Published in final edited form as:

Math Biosci. 2017 December ; 294: 71–84. doi:10.1016/j.mbs.2017.09.009.

Identification of microbiota dynamics using robust parameter estimation methods

Matthias Chung^{a,b}, Justin Krueger^a, and Mihai Pop^c

^aVirginia Tech, Department of Mathematics, 225 Stanger St, Blacksburg, Virginia ^bVirginia Tech, Computational Modeling and Data Analytics, Academy of Integrated Science, 8314 Paint Branch Dr, College Park, Maryland ^cUniversity of Maryland, Center for Bioinformatics and Computational Biology, 8314 Paint Branch Dr, College Park, Maryland

Abstract

The compositions of in-host microbial communities (microbiota) play a significant role in host health, and a better understanding of the microbiota's role in a host's transition from health to disease or vice versa could lead to novel medical treatments. One of the first steps toward this understanding is modeling interaction dynamics of the microbiota, which can be exceedingly challenging given the complexity of the dynamics and difficulties in collecting sufficient data. Methods such as principal differential analysis, dynamic flux estimation, and others have been developed to overcome these challenges. Despite their advantages, these methods are still vastly underutilized in fields such as mathematical biology, and one potential reason for this is their sophisticated implementation. While this paper focuses on applying principal differential analysis to microbiota data, we also provide comprehensive details regarding the derivation and numerics of this method and include a functional implementation for readers' benefit. For further validation of these methods, we demonstrate the feasibility of principal differential analysis using simulation studies and then apply the method to intestinal and vaginal microbiota data. In working with these data, we capture experimentally confirmed dynamics while also revealing potential new insights into the system dynamics.

Keywords

Parameter estimation; Differential equations; Lotka-Volterra models; Microbiota

Correspondence to: Justin Krueger.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

1.1. Biology of Microbial Communities

Bacteria are ubiquitous in our world, and play a key role in maintaining the health of our environment as well as the health of virtually all living organisms. The human associated microbial communities (human microbiota) have been shown to be at least associated with, if not causative of, several human diseases such as periodontitis [1], type 2 diabetes [2], atopic dermatitis [3], ulcerative colitis [4], Crohn's disease [5], and vaginosis [6]. Furthermore, time series data have shown that the host-associated microbiota undergo dynamic changes over time within a same individual, e.g., within the gut of a developing infant [7, 8], the gut, mouth, and skin of healthy adults [9], and within the vagina of reproductive age women [10]. The mechanisms that underlie these changes are currently not well understood, whether they represent fluctuations in the normal flora, or the transition from health to disease, and conversely from disease to health after treatment.

Understanding the role human-associated microbial communities play in health and disease requires the elucidation of the complex networks of interactions between the microbes and between microbes and the host, a challenging task due to our inability to directly observe bacterial interactions. Instead, researchers have reconstructed microbial networks based on indirect approaches, such as knowledge about the metabolic functions encoded in the genomes of the interacting partners [11], coexistence patterns across multiple samples [12], covariance of abundance across samples [13], or changes in abundance across time [14, 15]. Multiple mathematical formalisms have been used to reason about the resulting networks with examples including metabolic modeling through flux balance analysis [16], machine learning algorithms based on environmental parameters [17], and differential equation based models of interactions [15].

Here we focus on the latter, a flexible formalism that can model complex interaction patterns, including abundance-dependent interaction parameters [18]. While such modeling approaches have been developed since the 1980s in the context of wastewater treatment systems [19], their use in studying human-associated microbial communities has been limited, in no small part due to the specific characteristics of human microbiome data. First, the rate at which samples can be collected is severely limited by clinical and logistical factors, e.g., stool samples can be collected roughly on a daily basis, while subgingival plaque may only be feasibly collected at an interval of several months. Second, microbiome data are sparse, i.e., most organisms are undetected in most samples [20] due to the detection limits of sequencing-based assays as well as the high variability of the microbiota across the human population. Third, it is difficult if not impossible to directly measure environmental parameters, such as nutrient concentrations, that may impact the microbiota.

These features of the data derived from the human-associated microbiota lead to an ill-posed parameter estimation problem since multiple parameter sets may be consistent with the data. Numerical instabilities that result from specific parameter sets can also cause traditionally used estimation procedures to fail. Here we explore solutions to the parameter estimation problems in the context of the Lotka-Volterra formalism, which is described in more detail below.

1.2. Mathematical Modeling of Microbial Communities

We focus on a special type of differential equation based models of interactions, the Lotka-Volterra model, which is named after Alfred J. Lotka (1880–1949), an American mathematician, physical chemist, and statistician, and Vito Volterra (1860–1940), an Italian mathematician and physicist [21]. This model was originally developed in the context of predator-prey interactions; however, it can be generalized to more complex interactions. Let $\mathbf{y}$ be the time dependent state variable for the dynamics of n species with time variable t . Then the Lotka-Volterra system can be written as

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}) = \text{diag}(\mathbf{y})(\mathbf{b} + \mathbf{A}\mathbf{y}). \quad (1)$$

Here, $\text{diag}(\mathbf{y})$ is the diagonal matrix with the state variables $\mathbf{y} = [y_1, \dots, y_n]^T$ as its entries, and $\mathbf{b} = [b_1, \dots, b_n]^T \in \mathbb{R}^n$ is the intrinsic growth rate, which incorporates the natural birth and death rate of each species in a given environment. Negative b_i refers to a negative intrinsic growth rate and species i 's survival depends on the interaction with other species. The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ represents the dynamics of the relationships between the species and is often referred to as the interaction matrix. An element $a_{ij} = [\mathbf{A}]_{i,j}$ of $\mathbf{A}$ describes the influence of species j on the growth of species i . For $i = j$ and $a_{ij} < 0$, we classically consider species i to be a prey of predator j and vice versa for $a_{ij} > 0$. If for $i = j$ both $a_{ij} < 0$ and $a_{ji} < 0$, species i and j are competing for existence. On the other hand, if $i = j$ both $a_{ij} > 0$ and $a_{ji} > 0$, symbiotic behavior between species i and j is observed. If $a_{ij} = 0$, $i = j$, no direct interaction between species i and j exists.

This formalism allows the simulation of ecological systems and the study of the long-term behavior of these systems. For example, the equilibrium solution $\mathbf{y}_\infty = \mathbf{0}$ describes the extinction of all species. The equilibrium $\mathbf{y}_\infty = \mathbf{0}$ is unstable if and only if at least one intrinsic growth rate b_i is positive. All other biologically feasible, i.e., nonnegative, equilibrium solutions $\mathbf{y}_\infty$ of (1) are solutions of the equation $\mathbf{0} = \text{diag}(\mathbf{y})(\mathbf{b} + \mathbf{A}\mathbf{y})$. The real parts of the eigenvalues of $\mathbf{f}_y(\mathbf{y}_\infty)$ determine the stability of the additional equilibria. Here, $\mathbf{f}_y(\mathbf{y}_\infty)$ is the Jacobian of $\mathbf{f}$ with respect to $\mathbf{y}$ evaluated at $\mathbf{y}_\infty$ with $\mathbf{f}_y(\mathbf{y}) = \text{diag}(\mathbf{b}) + \text{diag}(\mathbf{y})\mathbf{A} + \text{diag}(\mathbf{A}\mathbf{y})$. Detailed analysis on population dynamics, persistence, and stability can be found in [22], and the specifics for the dynamics of Lotka-Volterra systems can be found in [23].

1.3. Assumptions and Prior Knowledge

The Lotka-Volterra system makes simplifying assumptions about the underlying biological system, in particular it assumes that the interaction between microbes is constant in time and independent of the abundance of the interacting partners. As a result, certain types of microbial interactions, e.g., quorum sensing, cannot be effectively modeled. For the sake of computational tractability, we restrict ourselves to this traditional definition of the Lotka-Volterra model, but extensions that allow more complex interaction modalities [18] can also be addressed by the computational parameter estimation framework described below.

The number of parameters of the Lotka-Volterra model is proportional to the square of the number of interacting partners, complicating the parameter estimation problem for complex datasets, especially when the number of samples is limited. Prior knowledge about the system is often available and can be used to mitigate the complexity of parameter estimation. In the context of host-associated microbiota, this prior knowledge may include:

Known Parameters. Intrinsic growth rates or specific interactions might be known or partially known prior to parameter estimation.

Grouping. A reduction in the size of the system in (1) can be achieved if species with similar behavior can be pooled together into a meta-species.

Biomass. Often a reasonable assumption is the total biomass in a dynamical system remains constant or is tightly regulated at all time.

Symmetry. Knowing the influence of species j on species i may simultaneously give information on both interaction parameters a_{ij} and a_{ji} .

Finite Carrying Capacities. It can be assumed that all species display logistic growth and have a finite carrying capacity in the absence of all other species.

Sparsity. For some biological systems one may assume that most species do not directly interact, constraining the set of solutions possible.

In this work, we use principal differential analysis, a previously established parameter estimation method [24, 25], to recover intrinsic information about the dynamical system given temporal density observations of the interacting species. In our application to microbial communities, this approach readily allows for the inclusion of prior knowledge like the examples listed above, which helps to address the ill-posedness of such problems.

2. Methods

2.1. Problem Statement

In order to validate and impel model predictions, the mathematical model needs to be compared to experimentally observed data $\mathbf{d} \in \mathbb{R}^m$. The estimation of parameters for dynamical systems is a key step in the analysis of biological systems. The point estimates give quantitative information about the system, and in the case of a Lotka-Volterra system specifically, the intrinsic growth rates and interaction dynamics between species. Let us assume intrinsic growth rates $\mathbf{b}$ and interaction dynamics $\mathbf{A}$ are unknown and are collected in the parameter vector $\mathbf{p} = [\mathbf{b}; \text{vec}(\mathbf{A})]$, where $\text{vec}(\mathbf{A})$ is the concatenation of the columns of $\mathbf{A}$, i.e., $\text{vec}(\mathbf{A}) = [a_{11}, \dots, a_{n1}, a_{12}, \dots, a_{nn}]^T$.

