or allometric trophic network (ATN) model (Boit et al., 2012; Brose et al., 2006), which offers powerful tools for predicting the dynamic interactions within ecosystems. Otherwise similar to the traditional Lotka–Volterra (Lotka, 1925; Volterra, 1926) predator–prey model, which describes the dynamics of interacting predator and prey populations by ordinary differential equations (ODEs), the ATN model adjusts predator consumption to prey density and relates the model parameters to species' body masses and their metabolic categories using allometric relationships (Williams et al., 2006). By encapsulating the principles of energy flow and biomass distribution across different trophic levels in a general manner, these models not only contribute to the development of basic theory of dynamically interacting species (Boit et al., 2012) but also provide critical insights into the mechanisms underlying ecological stability (Brose et al., 2006) and the impact of anthropogenic changes (e.g. Eloranta et al., 2023; Kuparinen et al., 2023; Perälä et al., 2023; Uusi-Heikkilä et al., 2022).

Despite the ATN model has recently gained a lot of interest in research (e.g. Ávila Thieme et al., 2021; Kath et al., 2018; Martinez, 2020, in addition to the aforementioned papers), only a few studies (Banks et al., 2016, 2017; Boit et al., 2012; Koen-Alonso & Yodzis, 2005) have considered whether the ATN model can predict abundances in a real system. These studies range from modelling part of an ecosystem (Koen-Alonso & Yodzis, 2005) with only a few species to a large whole lake system with 24 functional groups (guilds). Overall, these studies obtained good resemblance with the empirical data but also reported inaccuracy in some cases (Banks et al., 2016, 2017; Koen-Alonso & Yodzis, 2005).

To calibrate the ATN model to empirical data, the studies with trophic networks of a relatively small size and low number of ATN model parameters to be estimated (Banks et al., 2016, 2017; Koen-Alonso & Yodzis, 2005) utilized numerical optimization methods, while the authors of the larger model (Boit et al., 2012) reported to have calibrated their unknown parameters manually. Manual calibration usually refers to the comparison of results obtained by using different parameter values, chosen by the modeller. Specifically, the studies utilizing automatic calibration (Banks et al., 2016, 2017; Koen-Alonso & Yodzis, 2005) applied the ordinary least squares (OLS) or maximum likelihood estimation (MLE). While these studies included sensitivity analysis (Banks et al., 2016, 2017) or exploration of the likelihood for parameters in the neighbourhood of its maximum (Koen-Alonso & Yodzis, 2005) to address parameter uncertainty and its impact on the predicted dynamics, a strictly Bayesian approach has not been applied to the inverse problem of the ATN model. Furthermore, only Banks et al. (2016, 2017) considered how accurately the parameter values can actually be recovered from observations. When testing the inverse problem of the ATN model on synthetic data, they found that the solution did not necessarily give the same parameter values as they used to generate the data.

Overall, calibration of the ATN model to observations can face several kind of challenges. In addition to the computational challenges associated with the parameter inference of ODE models in general, the solution of the inverse problem of the ATN model can easily appear as non-unique, or, more generally, that the inverse problem of the ATN model is ill-posed (there is no uniqueness or stability of solutions; Kabanikhin, 2008). Indeed, as biological data are often very noisy and the ATN model can produce various kinds of dynamics from stable diminished oscillations to chaotic fluctuations and bifurcation (Williams & Martinez, 2004), the solution of the inverse problem can highly depend on, for example how much of the variation in our data we regard as 'noise' and how much we regard to orginate from oscillations or fluctuations in the underlying ATN dynamics. When using numerical optimization methods, parameter estimation for the ATN model may encounter challenges with multimodality, as the results by Banks et al. (2016, 2017) suggest. This occurs when multiple local optima exist, making it difficult to identify the global optimum. Traditional approaches like MLE typically aim to find a single set of parameter values that maximize the likelihood, which can result in selecting a local rather than the global optimum. These challenges can become considerable especially when the system is large and the ATN model have many parameters. On the other hand, the allometric relationships underlying the ATN model may enable reducing the set of estimated parameters to functional response parameters only (Boit et al., 2012). The functional response parameters are often notoriously difficult to measure directly owing to complex interactions, mechanisms and processes underlying feeding events (Williams, 2008). Nonetheless, they play a major role in defining the dynamics (Williams & Martinez, 2004), and the ATN inverse problem may still need additional constraints to reduce its ill-posedness (Wu et al., 2019). Moreover, to address the issue of multimodality, Bayesian methods, which aim to account for the full posterior distribution of parameters, might offer a potential advantage over methods that focus solely on parameter values near the maximum likelihood. However, the effectiveness of these methods

in fully capturing multimodal distributions can vary depending on the specific implementation.

The present study builds upon previous ATN studies and advancements in other fields to develop the first Bayesian approach for calibration of the ATN model to observed data. Aiming to reduce the computational cost of the problem, we resort to a variational inference (VI) method (Blei et al., 2017; Ghosh et al., 2021). Compared to Markov chain Monte Carlo (MCMC) methods, often applied in the Bayesian context, VI tends to be faster and easier to scale to large data (Blei et al., 2017). The performance of VI methods is obtained by using a surrogate for the posterior distribution of model parameters and by applying optimization methods to find the best parameter values of this surrogate distribution, given the data. By definition, these methods provide only approximations of the posterior, nonetheless, variational techniques have been found to provide good approximations in many applications, compared to MCMC results (Ghosh et al., 2021). A considerable challenge in VI, though, is the use of gradient-based optimization methods for ODE models as differentiation through the operations of numerical solvers can be computationally costly (Chen et al., 2018; Ghosh et al., 2021). For this reason, we combine VI with an optimization technique that does not use gradient information, specifically, an evolutionary optimization method. Such methods are population-based global optimization techniques that systematically pursue the identification of the best solution within a problem space using metaheuristics (Sloss & Gustafson, 2020) and only the function values. Often, evolutionary methods are used as last resort for inherently complex problems and when the size of problem domain is extreme, providing 'good enough' solutions in such circumstances (Sloss & Gustafson, 2020).

We demonstrate our method with synthetic data, generated by the ATN model, and explore how accurately the parameters of the ATN model can be recovered from these synthetic time series. Specifically, we compare our approach to OLS estimation. When using synthetic data, it is easy to evaluate the accuracy of the obtained parameter estimates as we know exactly the parameter values that produced the data. We generate the synthetic time series using trophic networks with 10 and 25 guilds, the latter size corresponding to the largest empirical trophic network against which the ATN model has been tested (Boit et al., 2012). To reduce the number of parameters to be estimated, we focus particularly on the functional response parameters as unknown. The number of estimated model parameters in our inverse problem corresponds to the studies by Boit et al. (2012) and Koen-Alonso and Yodzis (2005). Furthermore, we focus on systems with stabilized dynamics and minimal oscillations and constrain the problem accordingly. Overall, our study sets ground for parameter inference of the ATN and other complex ODE models from observed data in the Bayesian context.

