

Trend and fluctuations: Analysis and design of population dynamics measurements in replicate ecosystems

Doeke R. Hekstra,^{1,*} Simona Cocco,^{2,3} Remi Monasson,^{2,4} and Stanislas Leibler^{1,2}

¹*Center for Studies in Physics and Biology and the Laboratory of Living Matter, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA*

²*School of Natural Sciences, and The Simons Center for Systems Biology, The Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA*

³*Laboratoire de Physique Statistique de l'École Normale Supérieure, 24, Rue Lhomond, 75231 Paris Cedex 05, France*

⁴*Laboratoire de Physique Théorique de l'École Normale Supérieure, 24, Rue Lhomond, 75231 Paris Cedex 05, France*

(Received 12 August 2013; published 16 December 2013)

The dynamical evolution of complex systems is often intrinsically stochastic and subject to external random forces. In order to study the resulting variability in dynamics, it is essential to make measurements on replicate systems and to separate arbitrary variation of the average dynamics of these replicates from fluctuations *around* the average dynamics. Here we do so for population time-series data from replicate ecosystems sharing a common average dynamics or common trend. We explain how model parameters, including the effective interactions between species and dynamical noise, can be estimated from the data and how replication reduces errors in these estimates. For this, it is essential that the model can fit a variety of average dynamics. We then show how one can judge the quality of a model, compare alternate models, and determine which combinations of parameters are poorly determined by the data. In addition we show how replicate population dynamics experiments could be designed to optimize the acquired information of interest about the systems. Our approach is illustrated on a set of time series gathered from replicate microbial closed ecosystems.

DOI: [10.1103/PhysRevE.88.062714](https://doi.org/10.1103/PhysRevE.88.062714)

PACS number(s): 87.23.-n, 02.50.Tt, 45.30.+s, 89.75.-k

I. INTRODUCTION

Complex systems in physics and other fields, such as biology and sociology, are often characterized by intrinsically stochastic dynamics and subject to external random forces. At the same time, nontrivial trends are pervasive in those systems, be it in the laboratory, society, or nature. This raises a basic question when studying the variations in the activities or numbers of the system components. How can one separate the contribution of the trend to the observed dynamical evolution from the contribution of the various sources of noise? Traditionally, and mostly for unique time series, such a separation is “achieved” by using models with constant coefficients, which allow for only a few types of trends, such as constant average, logistic or exponential growth, or oscillations.

In reality, however, trends can have much more sophisticated time dependence, and deviations between the true average dynamics and the average dynamics imposed by constant-coefficient trend models will have the appearance of fluctuations around the inferred average dynamics. This will confound the analysis of true variability of the dynamics. The most direct approach to answering the question above is to replicate the experiment many times, with identical initial and experimental conditions. The average over the different replicate dynamics naturally defines the common trend of the system, while replicate-dependent deviations correspond to specific realizations of the stochastic fluctuations. Obviously, in practice, this approach is hindered by various limitations,

such as the errors in the measurements, the limited number of available replicates, and limited sampling. To which extent those limitations affect the determination of the common trend and the characterization of the statistical features of the fluctuations is a matter of concern.

The purpose of this paper is to address this general question, with a focus on the specific case of ecological systems. The dynamics of ecological systems are by their very nature stochastic, as a result of, for instance, random genetic and phenotypic changes and randomness in the timing of birth and death. External forces, such as the weather and the random immigration of other species, play crucial roles, too. Factors such as seasonal variation, depletion of nutrients, detailed chemistry, and any other factor not explicitly modeled may affect the average dynamics in various idiosyncratic ways, and their interplay can produce nontrivial average dynamics (a common trend). While replication of ecosystems is already fairly common, model ecosystems permitting excellent control of starting and experimental conditions are becoming increasingly available. Examples include the development of replicate closed ecosystems [1–3], multispecies chemostats [4–6], and gnotobiotic mouse experiments [7,8], which have been revolutionized by modern sequencing methods.