The general parameter estimation problem for any explicit first order ordinary differential equation (not just restricted to (1)) is stated as the constrained weighted least squares problem

$$(\hat{\mathbf{p}}, \hat{\mathbf{y}}) = \arg \min_{(\mathbf{p}, \mathbf{y})} \|\mathbf{W}(\mathbf{m}(\mathbf{y}(t; \mathbf{p})) - \mathbf{d})\|_2^2 + \alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{y})) \quad (2)$$

subject to $\mathbf{y}' = \mathbf{f}(t, \mathbf{y}, \mathbf{p})$,

with $\mathbf{y} : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. Additional constraints such as prior knowledge discussed above or relaxed inequality or equality constraints are gathered by the general statement $\alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{y}))$, where $\alpha > 0$ and $\mathcal{D}$ is a distance metric [26]. For every feasible parameter set $\mathbf{p}$ and initial state $\mathbf{y}_0 = \mathbf{y}(a)$ we assume that the conditions of the Theorem of Picard-Lindelöf are fulfilled. This means that a unique state variable $\mathbf{y}$ exists for any feasible choice $\mathbf{p}$ and $\mathbf{y}_0$. For our focus, the Lotka-Volterra system fulfills this condition for any finite $\mathbf{p}$ and $\mathbf{y}_0$. Further, $\mathbf{m} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a projection from the state space onto the measurement space of given data $\mathbf{d} = [d_1, \dots, d_m]^T \in \mathbb{R}^m$. For instance, observations $\mathbf{d}$ might not include all states at all time points, $\mathbf{d}$ might be in a frequency domain, or $\mathbf{d}$ may only be a combination of observed states. The precision matrix $\mathbf{W}^T \mathbf{W}$ is also referred to as the inverse covariance matrix, and the optimization problem (2) can be seen as a weighted least squares problem with weight matrix $\mathbf{W}$. Here, we assume that we have independent samples $\mathbf{d}$ and $\mathbf{W} = \text{diag}(\mathbf{w}) = \text{diag}([w_1, \dots, w_m]^T)$, with $w_j > 0$ for $j = 1, \dots, m$. If w_j is large, observation d_j plays an important role for the parameter estimation procedure. Otherwise, small w_j indicate a lesser role for observation d_j . The underlying statistical assumption for this parameter estimation problem is that the residuals are normally distributed and in the case of a diagonal $\mathbf{W}$ are uncorrelated [27].

Computationally, solving (2) may be challenging. A limited number of observations, high levels of noise in the data, large dynamical systems, non-linearity of the system, and a large number of unknown parameters are all such challenges. These challenges appear for Lotka-Volterra models of biological systems and make parameter estimation extremely difficult [28]. As a result, we desire a robust parameter estimation method, which means the method provides reliable and accurate parameter estimates [29], and under this definition, a lack of robustness may stem from three different sources: corrupted data, the numerical integration scheme, or from the numerical optimization method [30].

Next, we note the methods traditionally used to solve the parameter estimation problem, discuss the limitations of these methods, and present alternatives.

2.2. Single Shooting Methods

Typically, single shooting methods are utilized to solve (2) for biological systems [31, 32]. For single shooting methods, first, initial guesses for $\mathbf{p}^0$ and $\mathbf{y}_0^0$ are used to numerically solve the initial value problem (forward problem) using single- or multi-step methods such as Runge-Kutta and Adams-Bashforth methods [33, 34]. Next, the misfit between data and model is computed, and depending on the optimization strategy (e.g., gradient based strategies such as Gauss-Newton methods or direct search approaches such as the Nelder-Mead Method), a new set $(\mathbf{p}^1, \mathbf{y}_0^1)$ is chosen. This process continues until a $\hat{\mathbf{p}}$ and $\hat{\mathbf{y}}_0$ are found to fulfill pre-defined optimality criteria. Since most efficient optimization methods are typically local optimization methods, globalization is achieved, for example, by a Monte Carlo sampling of the search space, i.e., repeated local optimization with random initial

guesses [35]. The global minimizer chosen from the set of local minimizers is the local minimizer $(\hat{\mathbf{p}}, \hat{\mathbf{y}}_0)$ with minimal function value. Various other strategies for global optimization can be applied, such as simulated annealing [36], evolutionary algorithms [37], or particle swarm optimization methods [38].

It has been established that single shooting methods are not robust to initial guesses $\mathbf{p}^0$ and $\mathbf{y}_0^0$, which in this case refers to the methods' ability to successfully find minimizers as defined in (2). Notorious parameter estimation examples illustrating the lack of robustness can be found in [39, 40, 41], and various alternative methods have been developed to compensate for such lack of robustness.

2.3. Alternative Methods

Here we give a short overview of some of these alternative methods. While this overview is not comprehensive, the vast volume of publications on parameter estimation methods illustrates the exigency of these methods and that this is a field of active research [42, 43, 44, 45, 46, 47, 48, 49, 50]. We hope this overview may guide the interested reader towards state of the art parameter estimation methods within and beyond biological dynamical systems identification.

Multiple shooting methods [41], divide the relevant time interval for the model into several smaller subintervals, introduce initial conditions for each subinterval, and solve the individual initial value problems on each subinterval using finite difference techniques. Constrained optimization methods will ensure continuity of the optimized state variables when using multiple shooting methods. These methods have shown to introduce robustness when compared to single shooting methods by reformulating the problem as a constrained optimization [41].

Another class of methods known as dynamic flux estimation and incremental parameter estimation have shown robust recovery of parameter estimates [45, 46, 42]. These methods work in two distinct phases. In a first model-less phase, data is approximated using sufficiently smooth parameterized functions. In the second phase the dynamic flux estimation uses the continuous approximation to potentially uncouple the differential equation and efficiently approximate parameters within the dynamical system [42, 44]. The incremental parameter estimation is very similar in nature [45, 51, 52] and is designed for homogenous reaction kinetics with the focus on system identification. Both methods are well established and optimized for metabolic pathway systems and provide robust parameter estimates by avoiding any finite difference scheme.

Collocation methods are finite element methods that take the approach of approximating the solution to the differential equation using sufficiently smooth parametrized functions, for instance cubic splines [53]. As a result, satisfying the differential equation at a finite number of points will define an approximating solution, and finding the forward solution generally amounts to solving a well-posed algebraic system [54]. Within this class of finite element parameter estimation methods there exist two approaches. One approach estimates the model parameter and the finite element parameter simultaneously, which is often referred to as an "all-at-once" approach [49, 48, 40]. The other approach alternates between solving the

differential equation using collocation type methods and updating the model parameters [55, 56].

The method we use here is often referred to as *principal differential analysis* and is closely related to collocation methods but instead uses a least squares finite element approach [24, 57, 25], so the distinguishing feature is that the forward solution amounts to solving a least squares problem [58]. Principal differential analysis is also related to dynamic flux estimation, as it can be seen as an iterated or all-at-once dynamic flux estimation where the first and second phase are repeated or simultaneously computed until convergence.

We chose the principal differential analysis approach for the following reasons. Compared to collocation methods, our method does not guarantee that the model equation will be satisfied exactly at certain points. Although this may seem as a disadvantage, it allows us to decouple the parameterization of the approximating function and the evaluation points of the misfit function allowing for the use of efficient algorithms. Furthermore, since we acknowledge that any model is incomplete and at best an approximation, which is especially true for biological systems, this allows the optimal approximating function to display dynamics not captured in the model construction. Compared to dynamic flux estimation, the all-at-once data approximation and parameter estimation prevents eventual bias towards initial data approximation.

Despite the advantages of these robust methods, they are still vastly underutilized in their application, e.g., in mathematical biology. This encouraged us to provide a comprehensive derivation (Section 2.4), theoretical and numerical details (Sections 2.5 and 2.6 and Appendix A and Appendix B), and a Matlab implementation of our method (Appendix C) to interested readers in this community.

2.4. Principal Differential Analysis

In this manuscript we utilize principal differential analysis for parameter estimation. While this approach has been well-established as noted above [24, 25], we provide details of our approach for interested readers. We first reformulate (2) as

$$\min_{(\mathbf{p}, \mathbf{y})} \|\mathbf{W}(\mathbf{m}(t; \mathbf{y}) - \mathbf{d})\|_2^2 + \alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{y})) \quad (3)$$

$$\text{subject to } \|\mathbf{y}' - \mathbf{f}(t, \mathbf{y}, \mathbf{p})\|_{\mathcal{L}^p}^p = 0,$$

where $\|\cdot\|_{\mathcal{L}^p}$ is any appropriate integral norm with $p = 2$ for our later examples. The problems defined in (2) and (3) are equivalent since $\mathbf{y}$ is required to be continuous. Relaxing the differential equation constraint leads to

$$\min_{(\mathbf{p}, \mathbf{y})} \|\mathbf{W}(\mathbf{m}(t; \mathbf{y}) - \mathbf{d})\|_2^2 + \lambda \|\mathbf{y}' - \mathbf{f}(t, \mathbf{y}, \mathbf{p})\|_{\mathcal{L}^2}^2 + \alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{y})).$$

Here, $\lambda \geq 0$ and $\alpha \geq 0$ can be seen as either Lagrange multipliers or regularization parameters [26, 59]. When $\alpha = 0$ the parameter λ has the effect that if λ is small, the main contributor to the minimization process is the data misfit. If λ vanishes, i.e., $\lambda = 0$, only the data misfit term is influential and the model equation is not relevant, which leads to data overfitting [28, 24]. On the other hand, if λ is large, the weight of the misfit shifts to the model equations and may ultimately disregard the data fitting term leading to underfitting [28]. A similar interpretation holds for the influence of α and the combined influences of α and λ when $\alpha \geq 0$.