2 | METHODS

We proceed as follows: First, we generate theoretical trophic networks with the niche model (Section 2.1.1). Next, for each network,

we simulate synthetic trajectories of biomasses (Section 2.1.2) and assume them to be an analogue of the empirical trajectories that could be obtained in a natural system. For this step, we use the ATN model and parameterize it using allometric rules and values from previous studies. For these first steps, we require persistence and stability of the simulated networks and their population dynamics (Section 2.1.3). Finally, we add noise to the simulated trajectories (Section 2.1.4) and apply a variational inference technique (Section 2.2.1) and OLS (Section 2.2.2), combined with evolutionary optimization (Section 2.2.3), to infer model parameters from the synthetic time series of 'observed biomasses'. We compare the synthetic time series data and parameter values used to generate them with the predicted ones (Section 3).

2.1 | Synthetic data

2.1.1 | Trophic network structure

To simulate synthetic biomass time series, we first generated three synthetic trophic networks with 10 guilds and one with 25 guilds (TN1–TN4; Section S1.1) using the niche model (Williams & Martinez, 2000) and utilizing the function *nichemodel* of the Julia package *BioEnergeticFoodWebs* (Delmas et al., 2017; Poisot et al., 2017). For the *nichemodel* algorithm, we set the target connectance to 0.15, which is within the range of connectance of empirical networks (Martinez, 1992); the realized connectances are close to this. When generating the networks, we accepted only the ones in which all producers were connected to the network and there was a path from every consumer guild to a basal producer. Studying the parameter estimation problem of the ATN model with three networks of the same size and approximately the same connectance gives us insight about the possible variation in the results within similar networks.

2.1.2 | ATN dynamics

After generating synthetic trophic networks, we next simulated biomass time series using the bioenergetic model (Williams, 2008) with allometric coefficients (Brose et al., 2006). The model is composed of the following set of ODEs that describe the dynamics of biomasses for producers (indices $\mathcal{J}^{\text{prod}}$) and consumers ($\mathcal{J}^{\text{cons}}$):

$$\frac{dB_i}{dt} = r_i B_i G_i(\boldsymbol{B}) - \sum_{j \in \mathcal{J}_i^{\text{cons}}} \frac{x_j(M_j) y_{ji} B_j F_{ji}(\boldsymbol{B})}{e_{ji}}, \quad i \in \mathcal{J}^{\text{prod}}, \tag{1}$$

$$\begin{aligned}\frac{dB_i}{dt} = &-x_i(M_i) B_i + x_i(M_i) B_i \sum_{j \in \mathcal{J}_i^{\text{res}}} y_{ij} F_{ij}(\boldsymbol{B}) \\ &- \sum_{j \in \mathcal{J}_i^{\text{cons}}} \frac{x_j(M_j) y_{ji} B_j F_{ji}(\boldsymbol{B})}{e_{ji}}.\end{aligned} \tag{2}$$

Above, t denotes time in the chosen scale and $B_i(t)$, $i \in \mathcal{J}^{\text{prod}} \cup \mathcal{J}^{\text{cons}}$, denotes the biomass of guild i at time t (omitted).

$\mathcal{J}_i^{\text{cons}}$ and $\mathcal{J}_i^{\text{res}}$ denote the indices of guilds that consume or are resources of guild i, respectively. In Equation (1), r_i denotes the intrinsic growth rate of the producer guild i, and G_i is the logistic function

$$G_i(\boldsymbol{B}) = 1 - \frac{\sum_{j\in\mathcal{J}^{\text{prod}}}c_{ij}B_j}{K}, \quad i \in \mathcal{J}^{\text{prod}}, \tag{3}$$

where c_{ij} is a producer competition coefficient and K is a system-wide carrying capacity. Moreover, x_i denotes the metabolic rate of consumer i, which we described by

$$x_i(M_i) = a\left(\frac{M_i}{M_0}\right)^{-A}, \tag{4}$$

with constants $a, A > 0, M_i$ the average body mass within guild i and M_0 the body mass of a reference producer guild. We set the body masses of guilds to be constant within trophic levels and to increase with increasing trophic levels by defining

$$\frac{M_i}{M_0} = 100^N, \quad i \in \mathcal{J}^{\text{cons}}, \tag{5}$$

where $N \in \{1, 2, \dots\}$ is the distance to a basal producer. Furthermore, in Equations (1) and (2), y_{ij} is the maximum consumption rate of a consumer guild i when feeding on guild j, e_{ij} is the corresponding assimilation efficiency, equal to the fraction of biomass of guild j lost for guild i that is actually metabolized, and F_{ij} is a functional response (FR) that describes the realized fraction of the maximum rate of consumption. We defined the latter one as Williams (2008)

$$F_{ij}(\boldsymbol{B}) = \frac{\left(B_j/B0_{ij}\right)^{q_{ij}}}{1 + \sum_{k\in\mathcal{J}_i^{\text{res}}}\left(B_k/B0_{ik}\right)^{q_{ik}}}, \tag{6}$$

where q_{ij} is a FR exponent that defines the type of the effect, and $B0_{ij}$ is a half-saturation constant, that is the density of resource guild j at which consumer guild i achieves half its maximum feeding rate.

We chose to simulate our synthetic systems in the time scale in which the biomasses of producers would intrinsically double, starting from initial biomass $B_i(0) = 1 \; \forall i$, and set the parameter values following previous studies and in such a way that all the guilds persisted through the simulation period and the dynamics were stable (Table S5; Section 2.1.3).

We approximated the biomass dynamics (1)–(2) of the synthetic trophic networks numerically. The computational model was implemented in Julia using *Tsit*5 of the package *DifferentialEquations* (Rackauckas & Nie, 2017) for numerical solution of the ODE model.

2.1.3 | Persistence and stability

In the simulation of synthetic time series (Section 2.1.2), we first considered the ATN dynamics for 2000 time steps. In this, we regarded a guild as extinct when it exceeded a value close to zero within the simulation period. Specifically, we considered whether $B_i(t) < 10^{-6}$ for some t, where the threshold 10^{-6} was an arbitrary choice. For our analysis, we selected only networks in which none of the guilds went extinct during the simulation period, that is the number of guilds in the system remained the same as was initially set.

In addition, we required that the oscillation of the generated dynamics was diminish at the end of this simulation period. To achieve this, we required that the coefficient of variation (CV; standard deviation per mean) of the simulated dynamics was 0.001 at maximum during the last 100 time steps, for each guild. On the other hand, some oscillation occurred in all generated systems so that the gradients (1)–(2) were never zero (Section S1.2). Moreover, we also required that the mean of the simulated biomasses during the last 100 time steps and during the preceding 100 days did not differ considerably (their absolute difference was 0.001 at maximum).

Moreover, we considered possible bistability of the generated systems by simulating data with initial values $B_i(0) = 0.1, 0.3, 0.5$ and 0.7, set equal for all of the guilds. In this, we did not notice bistability in the simulated time series (Section S1.2).

2.1.4 | Training and test data

For parameter inference, or model training, we recorded the synthetic abundances during the time steps $t = 1970, 1971, \dots, 2000$. We created random variation to the training data by simulating the time series as

$$\mathcal{D}_{i,t} \sim \mathcal{N}\left(B_i(t), s_i\right), \quad i \in \mathcal{J}^{\text{prod}} \cup \mathcal{J}^{\text{cons}}, \quad t = 1970, \dots, 2000, \tag{7}$$

where $\mathcal{N}(x, s)$ denotes the Gaussian distribution with mean x and standard deviation s. To test the performance of the methods under different amounts of noise, we generated data using different values of CV (Table S5).