By providing multiple time series, experimental replication allows one to better separate temporal variation of the average density dynamics from density fluctuations around the average. However, to realize this advantage, one needs to be able to fit simple models to the data which are flexible enough to accommodate arbitrary average dynamics.

Here, we develop a method for the statistical analysis of population dynamics measurements for replicate ecosystems. Our approach is based on a combination of methods and ideas coming from the study of stochastic processes in statistical

*Present address: Green Center for Systems Biology, University of Texas Southwestern Medical Center, 6001 Forest Park Road, Dallas, TX 75390, USA; doeke.hekstra@utsouthwestern.edu

physics and probability theory, as well as from statistical Bayesian inference. A crucial point in our approach is the distinction between “measurement error” and “dynamical noise,” which both contribute to the observed density fluctuations around the average dynamics. *Measurement error* comprises both the effects of limited sampling and the possibility that the measured populations do not coincide with the dynamically relevant ones [9]. For example, one might observe individuals in only one life-cycle stage (e.g., adults) or in only one part of their habitat. *Dynamical noise* affects the population dynamics itself and can arise from endogenous sources, such as randomness in the number and timing of births and deaths or in behavior, and exogenous sources, such as fluctuations in the weather or the timing and nature of individuals entering the system of interest. Apparent dynamical noise can also result from more complex dynamics (e.g., chaos or interactions with unobserved species) not adequately captured by the deterministic part of a chosen model. Collectively, we will refer to measurement error and dynamical noise as “noise sources.”

To separate the effects of measurement error and dynamical noise, our approach is based on a *state space model* [10], described by two types of equations: (1) one or more so-called state equations which describe the system dynamics in terms of its true, underlying variables, say, $\mathbf{x} = (x_1, \dots, x_N)$ (bold symbols denote vectors and matrices), model parameters, say, θ , and dynamical noise, say, η , and (2) an “observation equation” describing the relation between the observed variables, say, \mathbf{y} , the underlying variables, \mathbf{x} , and measurement error, ξ . State space models have indeed become fairly commonplace in ecology [11–14].

Such a model defines the probability $P(\mathbf{y}|\theta)$ of the data given the model parameters. To infer the values of the parameters from the observed data we introduce the *posterior probability distribution* of the model parameters given the data, $P(\theta|\mathbf{y}) = P(\mathbf{y}|\theta) \cdot P(\theta)/P(\mathbf{y})$ by Bayes’s theorem. Here $P(\theta)$ is the so-called *prior probability distribution* of the model parameters, which specifies what is known *a priori* about the model parameters. $P(\mathbf{y})$ is the probability of the data under the model integrated over all possible values of the model parameters. Maximizing the unnormalized posterior probability, $P(\mathbf{y}|\theta) \cdot P(\theta)$, is thus equivalent to finding the most likely set of parameters given the data.

The use of state space models poses, however, a difficulty, since one needs to integrate over the possible values of the underlying (unobserved) variables, \mathbf{x} , to evaluate $P(\mathbf{y}|\theta) = \int d\mathbf{x} (P(\mathbf{y}|\mathbf{x}, \theta)P(\mathbf{x}|\theta))$. This integration is commonly done by so-called Markov-chain Monte Carlo methods (e.g., in WINBUGS [15]). Such methods essentially explore the “product space” of underlying variables and model parameters, $X \times \Theta$, by artificial Monte Carlo dynamics dependent on $P(\mathbf{x}, \mathbf{y}|\theta) \cdot P(\theta)$, and obtain estimates of the posterior distribution, $P(\theta|\mathbf{y})$. Alternatively, such integration can be performed using forward and backward recursion relations, known as filters and smoothers, which propagate estimates of the probability distribution of the underlying variables, \mathbf{x} , as a function of the data. This integration is then alternated with optimization over the model parameters. A rigorous example of such an approach is expectation maximization (see Ref. [16] for a nice introduction). For linear, Gaussian, discrete-time models,

the forward and backward recursions reduce to the so-called Kalman filter (for forward recursion only [17]) or Kalman smoother (forward and backward recursion, also known as the Rauch-Tung-Striebel smoother [18]). In ecology, approximate methods are more common, in which the likelihood function is simplified and, most commonly, the Kalman filter is used (for example, restricted maximum likelihood [14] and conditional maximum likelihood [19]).