Let $\mathcal{S}_\tau^3([a, b])$ be the set of cubic splines with knots $a = \tau_0 < \dots < \tau_k = b$ with a chosen set of boundary conditions, e.g., not-a-knot conditions. Then every $s \in \mathcal{S}_\tau^3([a, b])$ is uniquely determined by a set of parameters $\tilde{\mathbf{q}} = [\tilde{q}_0, \dots, \tilde{q}_k]^\top$. A vectorized spline $\mathbf{s}$ defined by $\mathbf{s} = [s_1, \dots, s_n]^\top$ with $s_j \in \mathcal{S}_\tau^3([a, b])$ for $j = 1, \dots, n$ is then uniquely determined by a parameter vector $\mathbf{q} = [\mathbf{q}_1, \dots, \mathbf{q}_n]^\top$. Approximating the state variable $\mathbf{y}$ by $\mathbf{s}$, discretizing the integral of the $\mathcal{L}^2$ -norm, and normalizing λ finally leads to the optimization problem

$$\min_{(\mathbf{p}, \mathbf{q})} \mathcal{J}(\mathbf{p}, \mathbf{q}) = \|\mathbf{W}(\mathbf{m}(t; \mathbf{s}(t; \mathbf{q})) - \mathbf{d})\|_2^2 + \frac{\lambda}{n\ell} \|\mathbf{s}'(\mathbf{T}; \mathbf{q}) - \mathbf{f}(\mathbf{T}, \mathbf{s}(\mathbf{T}; \mathbf{q}), \mathbf{p})\|_2^2 + \alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{s}(t; \mathbf{q}))), \quad (4)$$

where we define

$$\mathbf{s}'(\mathbf{T}; \mathbf{q}) = \begin{bmatrix} \mathbf{s}'(T_1; \mathbf{q}) \\ \vdots \\ \mathbf{s}'(T_\ell; \mathbf{q}) \end{bmatrix}, \quad \mathbf{f}(\mathbf{T}, \mathbf{s}(\mathbf{T}; \mathbf{q}), \mathbf{p}) = \begin{bmatrix} \mathbf{f}(T_1, \mathbf{s}(T_1; \mathbf{q}), \mathbf{p}) \\ \vdots \\ \mathbf{f}(T_\ell, \mathbf{s}(T_\ell; \mathbf{q}), \mathbf{p}) \end{bmatrix},$$

and $a = T_1 < \dots < T_\ell = b$ is a discretization of the interval $[a, b]$. Equation (4) is the maximum a-posteriori (MAP) estimator under the Gaussian assumption on the likelihood and on the prior for the model parameter $\mathbf{p}$ [27].

2.5. Method Analysis

Using this method we do not find a solution to the original problem statement, but we solve a “nearby” problem instead with the idea that solving (4) is a more robust approach. It is feasible to criticize that the optimization problem being solved is just an approximation to the solution of (2), but even in cases where the solution to (4) is not sufficiently accurate, this method can be used to efficiently precompute approximations for $\hat{\mathbf{p}}$ and $\hat{\mathbf{y}}_0$, which can then be used as initial guesses for single or multiple shooting methods.

One step in constructing our “nearby” problem is replacing the state variable $\mathbf{y}$ in the model with an approximation $\mathbf{s}$. In our case, we use cubic spline functions for $\mathbf{s}$, and for $\mathbf{s} = [s_1, \dots, s_n]^\top$ the approximation error for each $s_j \in \mathcal{S}_\tau^3([a, b])$, $j = 1, \dots, n$, is bounded using the theorem below.

Theorem 2.1 ([60])—Let m be a positive integer. For every $y \in \mathcal{C}^m([a, b])$ and for every integer $j \in \{1, \dots, \min(m, 4)\}$, the least maximum error satisfies the condition

$$\min_{s \in \mathcal{S}_{\tau}^3([a, b])} \|y - s\|_{\infty} \leq \frac{4!}{(4-j)!} \frac{1}{2^j} h^j \|y^{(j)}\|_{\infty},$$

where $h = \max \{ \tau_{i+1} - \tau_i : i = 0, \dots, k-1 \}$.

Since $y_j \in \mathcal{C}^1([a, b])$ for $j = 1, \dots, n$ the bound on the approximation error for each $s_j \in \mathcal{S}_{\tau}^3([a, b])$, $j = 1, \dots, n$, simplifies to

$$\min_{s \in \mathcal{S}_{\tau}^3([a, b])} \|y - s\|_{\infty} \leq 2h \|y'\|_{\infty}.$$

While the solving the “nearby problem” is a more robust approach than solving the problem given in (2), the dimension of the optimization problem increases. This can adversely affect the speed of the optimization step, but is also counteracted by improvements in computational efficiency elsewhere. One example is that optimization steps in (4) never require the calculation of the solution of the ODE model. Eliminating the need for the solution of the initial value problem removes a computationally intensive step in each optimization iteration and replaces the step with the analytic evaluation of the spline vector $\mathbf{s}$ and its time derivative $\mathbf{s}'$.

2.6. Computational Details

Problem (4) might be ill-posed, which means a solution does not exist, is not unique, or does not depend continuously on the data [61]. As a result, the regularization given by the inclusion of the term $\alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{y}))$ is required to obtain a meaningful solution and prevent overfitting [28, 62]. In the examples to follow, we include an approximation of the 1-norm on the interaction matrix parameters as our regularization.

Various methods have been proposed to include regularization and estimate the regularization parameters [62, 63, 45, 64], and we choose to use k -fold cross-validation, which is a standard method in statistics to get reliable regularization parameters [65, 66]. The reason we use cross-validation here is the method allows us to use a subset of known data to train our model by solving (4) and the remaining known data to test the model’s predictive abilities. This is particularly useful in biological applications such as the ones being discussed here because the available data are often limited and hard to collect. It is also important to note that the estimation of adequate regularization parameters is a necessary step that can dominate the computational costs, but automating the process can prevent over- or underfitting [67] of the data.

To numerically solve (4) with respect to $\mathbf{p}$ and $\mathbf{q}$, we utilize Gauss-Newton type methods. Using a Gauss-Newton approach guarantees (locally) fast convergence of the optimization problem (4) to a local minimum [26]. Similar parameter estimation approaches have shown that if the model parameter $\mathbf{p}$ enters the model linearly, only one system solve may be

required to get the optimal estimates [43, 47]. Optimization problem (4) may include nonlinear spline parameters $\mathbf{q}$ and nonlinear regularization on the parameters $\mathbf{p}$ though, so convergence in a single iteration is not always guaranteed.

Gauss-Newton type methods are generally also local optimization methods, so we empirically sample the parameter space and independently repeat the optimization process with various initial guesses to obtain a global minimum. We use a Latin hypercube sampling, but this approach can be adapted easily to different sampling methods such as a Monte Carlo sampling or a predetermined set of sample points.

Finally, Gauss-Newton type methods are gradient based methods so they require computing of various derivatives. It is important to note that we can find the structures for $\mathbf{m}$, $\mathbf{s}$, $\mathbf{f}$, $\mathbf{c}$ and their corresponding partial derivatives analytically, which improves computational efficiency. In our examples $\mathcal{D}$ is an approximation of the 1-norm, but for the simplicity of an example, suppose $\mathcal{D}$ is the two-norm. Then a standard optimization algorithm can be written as follows. Let the residuals of (4) be defined by

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{W}(\mathbf{m}(t; \mathbf{s}(t; \mathbf{q})) - \mathbf{d}) \\ \sqrt{\lambda}(\mathbf{s}'(\mathbf{T}; \mathbf{q}) - \mathbf{f}(\mathbf{T}, \mathbf{s}(\mathbf{T}; \mathbf{q}), \mathbf{p})) \\ \sqrt{\alpha} \mathbf{c}(\mathbf{p}, \mathbf{s}(t; \mathbf{q})). \end{bmatrix}$$

and the Jacobian of $\mathbf{r}$ be given by

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{r}_1}{\partial \mathbf{p}} & \frac{\partial \mathbf{r}_1}{\partial \mathbf{q}} \\ \frac{\partial \mathbf{r}_2}{\partial \mathbf{p}} & \frac{\partial \mathbf{r}_2}{\partial \mathbf{q}} \\ \frac{\partial \mathbf{r}_3}{\partial \mathbf{p}} & \frac{\partial \mathbf{r}_3}{\partial \mathbf{q}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{W} \mathbf{m}_s \mathbf{s}_q \\ -\sqrt{\lambda} \mathbf{f}_p(\mathbf{T}) & \sqrt{\lambda}(\mathbf{s}'_q(\mathbf{T}) - \mathbf{f}_s(\mathbf{T}) \mathbf{S}_q(\mathbf{T})) \\ \sqrt{\alpha} \mathbf{c}_p & \sqrt{\alpha} \mathbf{c}_s \mathbf{s}_q \end{bmatrix}$$

with the appropriate abbreviations

$$\mathbf{m}_s = \frac{\partial}{\partial \mathbf{s}} \mathbf{m}(t; \mathbf{s}), \quad \mathbf{f}_p(\mathbf{T}) = \frac{\partial}{\partial \mathbf{p}} \mathbf{f}(\mathbf{T}, \mathbf{s}, \mathbf{p}),$$

$$\mathbf{s}_q = \frac{\partial}{\partial \mathbf{q}} \mathbf{s}(t; \mathbf{q}), \quad \mathbf{f}_s(\mathbf{T}) = \frac{\partial}{\partial \mathbf{s}} \mathbf{f}(\mathbf{T}, \mathbf{s}, \mathbf{p}),$$

$$\mathbf{s}_q(\mathbf{T}) = \frac{\partial}{\partial \mathbf{q}} \mathbf{s}(\mathbf{T}; \mathbf{q}), \quad \mathbf{c}_p = \frac{\partial}{\partial \mathbf{p}} \mathbf{c}(\mathbf{p}, \mathbf{q}),$$

$$\mathbf{s}'_q(\mathbf{T}) = \frac{\partial}{\partial \mathbf{q}} \mathbf{s}'(\mathbf{T}; \mathbf{q}), \quad \mathbf{c}_s = \frac{\partial}{\partial \mathbf{s}} \mathbf{c}(\mathbf{p}, \mathbf{s}).$$

Then the gradient of the objective function $\mathcal{J}$ is given by $\mathbf{g} = \nabla \mathcal{J}(\mathbf{p}, \mathbf{q}) = 2\mathbf{J}^T \mathbf{r}$ and the Gauss-Newton approximation of the Hessian is $\mathbf{H} = \nabla^2 \mathcal{J}(\mathbf{p}, \mathbf{q}) \approx 2\mathbf{J}^T \mathbf{J}$. The details for the construction of the cubic splines in $\mathbf{s}$ and their partial derivatives can be found in Appendix A, and the details for the construction of $\mathbf{f}$ and its partial derivatives when using the Lotka-Volterra system as a model can be found in Appendix B.1.

We provide a fully functional Matlab implementation of our principal differential analysis method in Appendix C, including options for the regularization of ill-posed problems using cross-validation and for “global optimization” via Latin hypercube sampling. Furthermore, our Matlab code uses parallel computing to efficiently estimate parameters and contains an example problem to try.