Moreover, to test the predictive ability of the trained models, we simulated data for $t = 2001, \dots, 2030$, while introducing a disruption to the system at $t = 2003$. The disruption consisted of 50% reduction in biomasses for consumer guilds. Random variation was simulated for the test data sets similarly as above, using the same CV as for the training data.

2.2 | Parameter inference

We first estimated the half-saturation constants and FR exponents of the functional responses as well as the standard deviations of biomasses for the networks with 10 guilds (TN1–TN3; Table 1). Based on these results, we focused on the half-saturation constants for the larger network (TN4). The total number of estimated parameters varied among the tested networks (Table 1).

2.2.1 | Variational Bayes

For parameter inference in the Bayesian context, we took a variational approach based on the Bayes by Backpropagation method

TABLE 1 The number of guilds and the types and total number of estimated parameters for each synthetic trophic network we generated and tested for parameter estimation using the variational inference (VI) and ordinary least squares (OLS) methods.

ID	Number of guilds	Types of estimated parameters	Total number of estimated parameters
TN1	10	$B0^*$, q^*, s, σ	92 (36)
TN2			96 (38)
TN3			88 (34)
TN4	25	$B0^*$, σ	246 (123)

Note: For OLS, we estimated fewer parameters (marked by * and given in parenthesis) than for VI. The parameters are defined in Table S5 and Sections 2.1.2 and 2.2.1.

presented by Blundell et al. (2015) for artificial neural networks. To apply their method, the operations of the numerical solver used to approximate the solution of the ATN model (Equations 1 and 2) can be regarded as layers in an artificial neural network (Chen et al., 2018).

In VI, the posterior distribution of unknown parameters are approximated by a variational distribution. In addition to the variational distribution, we needed to define a prior distribution for the parameters and the loss function to be used in optimization. Otherwise, we followed the algorithm by Blundell et al. (2015) by including the numerical solution of the ATN model as one step in the procedure. Similar to Blundell et al. (2015), we carried out the optimization for unrestricted parameter values but transformed them to admissible ranges as another step in the algorithm.

Specifically, we defined the variational posterior distribution, $\mathcal{V}$, of the unknown parameters as a diagonal Gaussian distribution with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$. Using a factorized distribution in VI is known as mean-field approximation (Ghosh et al., 2021). To keep the standard deviation strictly positive, we parameterized it pointwise as $\sigma = \log(1 + \exp(\rho))$ (Blundell et al., 2015).

Given parameters $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$, we constructed our neural network model following Blundell et al. (2015):

1. Sample a vector $\boldsymbol{p}$ from $\mathcal{V}(\boldsymbol{\mu}, \boldsymbol{\rho})$.
2. Perform transformations to $\boldsymbol{p}$ to obtain parameter values $\boldsymbol{p}_{\text{trans}}$ that are in admissible intervals.
3. Try to obtain a solution from the ODE solver by using the transformed parameters $\boldsymbol{p}_{\text{trans}}$.
4. Evaluate a loss function $\mathcal{L}$ with the output from previous step.

To keep the estimated parameters of the ATN model in a predefined range between $\boldsymbol{p}_{\text{min}}$ and $\boldsymbol{p}_{\text{max}}$, we used the following modified sigmoid (logistic) activation function pointwise in step 2 above:

$$p_{\text{trans}} = p_{\text{min}} + \frac{p_{\text{max}} - p_{\text{min}}}{1 + \exp(-p)}. \tag{8}$$

Moreover, we defined the loss function in step 4 as (Blundell et al., 2015)

$$\mathcal{L}^B(\boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\rho}) = \log V(\boldsymbol{p} | \boldsymbol{\mu}, \boldsymbol{\rho}) - \log P(\boldsymbol{p}) - \log L(\mathcal{D} | \boldsymbol{p}), \tag{9}$$

where V denotes the variational posterior density, P denotes the joint prior density of the unknown model parameters and L is the likelihood of the data. Similar to the variational posterior, we set the prior to a diagonal Gaussian distribution. Since we simulated the data with absolute normal errors (7), the likelihood of the data reads as

$$L(\mathcal{D} | \boldsymbol{p}) = \prod_i \prod_t N(\mathcal{D}_{i,t} | B_i(t), s_i), \tag{10}$$

where $N(z | x, y)$ denotes the univariate Gaussian density with mean x and standard deviation y at z, and the biomass B_i is obtained from $\boldsymbol{p}$.

We resticted the search of parameter values to persistent trophic networks with stable well-behaving dynamics. Therefore, if in step 3,

- any of the guilds went extinct during the simulation period,
- there was too much oscillation (CV > 0.001) or
- any problems occurred in solving the system (1)–(2) numerically,

we set the value of the loss function to infinity in step 4.

We set the prior means of the estimated parameters to values (approximately) at the middle of the intervals used in simulation, or to values calculated from the simulated time series during the time steps 1970–2000 (Table S5). The actual prior means we, however, obtained by performing the inverse of the transformation (8). For all estimated parameters, we set the standard deviations of their priors to one, which gives relatively flat priors for these parameters (Section S2.7). Moreover, we restricted the estimates of half-saturation constants and FR exponents to the intervals used in simulation (Table S5). We also bounded the standard deviations of biomasses to reasonable ranges (Section S2.9).

2.2.2 | Ordinary least squares

To compare the performance of the variational Bayesian method (Section 2.2.1) to OLS approximation, we also estimated the functional response parameters using

$$\mathcal{L}^{\text{OLS}} = \sum_i \sum_t \left(B_i(t) - \mathcal{D}_{i,t} \right)^2, \tag{11}$$

as a loss function. In this, we did not carry out transformations for the parameters but instead, restricted the search space directly so that the parameters were in the same admissible ranges as in VI. Moreover, we set the loss function to infinity in case of unwanted ATN properties or behaviour, similarly as in VI. In OLS, there were fewer parameters to be estimated than in VI (Table 1).

2.2.3 | Training

To avoid costly computation of the gradient with respect to the parameters to be estimated, we utilized an evolutionary optimization

method to optimize the loss functions (9) and (11). Specifically, we applied the Covariance Matrix Adaptation Evolution Strategy (CMA-ES; Hansen, 2016), implemented in the Julia package *Evolutionary.jl* (Wilde et al., 2021). CMA-ES is a stochastic derivative-free numerical method for optimization of challenging functions that can be, for example, non-convex, ill-conditioned, multi-modal and noisy. CMA-ES uses a multivariate normal distribution as the search distribution (Hansen, 2016).