Hereafter, we choose to study a model that is linear in (the logarithms of) the underlying variables and has a Gaussian stochastic structure. This permits explicit integration over underlying variables and an analytical expression for the posterior density of the model parameters. Because of the availability of the methods mentioned, explicit calculation of the likelihood function is not often used. However, much like a partition function in statistical mechanics, a likelihood function provides a complete description of a statistical problem and can be used to illuminate any aspect of data analysis and experimental design. We show that explicit calculation of the posterior distribution is feasible for the common trend problem, which contains *two* layers of underlying variables: the true population densities in each replicate ecosystem and the common trend.

While there are arguably many nonlinearities in ecology, a linear model often provides a reasonable starting point. First, many of the questions addressed here in fitting a linear model to the data will be encountered *a fortiori* for nonlinear models. Second, a linear model can often be considered a linearization of a nonlinear model [19] and as such as a reasonable “null hypothesis” for the importance of nonlinear effects. Often, for complex systems, there is conceptual support for one kind or another of nonlinear effect contributing to the dynamics. The heterogeneity and multitude of scales present in complex systems can easily make a judgment of their relative importance impossible. Polansky *et al.* [20], for example, found no *statistical* support for the inclusion of nonlinearities in the dependence of per capita growth rate on density in *any* of 25 data sets examined. Third, for our data (below), a linear, Gaussian model with a time-dependent coefficient describes the data well, as we anticipated based on previous work [1].

Linear state space models with constant coefficients have received considerable attention both within and outside of ecology. An ecological example is the “stochastic exponential,” or “random-walk,” model, which does not contain a density-dependent term in its dynamics (Refs. [9,11,21,22] for discrete time and Ref. [23] for the continuous time case). Another common linear, Gaussian model is the so-called Gompertz model, which accommodates density dependence. The discrete-time, constant-coefficient Gompertz model often has been used to study ecology (Refs. [14,24,25] for single species and Ref. [19] for multiple species). Here we will examine a continuous-time Gompertz model with a time-dependent per-species term and a time-independent species-species interaction matrix.

The paper is organized as follows. To allow quick perusal, we begin both the Methods and Results sections with a brief overview. Readers not interested in the mathematical details of our methods can restrict themselves to reading the overview of the Methods section. In the Methods section, we formulate the likelihood function in detail for replicate time series modeled

using the continuous-time Gompertz model and describe the integration over underlying variables and maximization over model parameters. This allows us to determine maximum likelihood parameter estimates and the joint (multivariate) posterior distribution of the model parameters. We will apply our methods to time series we recently acquired from replicate closed ecosystems (CES, Ref. [1]). These experiments are briefly summarized at the end of the Methods section.

In the Results section we validate our method and discuss how one can estimate parameters, determine which parameter combinations are hard to estimate, compare models, and optimize experimental design.

Applied to our closed ecosystem data, we show that (1) a model with a time-dependent term is more suitable to describe variability in ecological dynamics than is a constant-coefficient approach; (2) the covariation between species in these closed ecosystems does not reflect the structure of instantaneous fluctuations in densities but rather the collective response of the interacting species to such fluctuations; (3) the net effective interactions between species nearly vanish for one linear combination of densities, that is, for one system dimension, or “ecomode,” the ecosystems lack a significant “restoring force”; and (4) this, in turn, causes the dynamics of each species in individual ecosystems to resemble a random walk around the common trend.