3. Results and Discussion

In this section we apply our approach to biological systems using simulation studies and previously collected and published data for both intestinal and vaginal microbiota and share our findings. We focused on the Lotka-Volterra system of differential equations as a model and included an unknown sparsity pattern in the interaction matrix as an assumption and constraint on the model.

3.1. Simulation Studies

On the interval $t \in [0, 10]$, we defined a four dimensional Lotka-Volterra system using the parameters and initial conditions

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -3 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & -0.6 & 0 & -0.2 \\ 0.6 & 0 & -0.6 & -0.2 \\ 0 & 0.6 & 0 & -0.2 \\ 0.2 & 0.2 & 0.2 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y}_0 = \begin{bmatrix} 5 \\ 4 \\ 3 \\ 2 \end{bmatrix}.$$

In this case, the interaction matrix was sparse with 37.5 percent of its entries being zero. Using the parameters and initial conditions, we numerically solved the initial value problem with the differential equation given by (1). In Figure 1 a phase plot of the solutions for States 1 through 4 indicates the system displayed chaotic dynamics, and it is inherently difficult for parameter estimation methods to find parameters of chaotic systems.

We then used the numerical solution to generate three sets of data with different levels of multiplicative noise (Study 1: 0 percent noise; Study 2: up to 10 percent noise; Study 3: up to 25 percent noise) and applied our parameter estimation method using each data set. The data for each study are shown in Figure 2, and the remaining details for the problem setup are included in Appendix B.2. The code provided in Appendix C also allows for setting up similar studies.

The state solutions of the system found using the optimal model parameters recovered the data very well in all three studies. The relative errors, defined as $e_r = \frac{1}{m} \left\| \frac{\mathbf{m}(\mathbf{y}) - \mathbf{d}}{\mathbf{d}} \right\|_1$ where

m is the number of elements in $\mathbf{d}$ and the division is element-wise, were approximately $e_r \approx 0.0331$, $e_r \approx 0.0926$, and $e_r \approx 0.1511$ for the three studies, respectively.

As Figure 3 demonstrates, we also compared our optimal model parameters to their true values, and the absolute errors in the optimal model parameters suggested the true model parameters and the system dynamics were recovered very well in all three studies.

By plotting the relative error between the spline solutions and the state solutions in Figure 4, all three studies additionally illustrated that the optimal spline functions also proved to be a good approximation of the state solutions found using the optimal model parameters.

As an unknown sparsity pattern was an assumption in our model for these simulation studies, our method's ability to capture the true sparsity pattern was relevant. Due to numerical and computational limitations, it was unlikely for any model parameter in the optimal set to be identically 0, so we instead considered any model parameters with a magnitude below a certain threshold, 10^{-3} in this case, to essentially be 0. The sparsity structure was perfectly preserved in the first simulation study, 87.5 percent recovered in the second, and 62.5 percent recovered in the third. This was what we would expect as the increase in data noise over the three simulation studies should have had an increasingly significant effect on the accuracy of the model parameters that our method returned.

The model's local sensitivity to the optimal model parameters and initial conditions, which is the solution of the initial value problem

$$\frac{d}{dt} \begin{bmatrix} \mathbf{y} \\ \frac{\partial \mathbf{y}}{\partial \mathbf{p}} \\ \frac{\partial \mathbf{y}}{\partial \mathbf{y}_0} \end{bmatrix} = \begin{bmatrix} \mathbf{f}(t, \mathbf{y}, \mathbf{p}) \\ \frac{\partial}{\partial \mathbf{p}} \mathbf{f}(t, \mathbf{y}, \mathbf{p}) \\ \frac{\partial}{\partial \mathbf{y}_0} \mathbf{f}(t, \mathbf{y}, \mathbf{p}) \end{bmatrix}, \begin{bmatrix} \mathbf{y}(a) \\ \frac{\partial}{\partial \mathbf{p}} \mathbf{y}(a) \\ \frac{\partial}{\partial \mathbf{y}_0} \mathbf{y}(a) \end{bmatrix} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{0} \\ \text{vec}(\mathbf{I}_n) \end{bmatrix}$$

over the interval $[a, b]$, was also of interest. To have some idea of the cumulative effect of a single parameter or initial condition on the entire n -state Lotka-Volterra system we calculated S_{p_i} for $i = 1, \dots, n^2 + n$ and $S_{y_{0,i}}$ for $i = 1, \dots, n$ with

$$S_{p_i} = \sum_{j=1}^n \int_a^b \left| \frac{\partial y_j(t)}{\partial p_i} \right| dt \quad \text{and} \quad S_{y_{0,i}} = \sum_{j=1}^n \int_a^b \left| \frac{\partial y_j(t)}{\partial y_{0,i}} \right| dt,$$

respectively, and provide the results in Figure 5.

The cumulative sensitivities remained consistent across the three simulation studies, which was to be expected given the consistency across the simulation studies in the optimal model parameters and initial conditions themselves.

3.2. Interactions within the Intestinal Microbiota

The generalized Lotka-Volterra formalism was recently used to explore the impact of the intestinal microbiota on the development of antibiotic-induced *Clostridium difficile* colitis [15]. This disease occurs in patients who have been treated with antibiotics and is

characterized by a marked shift of the intestinal microbiota towards a state dominated by the pathogen *Clostridium difficile*. In many cases health can only be restored through a fecal transplant, a procedure which restores the diversity of the microbiota. The mechanisms through which disease occurs, and through which the normal gut microbiota can prevent the over-growth of *C. difficile* are still not well understood.

In Stein et al. [15] the authors relied on a mouse model of *C. difficile* colitis to attempt to address these questions. They tracked the microbiota of mice across time and used the resulting data to estimate the parameters of a Lotka-Volterra model. Based on the resulting model they were able to provide new testable hypotheses about the factors that promote the overgrowth of *C. difficile* following a course of clindamycin. Here we used the same data and model and added the assumption of interaction sparsity to test our parameter estimation procedure and compare to the originally published results.

We focused on a subset of the Stein et al. data, specifically data originating from three mice who had not been subjected to any antibiotic interventions. The exact details, including how we setup our problem, can be found in the Appendix B.3. The original study characterized the abundance of five Operational Taxonomic Units (OTUs), representing clusters of sequences believed to originate from a same organism.

In our simulation, the spline solutions remained good approximations to the state solutions as demonstrated in Figure 6. The relative errors for the spline solutions for the *Blautia* and *Coprobacillus* OTUs were larger than the relative errors for the other five OTUs, but this was due to both the significantly smaller magnitude of the data for these OTUs and the magnitude of the weights for the data relative to the other OTUs. Among the three replicates for a single OTU, variations in the magnitude of the relative errors, e.g., in *Blautia*, *Unclassified Mollicutes*, and *Coprobacillus*, were explained by noticeable variations in the magnitude of the data across replicates.

As in the simulation studies, we also assess our method's ability to recover the data. The relative error between the state solutions using the optimal model parameters and the data

for all three mice, given by $e_r = \frac{1}{m} \left\| \frac{\mathbf{m}(\mathbf{y}) - \mathbf{d}}{\mathbf{d}} \right\|_1$ with m being the number of elements in $\mathbf{d}$ and the division being element-wise, was $e_r \approx 0.4594$. The relative error for the model published in [15] was $e_r \approx 3.6790$, indicating that our method more accurately captured the dynamics of the data. This fact was further confirmed by a visual comparison of the state solutions to the data in Figure 7.

The biological implication of the disagreement between our results and those originally published on the same data became apparent when examining graph representations of the Lotka-Volterra interaction matrices in Figure 8 and considering a recent paper from the same group providing experimental evidence for the role of the gut microbiota in the prevention of *C. difficile* infection [68]. Stein et al. originally concluded, on the basis of Lotka-Volterra modeling, that members of the *Coprobacillus* genus inhibit the growth of other members of the gut microbiome, which implied *Coprobacillus* is a stabilizing factor within the gut microbiome. In our own analysis of their data, we did not identify any strong interactions

between the *Coprobacillus* OTU and other organisms. Instead we observed inhibitory interactions of members of the *Lachnospiraceae* family with other gut microbes, which suggested that members of the *Undefined Lachnospiraceae* and *Unclassified Lachnospiraceae* groups are the more likely players involved in preventing *C. difficile* colonization. Buffie et al. confirmed this experimentally [68]. In their paper they showed that the *Lachnospiraceae* species *Clostridium scindens* can provide resistance to *C. difficile* colonization in a mouse model of *C. difficile* enterocolitis.

As with the simulation studies, we could calculate the cumulative local sensitivities with respect to the optimal model parameters and initial conditions for the intestinal microbiota model. Note that because of how the parameter estimation problem was setup (see Appendix B.3), the resulting model parameters for each of the three replicates were the same, but the optimal initial conditions differed. This meant the local sensitivities and hence the cumulative sensitivities could vary by replicate, yet the results in Figure 9 display consistency in the cumulative sensitivities across the replicates.

3.3. Interactions within the Vaginal Microbiota

Gajer et al. described the temporal dynamics of the human vaginal microbiome sampled twice per week over a 16-week period in 32 women [10]. Understanding the factors that drive community structure in this environment may provide insights into the stability of the system and the disruptions that lead to the development of bacterial vaginosis, a condition impacting millions of women in the United States. These data are more deeply sampled than the Stein et al. dataset described above, leading to a clearer picture of the system's dynamics.

We present here the results of analyzing the data obtained from subject 15 in the original study. We chose this particular subject because her data comprised most of the OTUs we determined to be likely to play an important role in the dynamics of the vaginal microbiota. The details of the analysis used to select this subject can be found in the Appendix B.4.

As shown in Figure 10, the results of the parameter estimation led to spline solutions that closely approximated the state solutions found using the optimal model parameters. The high relative error for *Lactobacillus jensenii*, *Ureaplasma*, and *Lactobacillus vaginalis* was attributed to the small magnitude of the state solutions in the second half of the time series.

The dynamics that result from the optimal Lotka-Volterra model also broadly matched the experimental data and captured the oscillation in the abundance observed for several of the organisms in this sample (*Lactobacillus crispatus* and *Lactobacillus otu5* in particular). See Figure 11. This result was perhaps surprising given that the oscillations are at least in part due to the physiological changes that occur during menstrual periods. The dynamics that the optimal Lotka-Volterra model captured indicate that inter-microbe interactions may play a role in these periodic changes as well.