We first explored the impact of different optimization settings on the results with the smaller networks (TN1–TN3). In all cases, we set the initial values of the estimated parameters to correspond the prior distribution of the model parameters (Table S5; Section 2.2.1). Moreover, we set the maximum number of iterations to 3000 in each call of the optimization function and restarted the search several times from the parameter values obtained in the previous call. Such a procedure of restarts is usual practice in evolutionary optimization. Here, we regarded those parameter values that minimized the loss function, when recorded at the end of each call, as the best estimates for the ATN parameters. For these smaller networks, we run CMA-ES with 50 restarts and the number of offspring set to 100, otherwise using the default parameter values in *Evolutionary.jl*. Within these restarts, the actual number of iterations carried out until having reached the minimum loss varied considerably case by case (Figure S9). Overall, the loss functions did not seemingly decrease at the end of iteration (Figures S7 and S8). In addition to this setting, we carried out sensitivity analysis using 50 and 200 offspring with 50 and 100 total number of calls to the optimization function, respectively. Large population sizes can help to avoid local optima, however, we did not find considerable differences in the results consistently when altering the number of offspring (Section 3; Sections S3 and S4). Based on these findings, we used 100 offspring for the larger network (TN4) and evaluated the results after 50 and 25 restarts in OLS and VI, respectively. The lower number of iterations for VI was chosen because of its higher computational cost compared to OLS (due to the higher number of parameters) and based on monitoring the values of the loss function during iteration.

As evolutionary optimization methods use a set of candidate values when searching the minimum of the loss function, it is possible to consider uncertainty of the parameter estimates and predictions by inspecting the whole population of solutions. Nonetheless, in many cases, the resulting set of predictions was highly dispersed (Section S2.3). As such, and for simplicity, we only considered the best candidate of the population of solutions in our parameter estimates and predictions.

When carrying out the model training on a regular laptop used simultaneously for other tasks, the computing time in VI was within days for the smaller networks (TN1–TN3, Figure S10) when using 100 offspring and 50 iterations, less for OLS because of the lower number of parameters. The number of offspring and iterations had a considerable impact on the execution time (Figure S10). For the larger network (TN4), the computation time was within weeks.

3 | RESULTS

3.1 | Parameter estimates

Overall, of the different types of parameters we estimated for the networks with 10 guilds (Table 1), we obtained the most accurate estimates for the half-saturation constants. When comparing the variational posterior means (Section 2.2.1) to the initial values (Table S5), the estimates improved, on average, for all data sets when using VI (Figure 1; Table 2). In OLS estimation, the error of estimates also decreased for all data sets, except for one, for which the mean absolute error stayed approximately the same. When considering relative errors, on average, the improved accuracy of the estimates was more pronounced (Figure 1; Table 2). However, in VI, the estimated standard deviations of the variational posteriors were often small for half-saturation constants (Figure S39) and, in many cases, the likelihood of the true value could be low (Figure S28). Nonetheless, uncertainty in the variational posteriors increased when variation in the data increased, on average (Figure S39.A), as expected.

For the larger network with 25 guilds (Table 1), the improved accuracy in the estimates of half-saturation constants became visible only when considering the relative errors (Table 2). The absolute errors of the estimates were, on average, slightly larger than of the initial values, for both VI and OLS. While the absolute errors were, on average, higher for VI than for OLS, the relative errors were remarkably smaller for VI. These differences may stem from the different scaling of the search space for these methods (Sections 2.2.1 and 2.2.2). Furthermore, the estimated standard deviations of the variational posteriors of half-saturation constants were higher for this larger network than for the smaller ones (Figure S39) and the likelihood of the true value was, on average, higher (Figure S28).

In most cases we tested with the smaller network size (Table 1), we were not able to obtain accurate estimates for the FR exponents and standard deviations of the biomasses, as the error of the estimate did not remarkably decrease, or could actually increase, compared to the error of the initial value (Figures S37 and S38). Nonetheless, the estimated standard deviations of the variational posteriors were larger for the FR exponents than for the half-saturation constants, reflecting the uncertainty about these parameters (Figures S39 and S28).

3.2 | Predictions about the dynamics

On average, the solution of our inverse problem yielded remarkably more accurate predictions about the underlying ATN dynamics than the initially set parameter values, particularly in terms of relative errors (Figure 2; Table 2; Figure S25). In OLS, the mean relative errors were high for some data sets (Figure S25), and here the largest errors appeared for low abundances (Figure S26).

Moreover, VI provided fairly accurate predictions about the total abundance (Figure S27; Table S6). Considering the underlying ATN model, the model predicted its dynamics less accurately. Naturally,

FIGURE 1 Absolute and relative error of the initial values (a, b) and of the estimates of half-saturation constants obtained by variational inference (VI; c, d) and ordinary least squares (OLS; e, f), averaged over the number of feeding links and for the synthetic trophic networks with ten guilds (TN1–TN3; Table 1). The estimates correspond to the minimum value of the loss function during optimization. For VI, the estimates correspond to the variational posterior means. The error is calculated as a difference to the parameter value used in simulation of the synthetic time series. The relative error compares this difference to the true parameter value.
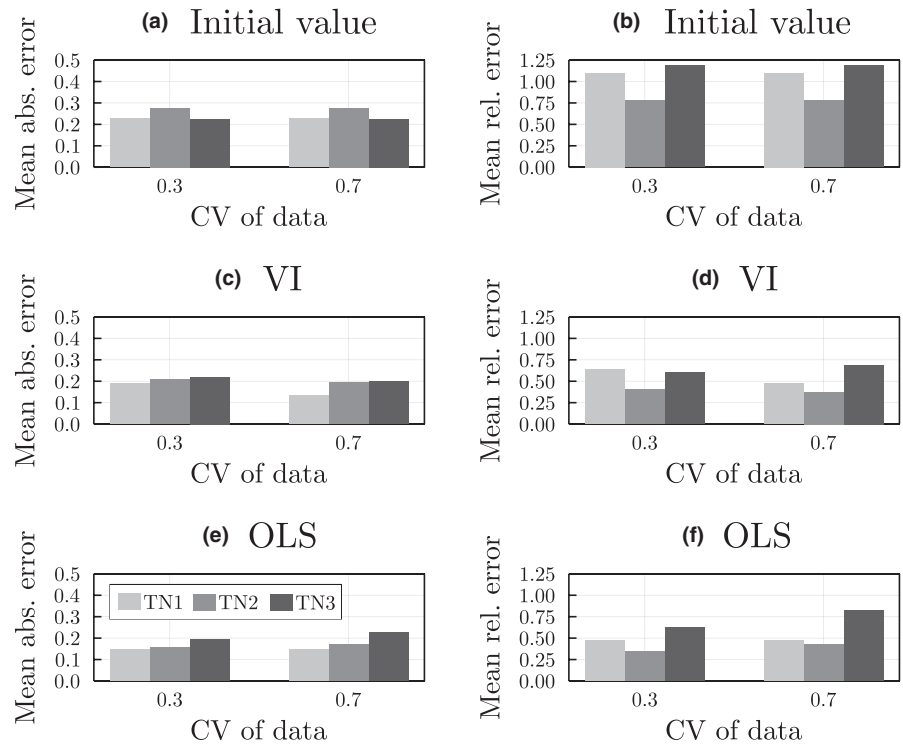
TABLE 2 Mean absolute and relative error of the initial values, the estimates of half-saturation constants ($B0$) and the predicted biomasses in the training and test sets, obtained by variational inference (VI) and ordinary least squares (OLS) for the synthetic trophic networks (TN1–TN4; Table 1).