II. METHODS

A. Overview

In the Methods section, we will describe the first step in the statistical analysis of replicate time series from stochastic systems subject to a common trend: calculation of the likelihood function. We will do so in the language of ecosystems with interacting species, but similar equations likely apply to many classes of complex systems. The approach is very similar whether one uses frequentist or Bayesian statistics. We will use the Gompertz model, a linear model of population dynamics. We will formulate the dynamical model and assumptions about measurement error and provide a physical intuition for the way the common trend is described (Sec. II B). We formulate the likelihood function and describe how we will address the two layers of unobserved values: of the dynamics in each individual ecosystem and of the common trend shared by the replicate systems (Sec. II C). To control the degree to which the common trend description can vary with time, we introduce a smoothing parameter, μ . The structure of the likelihood function allows us to integrate over the unobserved variables: that is, we can consider all possible values at once (Sec. II D). This greatly reduces the number of parameters that need to be estimated. Calculation of the likelihood function over the remaining parameters exploits a number of parameter transformations (Sec. II E) to make the likelihood function close to Gaussian. This final likelihood surface can be described by the location of its maximum and the Hessian matrix, which describes how steeply the likelihood function is peaked around the maximum in each direction. We will examine this matrix in detail in the Results section. Finally (Sec. II F), we briefly summarize closed ecosystem experiments.

1. Notation

We consider a general experiment of duration T , in which population densities (or numbers) N^{obs} are observed for K species, with index $k = 1, 2, \dots, K$. These densities are observed in S systems with index $s = 1, 2, \dots, S$, at N time points, $t_n \in [0, T]$, with $n = 1, 2, \dots, N$ the index over time points. The time points do *not* need to be regularly spaced. We start by transforming to logarithmic densities: $y_{kns} = \log N_{ks}^{\text{obs}}(t_n)$, indexed by species k , system s , and time point n . The general case in which densities may have been measured in different ecosystems at different times is described in the Supplemental Material [26], Sec. S3.

For simplicity, we will suppress indices that are implied. For example, \mathbf{y}_k refers to all observations of $\log N^{\text{obs}}$ for a given species k , an $NS \times 1$ vector; \mathbf{y}_s refers to all observations for a given system, s , a $KN \times 1$ vector; and \mathbf{y}_n refers to all observations at a given time point, t_n , a $KS \times 1$ vector. Unless specified otherwise, sums run over their entire possible range.

B. A state space model

We will make a few basic assumptions. First, we assume that the measurements do not affect the dynamics. Second, we assume that the dynamics are Markovian. That is, we assume that population density changes are a function of the present state of the ecosystem and not (also) of past states.

These assumptions lead to a state space approach in which the dynamics are separated into two parts: a state equation describing the (stochastic) dynamics of the system itself in terms of “true” underlying logarithmic densities, $\mathbf{x}(t) = \log N^{\text{dyn}}(t)$ [Eq. (1a)], and an observation equation describing the relation between observed and true logarithmic densities [Eq. (1b)],

$$\dot{\mathbf{x}}_s = \mathbf{g}(\mathbf{x}_s) + \boldsymbol{\eta}_s, \quad (1a)$$

$$\mathbf{y}_{ns} = \mathbf{h}(\mathbf{x}_s(t_n)) + \boldsymbol{\xi}_{ns}. \quad (1b)$$

Here $\mathbf{g}(\mathbf{x})$ represents the system dynamics and $\boldsymbol{\eta}_s$ additive dynamical noise with (instantaneous) covariance matrix $\mathbf{S}^d dt$. We further assume that the dynamical noise is Gaussian and uncorrelated over time. Such dynamical noise in the dynamics is a realistic assumption for our data [1]. For the description of measurements, we will assume $\mathbf{h}(\mathbf{x}) = \mathbf{x}$ and additive Gaussian measurement error, $\boldsymbol{\xi}$. This corresponds to log-normal measurement error in the observed densities, which seems appropriate for experiments in which one likely makes proportional errors in estimating the underlying density. We treat the variance of measurement error of each species, S_k^m , as an unknown model parameter and assume measurement errors are independent between species. That is, we assume the covariance matrix for measurement errors, \mathbf{S}^m , to be a diagonal matrix. It is straightforward to derive the general case instead.