The graph of the interaction matrix displayed in Figure 12 also highlighted that the dynamics of the optimal Lotka-Volterra model were due in part to several particular interactions. The OTU *Lactobacillus crispatus* had strong negative interactions with

Lactobacillus jensenii and *Lactobacillus otu5*, indicating a possible role of this organism in maintaining the stability of the vaginal microbiota. This observation was in agreement with previously reported epidemiological observations [69] suggesting *Lactobacillus crispatus* as a stabilizing factor in the human vaginal microbiota. The OTU *Lactobacillus otu5* also had positive interactions with multiple members of the vaginal microbiota suggesting this organism may produce compounds necessary to their growth, providing an initial insight into the potential function of this uncharacterized OTU.

The sensitivity analysis of the model parameters and initial conditions shown in Figure 13 suggested the Lotka-Volterra model for the vaginal microbiota was particularly sensitive to the parameters in the first column of the interaction matrix and not much else. This was a little deceiving because the perceived sensitivity was due to the large magnitude of the *Lactobacillus iners* data making it difficult to draw any conclusions about the sensitivity of the model to its parameters.

Another item of note is that when we performed a stability analysis of the optimal Lotka-Volterra model, we found the system has a stable equilibrium that is biologically relevant. The equilibrium abundances of the OTUs in order were approximately 2925.87 5.62, 0, 11.04, 22.23, 0, and 0. Extending the state solution in time, which is shown in Figure 14, showed that *Lactobacillus iners*, *Lactobacillus crispatus*, *Corynebacterium*, and *Lactobacillus otu5* oscillated around their equilibria with varying rates of decay in the magnitude of oscillations while *Lactobacillus jensenii*, *Ureaplasma*, and *Lactobacillus vaginalis* almost immediately decayed to theirs.

4. Conclusions

Using principal differential analysis eliminates the computational inefficiency and robustness-limiting step of numerically solving an initial value problem at every optimization step and also presents a natural framework for including prior knowledge or assumptions about the biological system being modeled. In defining and solving this new problem, we demonstrate that we have not sacrificed parameter or data recovery. This suggests this approach is a valuable tool, particularly for the increasingly complex parameter estimation problems like those found in systems biology.

Biologically speaking, principal differential analysis demonstrates the ability to closely model real biological data. In the case of the vaginal microbiota, the approach was able to capture the general abundance dynamics seen in the data and provide a full forward projection for the system. We would like to note though that the resulting model does not provide a full forward prediction in the case of the intestinal microbiota. The Lotka-Volterra system appears to be unstable leading to the uncontrolled growth, or the disappearance of certain taxa, phenomena not commonly observed in real data. In part this is due to the limitations of the Lotka-Volterra system and the simplifying assumptions made when choosing it as a model. Insufficient data both in terms of the relatively small number of samples and, more importantly, in terms of the sparse sampling rate also play a role. Unfortunately, these limitations are inherent as computational costs can require model

simplifications, experimental costs can limit the number of feasibly obtained samples, and the specific microbiome being sampled can potentially limit the sampling frequency.

Despite these limitations, we showed that we could identify interactions and their direction between members of the gut microbiome that were later confirmed by in vivo experiments. The approach we described here correctly identified the experimentally validated role of the *Lachnospiraceae* family as a stabilizing factor in the gut microbiome, in contrast with the suggestion that the *Coproacillus* OTU plays this role as predicted by a previous computational analysis. Likewise, our analysis correctly identifies the role of *Lactobacillus crispatus* as an important contributor to the stability of the vaginal microbiome. The network we infer also provides insights into the roles played by uncharacterized members of the vaginal community such as the organisms known as *Lactobacillus otu5*. Knowledge of the underlying interactions and their direction like these are an important aspect of guiding further biological experimentation.

Modeling approaches that exhibit robustness while having an appropriate theoretical framework are extremely useful in understanding host-associated microbial communities. The examples we provide here demonstrate how much information we can gather from experimental data when paired with the right inference approach, and we believe that approaches like principal differential analysis when paired with experimental work can help elucidate the role of host-associated microbes in health and disease. For future work, we will continue to build on what we have presented here by adding more tools and mechanisms, e.g., implementing alternative approximating functions, stochastic optimization methods and parameter confidence intervals, etc., to gain even more insights about the biological systems being analyzed.

Acknowledgments

The authors would like to thank Robert Torrence from Virginia Tech and Hector Bravo and Senthil Muthiah from the University of Maryland, College Park and the reviewers for providing their comments and feedback. This work was supported by the National Institute Of General Medical Sciences of the National Institutes of Health Award Number R21GM107683-01.

Appendix A

Cubic Spline Construction

Using gradient-based optimization methods for solving (4) requires the derivatives of the spline s with respect to the time and the coefficients $\mathbf{q}$. In particular we need to compute s , $s_{\mathbf{q}}$, s' , and $s'_{\mathbf{q}}$. Here, we provide the required derivations for the cubic spline function s and its derivatives.

Appendix A.1. Spline Definition

Let data points $a = t_0 < \dots < t_n = b$ and any corresponding real numbers q_j , $j = 0, \dots, n$ been given.

Let $s : [a, b] \rightarrow \mathbb{R}$ be a function with the interpolation properties

$$s(t_j)=q_j \quad \text{for } j=0, \dots, n \quad (\text{A.1})$$

that is equal to a cubic polynomial with coefficients a_j , b_j , c_j and d_j on each interval $[t_j, t_{j+1}]$, i.e.,

$$s(t) \Big|_{t \in [t_j, t_{j+1}]} = s_j(t) = a_j(t - t_j)^3 + b_j(t - t_j)^2 + c_j(t - t_j) + d_j \quad (\text{A.2})$$

for $t_j < t < t_{j+1}$, $j = 0, \dots, n-1$, and is twice differentiable, i.e.,

$$s'_j(t_{j+1}) = s'_{j+1}(t_{j+1}) \quad \text{and} \quad s''_j(t_{j+1}) = s''_{j+1}(t_{j+1}) \quad (\text{A.3})$$

for $j = 0, \dots, n-2$. We refer to s as a cubic spline [70].

Equations (A.1), (A.2) and (A.3) provide $4n-2$ conditions for the $4n$ coefficients a_j , b_j , c_j and d_j , $j = 0, \dots, n-1$. A spline s is now uniquely determined by choosing appropriate boundary conditions. Here, we choose not-a-knot boundary conditions, i.e., $s'''_0(t_1) = s'''_1(t_1)$ and $s'''_{n-2}(t_{n-1}) = s'''_{n-1}(t_{n-1})$. For efficient calculations of the coefficients a_j , b_j , c_j and d_j we define moments

$$m_j = s''(t_j)$$

for $j = 0, \dots, n$. Note that the second derivative of s is linear in each interval $[t_j, t_{j+1}]$, and for $t \in [t_j, t_{j+1}]$, the moments m_j have the expression

$$s''_j(t) = m_j \frac{t_{j+1} - t}{h_{j+1}} + m_{j+1} \frac{t - t_j}{h_{j+1}}$$

with $h_{j+1} = t_{j+1} - t_j$ for $j = 0, \dots, n-1$. Integrating and using equation (A.1) we get

$$s'_j(t) = -m_j \frac{(t_{j+1} - t)^2}{2h_{j+1}} + m_{j+1} \frac{(t - t_j)^2}{2h_{j+1}} + \frac{q_{j+1} - q_j}{h_{j+1}} - \frac{h_{j+1}}{6}(m_{j+1} - m_j)$$

$$s_j(t) = m_j \frac{(t_{j+1} - t)^3}{6h_{j+1}} + m_{j+1} \frac{(t - t_j)^3}{6h_{j+1}} + \left(\frac{q_{j+1} - q_j}{h_{j+1}} - \frac{h_{j+1}}{6}(m_{j+1} - m_j) \right) (t - t_j) + q_j - m_j \frac{h_{j+1}^2}{6}.$$

With the condition $s'_j(t_{j+1}) = s'_{j+1}(t_{j+1})$ for $j = 1, \dots, n-1$ we get the equation

$$\frac{h_j}{6}m_{j-1} + \frac{h_j+h_{j+1}}{3}m_j + \frac{h_{j+1}}{6}m_{j+1} = \frac{q_{j+1}-q_j}{h_{j+1}} - \frac{q_j-q_{j-1}}{h_j},$$

which can be rewritten as

$$(1-\lambda_j)m_{j-1} + 2m_j + \lambda_j m_{j+1} = \frac{6}{h_j+h_{j+1}} \left(\frac{q_{j+1}-q_j}{h_{j+1}} - \frac{q_j-q_{j-1}}{h_j} \right)$$

with $\lambda_j = \frac{h_{j+1}}{h_j+h_{j+1}}$ for $j = 1, \dots, n-1$.

Now the moments $\mathbf{m} = [m_0, \dots, m_n]^\top$ can be calculated by solving the linear system

$$\mathbf{A}\mathbf{m} = \boldsymbol{\beta}.$$

Here, $\boldsymbol{\beta} = [\beta_0, \dots, \beta_n]^\top$ with $\beta_j = \frac{6}{h_j+h_{j+1}} \left(\frac{q_{j+1}-q_j}{h_{j+1}} - \frac{q_j-q_{j-1}}{h_j} \right)$ for $j = 1, \dots, n-1$ and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_0 \\ \tilde{\mathbf{A}} \\ \mathbf{A}_n \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}$$

with

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1-\lambda_1 & 2 & \lambda_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \\ & & & & & 1-\lambda_{n-1} & 2 & \lambda_{n-1} \end{bmatrix}.$$

The entries $\mathbf{A}_0$, $\mathbf{A}_n$, β_0 , and β_n are specified by the boundary conditions. For the not-a-knot conditions,

$$\mathbf{A}_0 = \left[-\frac{1}{h_1}, \frac{1}{h_1} + \frac{1}{h_2}, -\frac{1}{h_2}, 0, \dots, 0 \right],$$

$$\mathbf{A}_n = \left[0, \dots, 0, -\frac{1}{h_{n-1}}, \frac{1}{h_{n-1}} + \frac{1}{h_n}, -\frac{1}{h_n} \right],$$

$$\beta_0 = \beta_n = 0.$$

With $\mathbf{m}$ given, coefficients a_j , b_j , c_j and d_j are determined by

$$d_j = q_j$$

$$c_j = \frac{q_{j+1} - q_j}{h_{j+1}} - \frac{(2m_j + m_{j+1})h_{j+1}}{6}$$

$$b_j = \frac{m_j}{2} \quad (\text{A.4})$$

$$a_j = \frac{m_{j+1} - m_j}{6h_{j+1}}$$

for $j = 0, \dots, n-1$.