Network	Error type	Initial	VI	OLS
TN1–TN3	Absolute for $B0$	0.23–0.28	0.13–0.22	0.14–0.23
	Relative for $B0$	0.78–1.19	0.37–0.69	0.35–0.83
TN4	Absolute for $B0$	0.23	0.27	0.24
	Relative for $B0$	0.91	0.65	0.90
TN1–TN3	Absolute for pred. in training set	0.073–0.096	0.0037–0.015	0.0017–0.020
	Relative for pred. in training set	1.30–14.46	0.053–0.14	0.043–4.8
TN4	Absolute for pred. in training set	0.18	0.06	0.02
	Relative for pred. in training set	62.02	0.26	61.89
TN1–TN3	Absolute for pred. in test set	0.074–0.093	0.0068–0.017	0.0060–0.018
	Relative for pred. in test set	1.35–13.13	0.13–0.26	0.12–7.1
TN4	Absolute for pred. in test set	0.18	0.06	0.02
	Relative for pred. in test set	71.07	0.27	94.39

Note: The errors were calculated as in Figures 1 and 2.

the guild-specific predictions varied in accuracy (Figures 3 and 4; Section S2.2).

4 | DISCUSSION

The present study lays the groundwork for Bayesian parameter estimation of the ATN model by developing a variational inference method and exploring how accurately the half-saturation constants of functional responses, particularly, can be inferred from observed data. For this, we used synthetic time series for which we know exactly the parameter values that produced the data. Compared to OLS, VI yielded estimates of similar accuracy. This accuracy varied

across networks, even for networks of similar size and connectance (mean relative error 37%–69% for VI; Table 2). Despite relatively high parameter estimation errors, the estimates overall yielded reasonably accurate predictions for both the data used for parameter estimation and for fresh data generated by introducing a disturbance to the ATN dynamics (mean relative errors ≤26% and ≤27%, respectively, for VI).

The results of the present study align with prior research on the invertibility of the ATN model. Banks et al. (2016, 2017) reported that the solution to their inverse problem on synthetic data did not consistently converge to the same parameter values as those used to generate the data. This observation was also evident in our study, reinforcing the idea that the inverse problem of the ATN model may
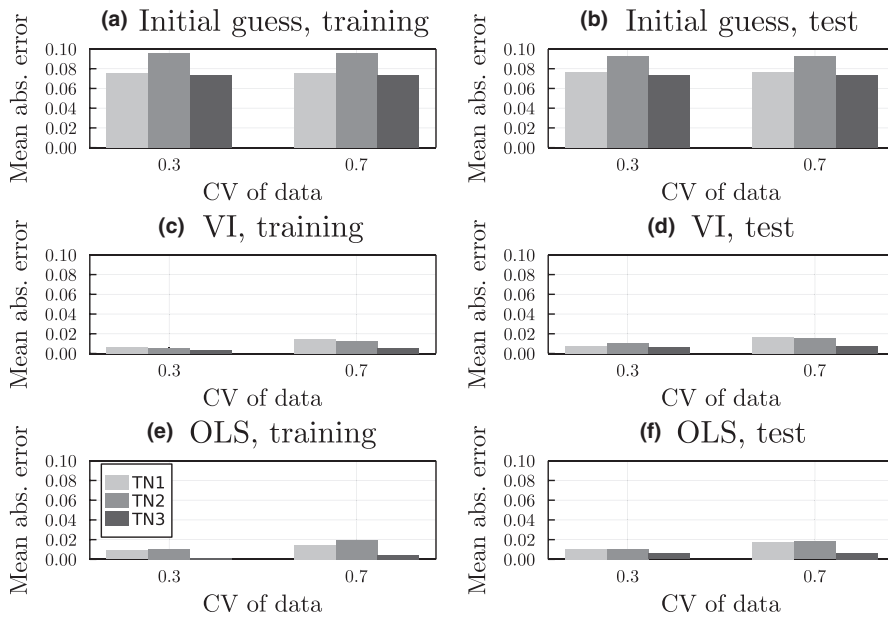
FIGURE 2 The mean absolute error of predicted allometric trophic network dynamics from the originally simulated dynamics within training and test data. The initial predictions (a, b) correspond to the values initially set for the functional response parameters in optimization. For predictions by variational inference (VI; c, d) and ordinary least squares (OLS; e, f), we used the values of functional response parameters corresponding to the minimum value of the loss function.

involve multiple local optima, making it challenging to identify the global optimum in the optimization landscape. This reflects the model's inherent complexity. Consequently, our results underscore the importance of testing various initial values and priors when calibrating the ATN model to observations in future studies. Rather than relying on a single solution, it may be prudent to consider a set of solutions that correspond to different scenarios, particularly when predicting to the future. Additionally, a multimodal variational distribution could better represent the posterior distribution of parameters than the unimodal approach we employed. Nonetheless, our study is the first to suggest a Bayesian method for the ATN model. In addition, while Banks et al. (2016, 2017) briefly addressed the invertibility of the ATN model, this issue has not been given full attention. We tested the proposed method and explored the model's invertibility in a simplified setting, laying a foundation for further exploration and refinements in future research.

We demonstrated our method with synthetic trophic networks of relatively moderate size, but our inverse problems are comparable to empirical studies of ATN modelling (Banks et al., 2016, 2017; Boit et al., 2012; Koen-Alonso & Yodzis, 2005). The smaller networks with 10 guilds had a similar number of unknown parameters (Table 1) as Koen-Alonso and Yodzis (2005) four-species model (38 estimated parameters), and more than (Banks et al., 2016, 2017) (six parameters). Our larger network with 25 guilds had more estimated parameters than Boit et al. (2012) reported to have manually calibrated for their 24-guild model (107 unknown functional response parameters). For these network sizes, we obtained results using a regular laptop, though the computing time was long, especially for the larger network. For networks larger than this or those with more unknown parameters, computing services tailored for high-demand tasks may alleviate the computational burden.

The present study examined the invertibility of the ATN model in a theoretical framework with a simplified approach. Specifically, we considered a relatively simple formulation of functional responses

and, especially for the larger network, limited the unknown model parameters to half-saturation constants. Such a setting nonetheless resembles a previous validation study of the ATN model against empirical data (Boit et al., 2012). In parameter inference, we constrained parameter values to predefined ranges, and the synthetic data we generated for model training had relatively time-invariant underlying ATN dynamics, which allowed us to further restrict the search space during parameter estimation and decrease ill-posedness. Identifiability is a common challenge in inverse problems and additional constraints can help regularize inversion results to consistent ranges (Wu et al., 2019). Our constraint aligns with findings that oscillations are seldom present in natural population dynamics (Lande et al., 2003), and by focusing on stable dynamics, we also reduced the number of latent variables, making it unnecessary to estimate the initial biomasses. While our modelling choices need to be addressed in an empirical context, the simplified framework serves to identify possibilities and challenges in ATN parameter estimation.

In this paper, we focused on a Bayesian approach to parameter inference, which allows for incorporation of prior knowledge and quantification of estimation uncertainty. For our inverse problem, we nonetheless only tested uninformative priors. Regarding the variational posteriors, the method often underestimated uncertainty although for functional responses, the estimates behaved as expected. This underestimation of variation is a recognized issue in mean field VI (Ghosh et al., 2021). When applying the method to empirical data, one potential solution is to account for parameter correlations within the variational posterior (Ghosh et al., 2021), although this increases the number of parameters to be estimated.