As an example of linear dynamics, $\mathbf{g}(\mathbf{x})$, we consider a Gompertz model [27], in which

$$\mathbf{g}(\mathbf{x}) = \mathbf{f}(t) - \mathbf{A} \cdot \mathbf{x}, \quad (2)$$

where we assume $\mathbf{f}(t)$ to depend on time but \mathbf{A} to be time independent.

The dynamical model in Eqs. (1a) and (2) is a “driven” version of the Ornstein-Uhlenbeck process [28] and is formally analogous to that of a set of S particles (with indices s) moving in K dimensions in the presence of a harmonic (quadratic) potential energy well with multivariate spring constant \mathbf{A} and center position $\mathbf{x}^*(t) = \mathbf{A}^{-1} \mathbf{f}(t)$ under overdamped conditions (that is, viscous forces instantaneously oppose any other forces). In this analogy, each particle represents an ecosystem. Each particle will tend to track the well (which itself can move with time) and experiences thermal fluctuations, η_s . If the eigenvalues of \mathbf{A} are large and positive, the average position of each particle will closely match that of the well, i.e., $\langle \mathbf{x}(t) \rangle \approx \mathbf{x}^*(t) = \mathbf{A}^{-1} \mathbf{f}(t)$. If, in contrast, the eigenvalues of \mathbf{A} are small (a shallow well), the well position $\mathbf{x}^*(t)$ at any instant becomes unimportant. Instead, $\langle \mathbf{x}(t) \rangle \approx \langle \mathbf{x}(0) \rangle + \int_0^t dt' \mathbf{f}(t')$, and individual time series are approximately random walks around $\langle \mathbf{x}(t) \rangle$.

The analogy has its limitations, since the eigenvalues of \mathbf{A} can be complex and its eigenvectors need not be orthogonal. An extensive ecological interpretation of this “particle in a well” picture of ecological dynamics is given by Ives *et al.* [19,29] for the case of the function \mathbf{f} being constant.

To fix terminology, we will call the matrix \mathbf{A} the “effective interaction matrix” because it describes the net effects that species densities exert on their own and other species’ growth rates. We call the expected replicate-average dynamics, $\langle \mathbf{x}(t) \rangle$, the “common trend.” As outlined above, the function $\mathbf{f}(t)$ can dictate the course of this common trend, in principle, in arbitrary ways. The time dependence of $\mathbf{f}(t)$ gives the model its flexibility. Because of this, we will call $\mathbf{f}(t)$ the “trend function.” Since population density measurements mostly occur at discrete time points, we will switch to a representation of $\mathbf{f}(t)$ by a discrete set of “trend variables.” We refer to these as trend *variables* rather than as trend parameters because we will treat the trend and the dynamics of individual replicates as two layers of unobserved dynamic quantities (variables). Fluctuations, then, are the deviations of the dynamics of individual replicates from the common trend.

C. Formulation of the likelihood function

Our first aim is to estimate the model parameters, $\theta = (\mathbf{A}, \mathbf{S}^d, \mathbf{S}^m)$, from the data, \mathbf{y} . From a classical perspective, one typically seeks to derive an estimator, such as a *maximum likelihood* (ML) estimator, for the model parameters. The likelihood is simply defined as

$$l_y(\theta) \equiv P(\mathbf{y}|\theta), \quad (3)$$

that is, the probability of the data given the model parameters. The maximum likelihood estimate maximizes $l_y(\theta)$. In *Bayesian statistics*, one estimates $P(\theta|\mathbf{y})$, the conditional probability of the model parameters given the data, also known as the *posterior distribution* of θ . By Bayes’s theorem,

$$P(\theta|\mathbf{y}) = P(\mathbf{y}|\theta) \cdot P(\theta) / P(\mathbf{y}) \propto P(\mathbf{y}|\theta) \cdot P(\theta), \quad (4)$$

where $P(\theta)$ is the so-called *prior distribution* of the model parameters and contains any constraints we put on our estimation *a priori*, based on prior knowledge or limitations imposed by the experimental design. $P(\mathbf{y})$ is a model-dependent, but parameter-independent, normalization factor. We will refer

to the product $P(\mathbf{y}|\theta) \cdot P(\theta)$ as the unnormalized posterior distribution of θ .