Appendix A.2. Derivatives of the Spline Function

We are interested in calculating $s_{\mathbf{q}}$, where $\mathbf{q} = [q_0, \dots, q_n]^T$. With equation (A.2) we get

$$s_{\mathbf{q}}(t)|_{t \in [t_j, t_{j+1}]} = \frac{ds_j}{d\mathbf{q}}(t) = \frac{d}{d\mathbf{q}}(a_j(t-t_j)^3 + b_j(t-t_j)^2 + c_j(t-t_j) + d_j) = (t-t_j)^3 \frac{da_j}{d\mathbf{q}} + (t-t_j)^2 \frac{db_j}{d\mathbf{q}} + (t-t_j) \frac{dc_j}{d\mathbf{q}} + \frac{dd_j}{d\mathbf{q}}.$$

First, the coefficients a_j , b_j , c_j and d_j depend on the moments $\mathbf{m}$. The derivative

$$\frac{d\mathbf{m}}{d\mathbf{q}} = \frac{d}{d\mathbf{q}}(\mathbf{A}^{-1}\boldsymbol{\beta}) = \mathbf{A}^{-1} \frac{d\boldsymbol{\beta}}{d\mathbf{q}} = \mathbf{A}^{-1} \mathbf{B} \text{ with}$$

$$\mathbf{B} = \frac{d\boldsymbol{\beta}}{d\mathbf{q}} = \begin{bmatrix} 0 & 0 & 0 & & & & \\ -\mu_1\lambda_1 & \mu_1 & -\mu_1(1-\lambda_1) & & & & \\ & -\mu_2\lambda_2 & \mu_2 & -\mu_2(1-\lambda_2) & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -\mu_{n-1}\lambda_{n-1} & \mu_{n-1} & -\mu_{n-1}(1-\lambda_{n-1}) & \\ & & & 0 & 0 & 0 & \end{bmatrix}$$

for the not-a-knot boundary conditions and for $\mu_j = -\frac{6}{h_j h_{j+1}}$, $j = 1, \dots, n-1$. We can

calculate $\frac{d\mathbf{m}}{d\mathbf{q}}$ by solving the linear systems $\mathbf{A}\mathbf{M} = \mathbf{B}$ where $\mathbf{M} = \frac{d\mathbf{m}}{d\mathbf{q}}$.

Finally, the derivatives of the coefficients with respect to $\mathbf{q}$, i.e., $\frac{da}{d\mathbf{q}}$, $\frac{db}{d\mathbf{q}}$, $\frac{dc}{d\mathbf{q}}$, and $\frac{dd}{d\mathbf{q}}$ with $\mathbf{a} = [a_0, \dots, a_{n-1}]^T$, $\mathbf{b} = [b_0, \dots, b_{n-1}]^T$, $\mathbf{c} = [c_0, \dots, c_{n-1}]^T$ and $\mathbf{d} = [d_0, \dots, d_{n-1}]^T$, can be computed using (A.4) and are given by

$$\frac{dd}{d\mathbf{q}} = \mathbf{E}_n$$

$$\frac{dc}{d\mathbf{q}} = \mathbf{H}_{-1} \hat{\mathbf{E}} \ddot{\mathbf{Y}} (\mathbf{E}_0 - \mathbf{E}_n) - \frac{1}{6} \mathbf{H}_1 \hat{\mathbf{E}} \ddot{\mathbf{Y}} (2\mathbf{E}_0 + \mathbf{E}_n) \mathbf{M}$$

$$\frac{db}{d\mathbf{q}} = \frac{1}{2} \mathbf{E}_n \mathbf{M}$$

$$\frac{da}{d\mathbf{q}} = \frac{1}{6} \mathbf{H}_{-1} \hat{\mathbf{E}} \ddot{\mathbf{Y}} (\mathbf{E}_0 - \mathbf{E}_n) \mathbf{M},$$

where $\mathbf{E}_0 = [\mathbf{0}, \mathbf{I}_n]$, $\mathbf{E}_n = [\mathbf{I}_n, \mathbf{0}]$, $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{H}_1 = [h_1, \dots, h_n]^T \otimes [1, \dots, 1]$, and $\mathbf{H}_{-1} = [1/h_1, \dots, 1/h_n]^T \otimes [1, \dots, 1]$. The symbols $\otimes$ and $\odot$ denote the Kronecker product the Hadamard product, respectively.

The further required derivatives s' and s'_q can now be easily computed by

$$s'(t)|_{t \in [t_j, t_{j+1}]} = 3a_j(t - t_j)^2 + 2b_j(t - t_j) + c_j$$

and

$$s'_q(t)|_{t \in [t_j, t_{j+1}]} = 3(t - t_j)^2 \frac{da_j}{d\mathbf{q}} + 2(t - t_j) \frac{db_j}{d\mathbf{q}} + \frac{dc_j}{d\mathbf{q}}.$$

Appendix B

Additional Computational Details

To run this method with a gradient-based optimization method on a problem that uses a Lotka-Volterra model with assumptions on sparsity, we require specific derivatives of the model and a mathematical formulation of the sparsity assumption. Here, we include these details and also our step-by-step approaches to both the simulation studies and the intestinal microbiota example.

Appendix B.1. Lotka-Volterra Models and Sparsity Constraint Details

Using the Lotka-Volterra system (1) as our model, we assume the parameters $\mathbf{p} = [\mathbf{b}; \text{vec}(\mathbf{A})]$ are considered unknown. Let us also replace the model state variable $\mathbf{y}$ with its approximation $\mathbf{s}$. The derivatives of $\mathbf{f}$ with respect to $\mathbf{s}$ and $\mathbf{p}$ are given by

$$\mathbf{f}_{\mathbf{s}} = \text{diag}(\mathbf{b}) + \text{diag}(\mathbf{s})\mathbf{A} + \text{diag}(\mathbf{A}\mathbf{s})$$

$$\mathbf{f}_{\mathbf{p}} = \begin{bmatrix} \text{diag}(\mathbf{s}), & \mathbf{s}^{\top} \otimes \text{diag}(\mathbf{s}) \end{bmatrix}$$

with $\otimes$ denoting the Kronecker product.

To include a sparsity constraint on the interaction matrix $\mathbf{A}$ in (4), the constraint term $\alpha \mathcal{D}(\mathbf{c}(\mathbf{p}, \mathbf{s}(t, \mathbf{q})))$ becomes $\alpha \|\text{vec}(\mathbf{A})\|_1$. In this case $\mathcal{D}$ is the one-norm, and $\mathbf{c}$ is the function that maps $\mathbf{p}$ to a vector of the parameters in $\mathbf{A}$. Note that this constraint is not differentiable everywhere, but one way to overcome this is by approximating the 1-norm using a smooth function. The approximation we use is

$$\|\text{vec}(\mathbf{A})\|_1 \approx \sum_{i=n+1}^{n^2+n} H_{\varepsilon}(p_i) \quad (\text{B.1})$$

where H_{ε} is the Huber function defined by

$$H_{\varepsilon}(x) = \begin{cases} x - \frac{\varepsilon}{2}, & |x| \geq \varepsilon \\ \frac{x^2}{2\varepsilon}, & |x| < \varepsilon. \end{cases}$$

The idea here is that the function $\|\text{vec}(\mathbf{A})\|_1$ is approximated by a smooth quadratic curve near its corners with “near” being defined by the choice of ε . Another important note regarding this modification to the objective function is that the effect on the numerics of the method. The data-fitting and model-fitting contributions to the function, gradient, and Hessian terms can be calculated as before. The contributions of the sparsity constraint term, however, require the first and second derivatives with respect to $\mathbf{p}$ and $\mathbf{q}$ of the approximation given in (B.1).

Appendix B.2. Simulation Study Details

Given the Lotka-Volterra system with parameters and initial conditions

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -3 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & -0.6 & 0 & -0.2 \\ 0.6 & 0 & -0.6 & -0.2 \\ 0 & 0.6 & 0 & -0.2 \\ 0.2 & 0.2 & 0.2 & 0 \end{bmatrix}, \quad \text{and } \mathbf{y}_0 = \begin{bmatrix} 5 \\ 4 \\ 3 \\ 2 \end{bmatrix},$$

we numerically solve the initial value problem and let $\mathbf{y}_{\text{true}}$ denote this forward solution. We then collect the values of $\mathbf{y}_{\text{true}}$ at the times given by the uniform discretization $0 = t_1 < \dots < t_{20} = 10$ and perturb them to generate the data $\mathbf{d} = \text{vec}(\mathbf{D}) = \text{vec}([\mathbf{d}_1, \dots, \mathbf{d}_{20}])$. Here, $\mathbf{d}_j = (1 + \boldsymbol{\varepsilon}_j) \mathbf{y}_{\text{true},j}$ for $j = 1, \dots, 20$ with $\boldsymbol{\varepsilon}_j$ representing a scaled vector of independent and identically Beta distributed noise, i.e., $\boldsymbol{\varepsilon}_j \sim \gamma \cdot (\text{Beta}(2, 2) - 1/2)$. We conduct three studies, each with different noise level scales (Study 1: $\gamma = 0$; Study 2: $\gamma = 0.1$; Study 3: $\gamma = 0.25$).

In the objective function, the projection $\mathbf{m}$ is the identity projection for all three studies, but the weight matrix $\mathbf{W}$ is different for each study because it depends on the data and is taken to be $\mathbf{W} = \mathbf{I}_{20} \otimes \text{diag}([w_1, \dots, w_4]^T)$ where $\mathbf{I}_{20}$ is the 20×20 identity matrix and $w_i = 10 / \sigma(\mathbf{D}_i)$, $i = 1, \dots, 4$, is the linearly scaled weighting of the inverse standard deviation of state i 's time-series data. Additionally, we use the standard constraint term in the objective function with a sparsity constraint on the interaction matrix $\mathbf{A}$, which was defined earlier. For each study, we separately sample 1,000 (λ, α) -pairs from the square $[1, 100] \times [0.01, 1]$ and choose a pair using leave-one-out cross-validation. The (λ, α) -pairs are approximately (1.1416, 0.01261), (5.5098, 0.04584), (23.8228, 0.87599) for studies 1, 2, and 3, respectively.