To invert the ATN model, we applied the Bayes by Backpropagation method suggested by Blundell et al. (2015) for variational Bayesian inference of artificial neural networks but used evolutionary optimization for training. We chose evolutionary optimization instead of a gradient-based method because computing the gradient with respect to the parameters is often costly for ODE models (Chen et al., 2018;
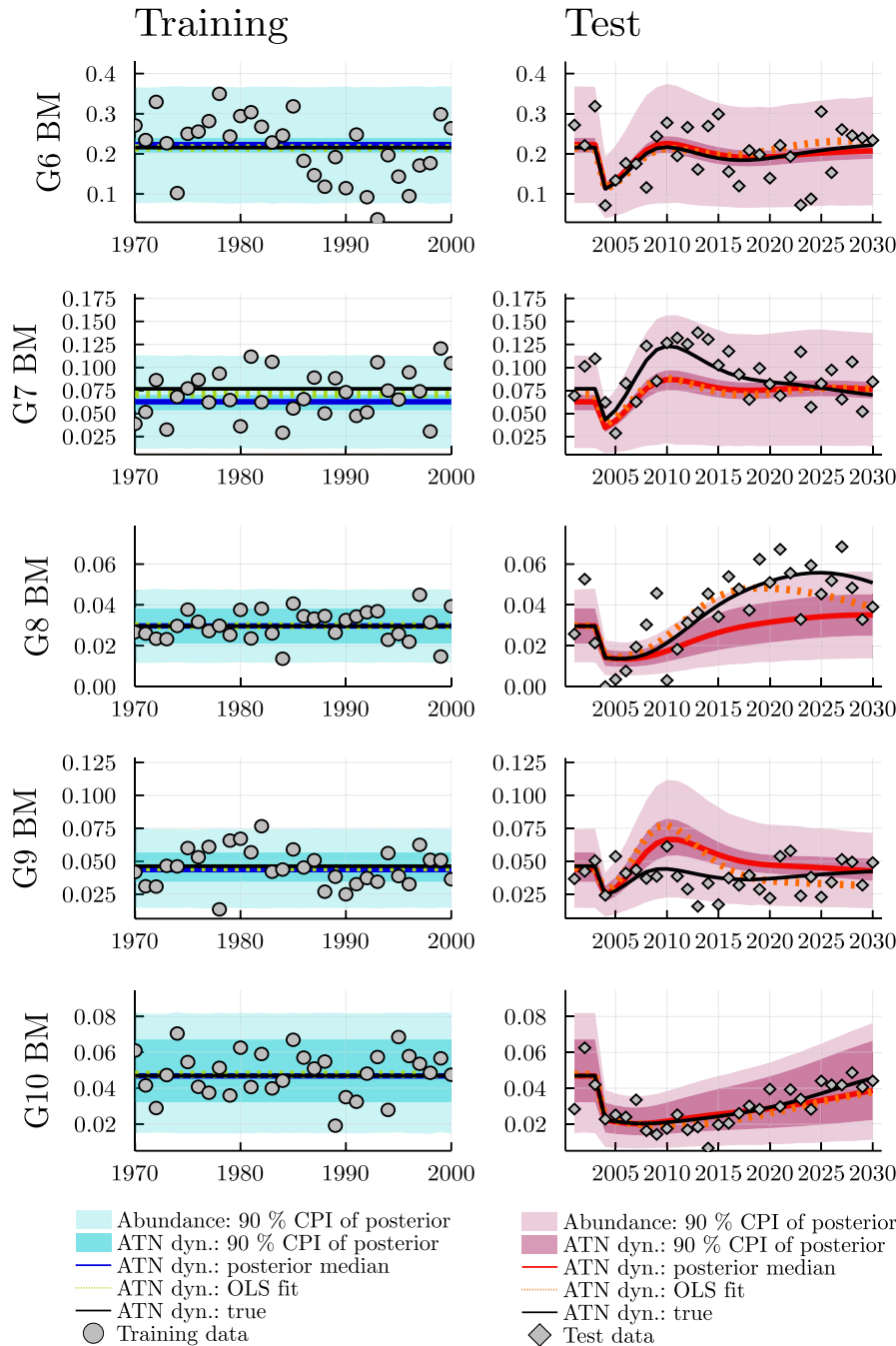
FIGURE 3 Guild-specific predicted total abundances and underlying allometric trophic network (ATN) dynamics for the training (figures on the left) and test (figures on the right) biomass (BM) time series that were simulated using the synthetic trophic network 3 and the coefficient of variation 0.3. For this data set, the mean error in predicted ATN dynamics was lowest among the generated data sets (Figure 2). The figure shows the true simulated biomasses and the predictions by variational inference and ordinary least squares (OLS) fit for five of the 10 guilds (G1–G5). For the Bayesian predictions, we used 3×10^4 samples from the posterior distributions of model parameters (Section S2.7). CPI, central probability interval.
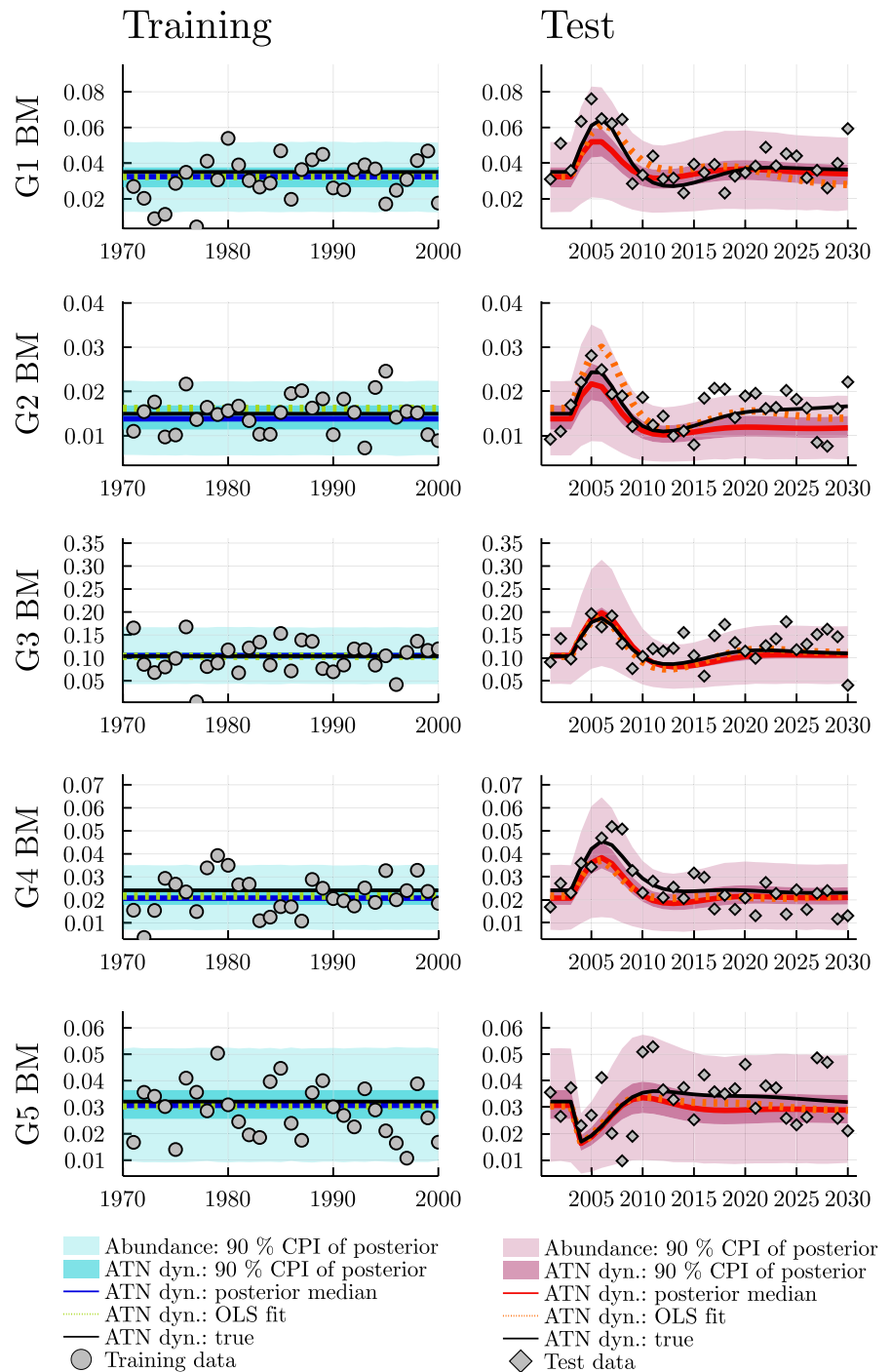
Ghosh et al., 2021). For example, algebraic differentiation (AD) is commonly applied in neural network training but can be computationally expensive when used through ODE solver (Chen et al., 2018; Ghosh et al., 2021). This also appeared in our initial trials for the ATN model. Evolutionary optimization methods, being black-box approaches, provided flexibility in implementing the ATN model and optimizing the computer program for speed. Nonetheless, efficient AD techniques for ODE models have been recently suggested (Chen et al., 2018; Ghosh et al., 2021). The efficiency (Chen et al., 2018) comes from backpropagating through an ODE solver without access to its internal operations. Overall, the field of machine learning