It is obvious that the likelihood function, $l_y(\theta) \equiv P(\mathbf{y}|\theta)$, plays a central role in parameter estimation in either perspective. To develop the likelihood function, however, one needs to account not just for the data, \mathbf{y} , and the model parameters, θ , but also for the two types of underlying variables, the true logarithmic densities, \mathbf{x} , and the unknown common trend, described by \mathbf{f} . Central to our approach is that we explicitly integrate $l_y(\theta)$ over the underlying variables. That is,

$$l_y(\theta) = \int d\mathbf{f} \int d\mathbf{x} (P(\mathbf{y}|\mathbf{x}, \theta) \cdot P(\mathbf{x}|\mathbf{f}, \theta) \cdot P(\mathbf{f}|\theta)). \quad (5)$$

We will first obtain expressions for each of the three factors in the integrand of this expression. Then we integrate over the two types of underlying variables, \mathbf{x} and \mathbf{f} , successively.

First, $P(\mathbf{y}|\mathbf{x}, \theta) = \prod_{n,s} P(y_{ns} | x_{ns}, \theta)$, with each $P(y_{ns} | x_{ns}, \theta) \sim \mathcal{N}(x_{ns}, \mathbf{S}^m)$, a multivariate normal distribution. We consider dynamics without memory in $\dot{\mathbf{x}}(t)$, as mentioned above. Hence, $P(\mathbf{x}|\mathbf{f}, \theta) = \prod_{s,n < N} P(x_{n+1,s} | x_{ns}, \mathbf{f}, \theta)$. We obtain these dynamical transition probabilities, $P(x_{n+1,s} | x_{ns}, \mathbf{f}, \theta)$, by integration of Eq. (1a) over a time interval $\Delta t_n = (t_{n+1} - t_n)$ [30]. (Note that we omit a prior distribution on \mathbf{x}_1 , although this would be trivial to include.) For the Gompertz model the transition probabilities are Gaussian, with expectation value as follows:

$$\begin{aligned} \langle \mathbf{x}_{n+1,s} | \mathbf{x}_{ns}, \mathbf{f}, \theta \rangle &= \mathbf{B}_n \cdot \mathbf{x}_{ns} + \boldsymbol{\phi}_n \Delta t_n \quad \text{with} \\ \mathbf{B}_n &= e^{-\mathbf{A} \Delta t_n} \quad \text{and} \\ \boldsymbol{\phi}_n &= \frac{1}{\Delta t_n} \int_{t_n}^{t_{n+1}} e^{-\mathbf{A}(t_{n+1}-t')} \mathbf{f}(t') dt' \end{aligned} \quad (6)$$

and covariance matrix

$$\Gamma_n \equiv \text{Cov}(\mathbf{x}_{n+1,s} | \mathbf{x}_{ns}, \theta) = \int_{t_n}^{t_{n+1}} e^{-\mathbf{A}(t_{n+1}-t')} \mathbf{S}^d e^{-\mathbf{A}^\top(t_{n+1}-t')} dt'. \quad (7)$$

Γ_n can be calculated analytically by so-called matrix fraction decomposition ([31], summarized in the Supplemental Material [26], Sec. S1.) Note that, in the limit of weak effective interactions ($A_{kl} \rightarrow 0$ for all k, l), one obtains $\Gamma_n \rightarrow \mathbf{S}^d \Delta t_n$ and $\boldsymbol{\phi}_n \rightarrow \mathbf{f}_n = \frac{1}{\Delta t_n} \int_{t_n}^{t_{n+1}} \mathbf{f}(t) dt$. From this point onward, we will take the $\boldsymbol{\phi}_n$, rather than $\mathbf{f}(t)$, as the variables describing the common trend.