For each study, we then separately sample the parameter space 10,000 times using a Latin hypercube sampling and perform local optimizations using the Gauss-Newton method with each sample serving as an initial parameter set. The global minimizer is the local minimizer that most minimizes the objective function.

Appendix B.3. Intestinal Microbiota Details

The data collected consists of the abundance levels for eleven operational taxonomic units (OTUs) on days 1, 3, 7, 14, and 21 for each of the three mice. We eliminate any OTU that was not present in a measurable amount at all time points for any of the mice, which reduces our data to seven OTUs labeled *Blautia*, *Barnesiella*, *Unclassified Mollicutes*, *Undefined Lachnospiraceae*, *Unclassified Lachnospiraceae*, *Coprobaecillus*, and *Other*. Here, *Other* is the eleventh original OTU and is the collection of bacteria not assigned to any of the other ten original OTUs.

For our method we use all 21 (seven OTUs for three mice) time-series as the data, but we model the seven OTU interactions using a single, seven state Lotka-Volterra model given by

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}) = \text{diag}(\mathbf{y})(\mathbf{b} + \mathbf{A}\mathbf{y}), \mathbf{y}(1) = \mathbf{y}_0.$$

This then means that the model given to the objective function is actually

$$\tilde{\mathbf{y}}' = \tilde{\mathbf{f}}(\tilde{\mathbf{y}}) = \text{diag}(\tilde{\mathbf{y}})(\tilde{\mathbf{b}} + \tilde{\mathbf{A}}\tilde{\mathbf{y}}), \tilde{\mathbf{y}}(1) = \tilde{\mathbf{y}}_0$$

with

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ \mathbf{y}^3 \end{bmatrix}, \tilde{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ \mathbf{b} \\ \mathbf{b} \end{bmatrix}, \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & & \\ & \mathbf{A} & \\ & & \mathbf{A} \end{bmatrix}, \tilde{\mathbf{y}}_0 = \begin{bmatrix} \mathbf{y}_0^1 \\ \mathbf{y}_0^2 \\ \mathbf{y}_0^3 \end{bmatrix}.$$

The superscripts indicate the data separately collected from each of the three mice, so our method approximates 21 different state variables, corresponding to the 21 total time-series, with spline functions. Those approximations, however, are all governed by a single seven-state Lotka-Volterra system defined by the intrinsic growth vector $\mathbf{b}$ and the interaction matrix $\mathbf{A}$.

The projection $\mathbf{m}$ in the objective function is the identity projection, and the weight matrix $\mathbf{W}$ is defined as $\mathbf{W} = \text{diag}(\text{vec}([\tilde{\mathbf{W}}^T, \tilde{\mathbf{W}}^T, \tilde{\mathbf{W}}^T]^T))$ where $\tilde{w}_{ij} = 1/(100 \times \sigma(d_{ij1}, d_{ij2}, d_{ij3}))$, $i = 1, \dots, 7, j = 1, \dots, 5$, is the linearly scaled weighting of the inverse standard deviation across replicates. We also replace the standard constraint term in the objective function with a sparsity constraint on the interaction matrix $\mathbf{A}$, which was defined earlier. To find the regularization parameters λ and α , we sample 100 (λ, α) -pairs from the square $[1, 100] \times [10^{-6}, 10^{-4}]$ and choose a pair using 12-fold cross-validation. The (λ, α) -pair is approximately $(2.6727, 5.7508 \times 10^{-6})$.

We then sample the parameter space 1,000 times using a Latin hypercube sampling and use these samples as initial parameter sets for local optimizations by the Gauss-Newton method. The global minimizer is the local minimizer that most minimizes the objective function.

Appendix B.4. Vaginal Microbiota Details

The data collected for Subject 15 in this experiment consists of abundance levels for 330 OTUs on days 1, 5, 8, 12, 15, 22, 26, 33, 36, 40, 43, 47, 50, 57, 61, 64, 68, 71, 75, 78, 82, 85, 89, 92, 96, 99, 103, 106, and 110. We eliminate any OTU whose leverage score is below 0.9 and then further eliminate any OTU with fewer than 20 nonzero abundance levels over the 29 time points. This leaves us with the seven OTUs labeled *Lactobacillus iners*, *Lactobacillus crispatus*, *Lactobacillus jensenii*, *Corynebacterium*, *Lactobacillus otu5*, *Ureaplasma*, and *Lactobacillus vaginalis*. We model the dynamics of these seven OTUs using the generalized Lotka-Volterra system.

The projection $\mathbf{m}$ in the objective function is the identity projection, and the weight matrix $\mathbf{W}$ is taken to be $\mathbf{W} = \mathbf{I}_{29} \otimes \text{diag}([w_1, \dots, w_7]^T)$ where $\mathbf{I}_{29}$ is the 29×29 identity matrix and $w_i = 1/(100 \times \sigma(\mathbf{D}_i))$, $i = 1, \dots, 7$, is the linearly scaled weighting of the inverse standard deviation of state i 's time-series data. We also include a sparsity constraint term on the interaction matrix $\mathbf{A}$ in the objective function.

In this case, we avoid any cross-validation and assign values to the regularization parameters, so the (λ, α) -pair is $(1 \times 10^{-2}, 1 \times 10^{-3})$. We then sample the parameter space 10,000 times using a Latin hypercube sampling and use the parameters as initial points for local Gauss-Newton optimizations. The global minimizer is the local minimizers that returns the smallest objective function.

Appendix C

MATLAB Code

The MATLAB code used for this work can be found in the online version.

References

1. Liu B, Faller LL, Klitgord N, Mazumdar V, Ghodsi M, Sommer DD, Gibbons TR, Treangen TJ, Chang YC, Li S, Stine OC, Hasturk H, Kasif S, Segrè D, Pop M, Amar S. Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS One*. 2012; 7(6):e37919. [PubMed: 22675498]
2. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012; 490(7418):55–60. [PubMed: 23023125]
3. Kong HH, Oh J, Deming C, Conlan S, Grice EA, Beatson MA, Nomicos E, Polley EC, Komarow HD, Program NCS, Murray PR, Turner ML, Segre JA. Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res*. 2012; 22(5):850–859. [PubMed: 22310478]
4. Zella G, Hait E, Glavan T, Gevers D, Ward D, Kitts C, Korzenik J. Distinct microbiome in pouchitis compared to healthy pouches in ulcerative colitis and familial adenomatous polyposis. *Inflamm Bowel Dis*. 2010; 17(5):1092–1100. [PubMed: 20845425]
5. Macfarlane S, Steed H, Macfarlane GT. Intestinal bacteria and inflammatory bowel disease. *Crit Rev Clin Lab Sci*. 2009; 46(1):25–54. [PubMed: 19107650]
6. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, Mc-Culle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A*. 2011; 108(Suppl 1):4680–4687. [PubMed: 20534435]
7. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biol*. 2007; 5(7):e177. [PubMed: 17594176]
8. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*. 2011; 108(Suppl 1):4578–4585. [PubMed: 20668239]
9. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R. Moving pictures of the human microbiome. *Genome Biol*. 2011; 12(5):R50. [PubMed: 21624126]
10. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UM, Zhong X, Koenig SS, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med*. 2012; 4(132):132ra52.
11. Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc Natl Acad Sci U S A*. 2013; 110(31):12804–12809. [PubMed: 23858463]
12. Chaffron S, Rehrauer H, Pernthaler J, von Mering C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res*. 2010; 20(7):947–959. [PubMed: 20458099]
13. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol*. 2012; 8(9):e1002687. [PubMed: 23028285]
14. Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics*. 2006; 22(20):2532–2538. [PubMed: 16882654]

15. Stein RR, Bucci V, Toussaint NC, Buffie CG, Räscher G, Pamer EG, Sander C, Xavier JB. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol*. 2013; 9(12):e1003388. [PubMed: 24348232]
16. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, Stahl DA. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol*. 2007; 3:92. [PubMed: 17353934]
17. Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, Somerfield P, Fuhrman JA, Field D. Defining seasonal marine microbial community dynamics. *ISME J*. 2012; 6(2):298–308. [PubMed: 21850055]
18. Trosvik P, Rudi K, Naes T, Kohler A, Chan KS, Jakobsen KS, Stenseth NC. Characterizing mixed microbial population dynamics using time-series analysis. *ISME J*. 2008; 2(7):707–715. [PubMed: 18385770]
19. Henze M, Grady CPL Jr, Gujer W, Marais GVR, Matsuo T. A general model for single-sludge wastewater treatment systems. *Water Res*. 1987; 21(5):505–515.
20. Paulson JN, Stine OC, Bravo H, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods*. 2013; 10(12):1200–1202. [PubMed: 24076764]
21. Bacaër, N. A short history of mathematical population dynamics. Springer-Verlag; London: 2011.
22. Sell, G., You, Y. Dynamics of evolutionary equations. 1. Springer-Verlag; New York: 2002.
23. Takeuchi, Y. Global dynamical properties of Lotka-Volterra systems. World Scientific Publishing; Singapore: 1996.
24. Poyton A, Varziri M, McAuley K, McLellan P, Ramsay J. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & Chemical Engineering*. 2006; 30(4):698–708.
25. Ramsay J. Principal differential analysis: data reduction by differential operators. *J R Stat Soc Series B Methodol*. 1996; 58(3):495–508.
26. Nocedal, J., Wright, SJ. Numerical optimization. 2. Springer-Verlag; New York: 2006.
27. Calvetti, D., Somersalo, E. An introduction to Bayesian scientific computing. 1. Springer-Verlag; New York: 2007.
28. Aster, R., Borchers, B., Thurber, C. Parameter estimation and inverse problems. 2. Academic Press; Waltham (MA): 2012.
29. Rodriguez-Fernandez M, Mendes P, Banga JR. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*. 2006; 83(2):248–265. [PubMed: 16236429]
30. Kostina E. Robust parameter estimation in dynamic systems. *Optimization and Engineering*. 2004; 5(4):461–484.
31. Stoer, J., Bulirsch, R. Introduction to numerical analysis. 3. Springer-Verlag; New York: 2002.
32. Bandara S, Schlöder J, Eils R, Bock H, Meyer T. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput Biol*. 2009; 5(11):e1000558. [PubMed: 19911077]
33. Hairer, G., Nørsett, S., Wanner, E. Solving ordinary differential equations I: nonstiff problems. 2. Springer-Verlag; Berlin: 1993.
34. Hairer, G., Wanner, E. Solving ordinary differential equations II: stiff and differential-algebraic problems. 2. Springer-Verlag; Berlin: 1996.
35. Conrad M, Hubold C, Fischer B, Peters A. Modeling the hypothalamus-pituitary-adrenal system: homeostasis by interacting positive and negative feedback. *J Biol Phys*. 2009; 35(2):149–162. [PubMed: 19669558]
36. Kirkpatrick S, Gelatt C Jr, Vecchi M. Optimization by simulated annealing. *Science*. 1983; 220(4598):671–680. [PubMed: 17813860]
37. Simon, D. Evolutionary optimization algorithms. John Wiley & Sons; Hoboken (NJ): 2013.
38. Kennedy J, Eberhart R. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on*. 1995; 4:1942–1948.
39. Bulirsch R. Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung. Report der Carl-Cranz-Gesellschaft. 251