advances rapidly (Sun et al., 2019) and it may be worth considering the latest development when applying the proposed approach.

Finally, while the variational inference approach we propose may appear considerably more complex than the OLS analysis conducted in parallel, it offers a crucial advantage: the ability to quantify uncertainty in parameter estimates. OLS, as it is, provides only point estimates without addressing this uncertainty. Previous studies have incorporated sensitivity analysis or explored the likelihood around parameter estimates (Banks et al., 2016, 2017; Koen-Alonso & Yodzis, 2005) to assess parameter uncertainty, but these methods do not directly estimate the full uncertainty in the way a Bayesian

FIGURE 4 Guild-specific predicted total abundances and underlying allometric trophic network (ATN) dynamics for the training (figures on the left) and test (figures on the right) biomass (BM) time series that were simulated using the synthetic trophic network 3 and the coefficient of variation 0.3. The figure shows the true simulated biomasses and the predictions by variational inference and ordinary least squares (OLS) fit for five of the 10 guilds (G6–G10). CPI, central probability interval.

approach can. For a complex model like the ATN, assessing uncertainty is inherently challenging, and we expect that doing so within a frequentist framework would also involve significant complexity—a task beyond the scope of this paper.

Research on the inverse problem of the ATN model is limited (Banks et al., 2016, 2017; Boit et al., 2012; Koen-Alonso & Yodzis, 2005), particularly in the Bayesian context. The present study builds upon previous studies of ATN validation against empirical data (Banks et al., 2016, 2017; Boit et al., 2012; Koen-Alonso & Yodzis, 2005) and research in machine learning (Blundell et al., 2015; Hansen, 2016) to develop a variational Bayesian method for parameter inference of the ATN model from observed data. With synthetic

data, we demonstrate the challenges in ATN parameter estimation and identify possibilities to alleviate these issues. While developed particularly for the ATN model, our method is a potential tool for parameter inference also for other ODE models.

AUTHOR CONTRIBUTIONS

Maria Tirronen and Anna Kuparinen conceived the research ideas; Maria Tirronen designed the methodology, implemented the software, carried out the analysis, led the writing of the original manuscript and conducted all subsequent revisions; Anna Kuparinen took part in the writing of the original manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14447.

DATA AVAILABILITY STATEMENT

The computer codes for data simulation, parameter estimation, analyses and visualization are available at https://doi.org/10.5281/zenodo.13957315 (Tirronen, 2024).

ORCID

Maria Tirronen https://orcid.org/0000-0001-6052-3186
Anna Kuparinen https://orcid.org/0000-0002-7807-8946

REFERENCES

Ávila Thieme, M., Corcoran, D., Perez-Matus, A., Wieters, E., Navarrete, S., Marquet, P., & Valdovinos, F. (2021). Alteration of coastal productivity and artisanal fisheries interact to affect a marine food web. *Scientific Reports, 11*, 1765. https://doi.org/10.1038/s41598-021-81392-4

Banks, H., Banks, J., Bommarco, R., Curtsdotter, A., Jonsson, T., & Laubmeier, A. (2016). *Parameter estimation for an allometric food web model*. Technical report. Center for Research in Scientific Computation technical report CRSC-TR16-03.

Banks, H., Banks, J., Bommarco, R., Curtsdotter, A., Jonsson, T., & Laubmeier, A. (2017). Parameter estimation for an allometric food web model. *International Journal of Pure and Applied Mathematics, 114*, 143–160. https://doi.org/10.12732/ijpam.v114i1.12

Barros, C., Luo, Y., Chubaty, A. M., Eddy, I. M. S., Micheletti, T., Boisvenue, C., Andison, D. W., Cumming, S. G., & McIntire, E. J. B. (2023). Empowering ecological modellers with a PERFICT workflow: Seamlessly linking data, parameterisation, prediction, validation and visualisation. *Methods in Ecology and Evolution, 14*(1), 173–188. https://doi.org/10.1111/2041-210X.14034

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association, 112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural networks. *arXiv*. https://doi.org/10.48550/ARXIV.1505.05424

Bodner, K., Rauen Firkowski, C., Bennett, J. R., Brookson, C., Dietze, M., Green, S., Hughes, J., Kerr, J., Kunegel-Lion, M., Leroux, S. J., McIntire, E., Molnár, P. K., Simpkins, C., Tekwa, E., Watts, A., & Fortin, M.-J. (2021). Bridging the divide between ecological forecasts and environmental decision making. *Ecosphere, 12*(12), e03869. https://doi.org/10.1002/ecs2.3869

Boit, A., Martinez, N. D., Williams, R. J., & Gaedke, U. (2012). Mechanistic theory and modelling of complex food-web dynamics in Lake Constance. *Ecology Letters, 15*(6), 594–602. https://doi.org/10.1111/j.1461-0248.2012.01777.x

Brose, U., Williams, R. J., & Martinez, N. D. (2006). Allometric scaling enhances stability in complex food webs. *Ecology Letters, 9*(11), 1228–1236. https://doi.org/10.1111/j.1461-0248.2006.00978.x

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31, pp. 6571–6583). Curran Associates, Inc.