40. Bock, HG. Numerical Treatment of Inverse Problems in Differential and Integral Equations. Vol. 2. Birkenhäuser, Boston: 1983. p. 95-121.Ch. Recent Advances in Parameteridentification Techniques for O.D.E
41. Bock, H., Plitt, K. A multiple shooting algorithm for direct solution of optimal control problems; Proceedings., 9th IFAC World Congress; 1984. p. 243-247.
42. Goel G, Chou I-C, Voit EO. System estimation from metabolic time-series data. *Bioinformatics*. 2008; 24(21):2505–2511. [PubMed: 18772153]
43. Baake E, Baake M, Bock HG, Briggs KM. Fitting ordinary differential equations to chaotic data. *Physical Review A*. 1992; 45(8):5524.
44. Chou I-C, Voit EO. Estimation of dynamic flux profiles from metabolic time series data. *BMC Systems Biology*. 2012; 6(1):84. [PubMed: 22776140]
45. Brendel M, Bonvin D, Marquardt W. Incremental identification of kinetic models for homogeneous reaction systems. *Chemical Engineering Science*. 2006; 61(16):5404–5420.
46. Vercammen D, Logist F, Van Impe J. Dynamic estimation of specific fluxes in metabolic networks using non-linear dynamic optimization. *BMC Systems Biology*. 2014; 8(1):132. [PubMed: 25466625]
47. Kallrath, J., Schlöder, JP., Bock, HG. Qualitative and Quantitative Behaviour of Planetary Systems. Springer; New York: 1993. p. 353-371.Ch. Least squares parameter estimation in chaotic differential equations
48. Vasantharajan S, Biegler LT. Simultaneous strategies for optimization of differential-algebraic systems with enforcement of error criteria. *Computers & Chemical Engineering*. 1990; 14(10): 1083–1100.
49. Tjoa I, Biegler L. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Computers and Chemical Engineering*. 1991; 15(10):679–690.
50. Voit, EO. A First Course in Systems Biology. Garland Science; New York: 2013.
51. Liu Y, Gunawan R. Parameter estimation of dynamic biological network models using integrated fluxes. *BMC Syst Biol*. 8(127)
52. Jia G, Stephanopoulos G, Gunawan R. Incremental parameter estimation of kinetic metabolic network models. *BMC Syst Biol*. 6(142)
53. Hanke M. On a least-squares collocation method for linear differential-algebraic equations. *Numerische Mathematik*. 1988; 54(1):79–90.
54. Davis ME. Numerical methods and modeling for chemical engineers. Courier Corporation. 2013
55. Banks HT, Kareiva PM, Murphy KA. Parameter estimation techniques for interaction and redistribution models: a predator-prey example. *Oecologia*. 1987; 74(3):356–362. [PubMed: 28312473]
56. Banks H, Crowley J, Kunisch K. Cubic spline approximation techniques for parameter estimation in distributed systems. *IEEE Transactions on Automatic Control*. 1983; 28(7):773–786.
57. Chung M, Haber E. Experimental design for biological systems. *SIAM J. Control and Optim.* 2012; 50(1):471–489.
58. Bochev, PB., Gunzburger, MD. Least-squares finite element methods. Vol. 166. Springer; New York: 2009.
59. Vogel, CR. Computational methods for inverse problems. Society for Industrial and Applied Mathematics; Philadelphia: 2002.
60. Powell, MJD. Approximation theory and methods. Cambridge University Press; Cambridge (UK): 1981.
61. Hadamard, J. Lectures on Cauchy's Problem in Linear Differential Equations. Yale University Press; New Haven: 1923.
62. Hansen, PC. Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. SIAM; Philadelphia: 1998.
63. Mhamdi, A., Marquardt, W. An inversion approach to the estimation of reaction rates in chemical reactors; Control Conference (ECC), 1999 European, IEEE; 1999. p. 3041-3046.
64. Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*. 1979; 21(2):215–223.

65. Geisser S, WF E. A predictive approach to model selection. J Am Stat Assoc. 1979; 74(365):153–160.
66. Chung M, Long Q, Johnson BA. A tutorial on rank-based coefficient estimation for censored data in small-and large-scale problems. Statistics and Computing. 2013; 23(5):601–614. [PubMed: 23956500]
67. Refaeilzadeh, P., Tang, L., Liu, H. Cross-validation, in: Encyclopedia of Database Systems. Springer; New York: 2009. p. 532-538.
68. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MR, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature. 2015; 517(7533):205–208. [PubMed: 25337874]
69. Verstraelen H, Verhelst R, Claeys G, De Backer E, Temmerman M, Vaneechoutte M. Longitudinal analysis of the vaginal microflora in pregnancy suggests that *L. crispatus* promotes the stability of the normal vaginal microflora and that *L. gasseri* and/or *L. iners* are more conducive to the occurrence of abnormal vaginal microflora. BMC Microbiol. 9(116)
70. de Boor, C. A practical guide to splines. 1. Springer-Verlag; New York: 2001.

Highlights

- A robust parameter estimation method for biological systems is proposed.
- The method can include prior knowledge of and assumptions on biological systems.
- Example applications show the usefulness of using appropriate inference methods.
- Insights from this method can guide future experiments studying biological systems.

Figure 1. Extended numerical solution of a Lotka-Volterra system. The plotted solution is from time $t = 0$ to $t = 200$ and demonstrates the chaotic behavior of the system

Figure 2. Numerical solution of a Lotka-Volterra system and data points for simulation studies. The black curves indicate the numerical solutions used to generate the data. The blue dots are the data with no multiplicative noise. The red dots are the data with up to 10 percent multiplicative noise, and the yellow dots are the data with up to 25 percent multiplicative noise

Figure 3. Absolute errors of optimal model parameters and initial conditions. The images show the absolute errors in the recovery of the model parameters and initial conditions for all three simulation studies

Figure 4.

Relative error between spline solutions and state solutions. Each row shows the relative error for a single state for all three simulation studies. The relative error between the spline

solution and state solution is calculated as $\frac{s - y}{y}$ over the time interval $[0, 10]$ with s denoting the optimal spline approximation and y denoting the numerical solution of the optimal Lotka-Volterra system

Figure 5.

Cumulative local sensitivities of optimized model parameters and initial conditions. The images display the cumulative local sensitivities for every optimal model parameter and initial condition for all three simulation studies

Figure 6.

Relative error between spline solutions and state solutions. Each row shows the relative error for a single OTU for all three replicate experiments. The relative error between the spline

solution and state solution is calculated as $\frac{s - y}{y}$ over the time interval $[1, 21]$ with s denoting the optimal spline approximation and y denoting the numerical solution of the optimal Lotka-Volterra system. Time is measured in days

Figure 7.

Comparison of state solutions. Each row compares the state solutions for a single OTU for all three replicate experiments using our parameter estimation results and those found in Stein et al. [15]. The black dots indicate the provided experimental data. The blue curves mark the state solutions found using our optimal Lotka-Volterra system, and the red curves denote the state solutions found using the optimal Lotka-Volterra system published by Stein et al. Time is measured in days, and abundance is measured in 10^{11} DNA copies per cubic centimeter

Figure 8.

Comparison of interaction matrices. This figure compares (A) the interaction matrix from our optimal Lotka-Volterra system to (B) a subset of the interaction matrix published by Stein et al. [15]. In the graphs, entry a_{ij} in the interaction matrix is given as a directed edge from node j to node i . The value of the entry a_{ij} is given by the color of the edge. OTUs: 1- *Blautia*, 2- *Barnesiella*, 3- *Unclassified Mollicutes*, 4- *Undefined Lachnospiraceae*, 5- *Unclassified Lachnospiraceae*, 6- *Coprobacillus*, and 7- *Other*

Figure 9.

Cumulative local sensitivities of optimized model parameters and initial conditions. The images display the cumulative local sensitivities for every optimal model parameter and initial condition for all three replicate experiments

Figure 10.

Relative error between spline solutions and state solutions. Each plot shows the relative error for a single OTU. The relative error between the spline solution and state solution is

calculated as $\frac{s - y}{y}$ over the time interval $[1, 110]$ with s denoting the optimal spline approximation and y denoting the numerical solution of the optimal Lotka-Volterra system. Time is measured in days

Figure 11.

Comparison of state solution and data. Each plot shows the state solution using the optimal Lotka-Volterra system for a single OTU against the provided experimental data. The curves mark the state solutions and the black dots indicate the provided experimental data. Time is measured in days, and the abundance is the number of 16S rRNA gene sequence reads

Math Biosci. Author manuscript; available in PMC 2018 December 01.

Figure 13.

Cumulative local sensitivities of optimized model parameters and initial conditions. The images display the cumulative local sensitivities for every optimal model parameter and initial condition for subject 15

Figure 14.

Extended state solution. The images display the state solution of the optimal Lotka-Volterra model up to 500 days