Delmas, E., Brose, U., Gravel, D., Stouffer, D. B., & Poisot, T. (2017). Simulations of biomass dynamics in community food webs. *Methods in Ecology and Evolution, 8*, 881–886. https://doi.org/10.1111/2041-210X.12713

Dietze, M. (2017). *Ecological forecasting*. Princeton University Press. ISBN: 9781400885459.

Eloranta, A. P., Perälä, T., & Kuparinen, A. (2023). Effects of temporal abiotic drivers on the dynamics of an allometric trophic network model. *Ecology and Evolution, 13*(3), e9928. https://doi.org/10.1002/ece3.9928

Ghosh, S., Birrell, P., & De Angelis, D. (2021). Variational inference for nonlinear ordinary differential equations. In A. Banerjee & K. Fukumizu (Eds.), *The 24th international conference on artificial intelligence and statistics, AISTATS 2021, April 13–15, 2021, Virtual Event, volume 130 of Proceedings of Machine Learning Research* (pp. 2719–2727). PMLR. http://proceedings.mlr.press/v130/ghosh21b.html

Hansen, N. (2016). The CMA evolution strategy: A tutorial. *arXiv*. https://doi.org/10.48550/arXiv.1604.00772

Kabanikhin, S. I. (2008). Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems, 16*(4), 317–357. https://doi.org/10.1515/JIIP.2008.019

Kath, N., Boit, A., Guill, C., & Gaedke, U. (2018). Accounting for activity respiration results in realistic trophic transfer efficiencies in allometric trophic network (ATN) models. *Theoretical Ecology, 11*, 453–463. https://doi.org/10.1007/s12080-018-0378-z

Koen-Alonso, M., & Yodzis, P. (2005). Multispecies modelling of some components of the marine community of northern and central Patagonia, Argentina. *Canadian Journal of Fisheries and Aquatic Sciences, 62*(7), 1490–1512. https://doi.org/10.1139/f05-087

Kuparinen, A., Uusi-Heikkilä, S., Perälä, T., Ercoli, F., Eloranta, A. P., Cremona, F., Nõges, P., Laas, A., & Nõges, T. (2023). Generalist invasion in a complex lake food web. *Conservation Science and Practice, 5*(6), e12931. https://doi.org/10.1111/csp2.12931

Lande, R., Engen, S., & Saether, B.-E. (2003). *Stochastic population dynamics in ecology and conservation*. Oxford Unversity Press. https://doi.org/10.1093/acprof:oso/9780198525257.001.0001. ISBN: 0198525257.

Lotka, A. J. (1925). *Elements of physical biology*. Williams and Wilkins Company.

Martinez, N. D. (1992). Constant connectance in community food webs. *The American Naturalist, 139*(6), 1208–1218. https://doi.org/10.1086/285382

Martinez, N. D. (2020). Allometric trophic networks from individuals to socio-ecosystems: Consumer–resource theory of the ecological elephant in the room. *Frontiers in Ecology and Evolution, 8*, 92. https://doi.org/10.3389/fevo.2020.00092

Perälä, T., Pesari, S. N. K., Ahti, P., Lehtinen, S. O., Schoenrock, K., & Kuparinen, A. (2023). Non-trophic interactions amplify kelp harvest-induced biomass oscillations and biomass changes in a kelp forest ecological network model. *Marine Ecology Progress Series, 722*, 11–36. https://doi.org/10.3354/meps14438

Poisot, T., Delmas, E., Shah, V. B., Kelman, T., & Clegg, T. (2017). *PoisotLab/BioEnergeticFoodWebs.jl*. https://doi.org/10.5281/zenodo.401053

Rackauckas, C., & Nie, Q. (2017). DifferentialEquations.jl—A performant and feature-rich ecosystem for solving differential equations in Julia. *Journal of Open Research Software*, *5*(1), 15. https://openresearchsoftware.metajnl.com/articles/10.5334/jors.151

Sloss, A. N., & Gustafson, S. (2020). 2019 evolutionary algorithms review. In W. Banzhaf, E. Goodman, L. Sheneman, L. Trujillo, & B. Worzel (Eds.), *Genetic programming theory and practice XVII. Genetic and evolutionary computation* (pp. 307–344). Springer International Publishing. https://doi.org/10.1007/978-3-030-39958-0_16. ISBN: 9783030399580.

Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2019). A survey of optimization methods from a machine learning perspective. *arXiv*. https://doi.org/10.48550/arXiv.1906.06821

Tirronen, M. (2024). Variational Bayesian inference for allometric trophic network models. *Zenodo*, https://doi.org/10.5281/zenodo.13957315

Uusi-Heikkilä, S., Perälä, T., & Kuparinen, A. (2022). Fishing triggers trophic cascade in terms of variation, not abundance, in an allometric trophic network model. *Canadian Journal of Fisheries and Aquatic Sciences*, *79*(6), 947–957. https://doi.org/10.1139/cjfas-2021-0146

Valderrama-Bahamóndez, G., & Fröhlich, H. (2019). MCMC techniques for parameter estimation of ODE based models in systems biology. *Frontiers in Applied Mathematics and Statistics*, *5*, 55. https://doi.org/10.3389/fams.2019.00055

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature*, *118*(2972), 558–560. https://doi.org/10.1038/118558a0

Wilde, A., Thiem, D., Leo, A., Gupta, A., Haselgrove, D. M., Honeypot95, Matago, Rogerluo, & Churavy, V. (2021). Evolutionary.jl (v0.10.0). Zenodo. https://doi.org/10.5281/zenodo.5110647

Williams, R. (2008). Effects of network and dynamical model structure on species persistence in large model food webs. *Theoretical Ecology*, *1*, 141–151. https://doi.org/10.1007/s12080-008-0013-5

Williams, R., Brose, U., & Martinez, N. (2006). Homage to Yodzis and Innes 1992: Scaling up feeding-based population dynamics to complex ecological networks. In N. Rooney, K. S. McCann, & D. L. G. Noakes (Eds.), *From energetics to ecosystems: The dynamics and structure of ecological systems* (pp. 37–51). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5337-5_2. ISBN: 9781402053368.

Williams, R., & Martinez, N. (2000). Simple rules yield complex food webs. *Nature*, *404*, 180–183. https://doi.org/10.1038/35004572

Williams, R. J., & Martinez, N. D. (2004). Stabilization of chaotic and non-permanent food-web dynamics. *The European Physical Journal B*, *38*, 297–303. https://doi.org/10.1140/epjb/e2004-00122-1

Wu, J., Wang, J.-X., & Shadden, S. C. (2019). Adding constraints to Bayesian inverse problems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 1666–1673. https://doi.org/10.1609/aaai.v33i01.33011666

Yodzis, P., & Innes, S. (1992). Body size and consumer-resource dynamics. *The American Naturalist*, *139*(6), 1151–1175. https://doi.org/10.1086/285380

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1: Synthetic trophic networks.

Appendix S2: Results of model fitting.

Appendix S3: Sensitivity analysis, 50 offspring, 50 iterations.

Appendix S4: Sensitivity analysis, 200 offspring, 100 iterations.