A description of the common trend can take up many variables. To control the effective number of variables used, we introduce a smoothness condition in $P(\boldsymbol{\phi}|\theta) \equiv P(\mathbf{f}|\theta)$, which penalizes curvature (“flexibility”) in the common trend [32] (this is also known as “regularization”),

$$P(\boldsymbol{\phi}|\theta) \propto \exp \left(-\frac{\mu^2}{2} \sum_{k,n < N} \frac{(\phi_{k,n+1} - \phi_{kn})^2}{\Delta t_{n+1} + \Delta t_n} \right), \quad (8)$$

where μ is a smoothing parameter that determines the smoothness of the inferred common trend of each species. From a Bayesian perspective the smoothness condition can instead be considered a prior distribution for the unknown common trend. We examine the choice of μ in Sec. III D. A complete specification of the Bayesian unnormalized posterior distribution requires formulating prior distributions for the

TABLE I. Integration over \mathbf{x} . $\mathbf{Q}^{(s)}$, \mathbf{u}_s , $C^{(s)}$, and $C^{(0)}$ are defined in the description of Eq. (10) and $\mathbf{\Gamma}$ in Eq. (7). Here, we suppressed species indices. For example, $\mathbf{Q}_{nn'}^{(s)}$, α_n , and β_n are $K \times K$ (sub-)matrices, while each \mathbf{u}_{ns} is a $K \times 1$ vector. In addition, any transposes or inverses of single-index variables imply transposition or inversion of the indexed variable, not vice versa. That is, $\mathbf{\Gamma}_n^{-1} = (\mathbf{\Gamma}_n)^{-1}$, and not $(\mathbf{\Gamma}^{-1})_n$. Also, $\mathbf{y}_{ns}^\top = (\mathbf{y}_{ns})^\top$. Note that $\delta_{nn'} = 1$ for $n = n'$ and 0 otherwise. Likewise, $\mathbb{1}_{(\text{condition})} = 1$ if the condition is satisfied, and 0 otherwise.

$$\begin{aligned} \mathbf{Q}_{nn'}^{(s)} &= \delta_{nn'} \mathbb{1}_{(n>1)} \mathbf{\Gamma}_{n-1}^{-1} - \delta_{n-1,n'} \mathbb{1}_{(n>1)} \beta_{n-1} - \delta_{n+1,n'} \mathbb{1}_{(n<N)} \beta_n^\top + \delta_{nn'} \mathbb{1}_{(n<N)} \alpha_n + \delta_{nn'} (\mathbf{S}^m)^{-1} \\ \mathbf{u}_{ns} &= \mathbb{1}_{(n>1)} \mathbf{\Gamma}_{n-1}^{-1} \phi_{n-1} \Delta t_{n-1} - \mathbb{1}_{(n<N)} \beta_n^\top \phi_n \Delta t_n + (\mathbf{S}^m)^{-1} \mathbf{y}_{ns} \\ C^{(s)} &= \sum_{n=1}^{N-1} \Delta t_n \phi_n^\top \mathbf{\Gamma}_n^{-1} \phi_n \Delta t_n + \sum_{n=1}^N \mathbf{y}_{ns}^\top (\mathbf{S}^m)^{-1} \mathbf{y}_{ns} + \sum_{n=1}^{N-1} \log |\mathbf{\Gamma}_n| + N \log |\mathbf{S}^m| + K(2N-1) \log(2\pi) \\ C^{(0)} &= \mu^2 \sum_{n=1}^{N-1} \sum_{n'=1}^{N-1} ((\phi_n - \phi_n^0)^\top \mathbf{D}_{nn'} (\phi_{n'} - \phi_{n'}^0)) - \log |\mu^2 \mathbf{D}| + K(N-1) \log(2\pi) \\ \alpha_n &= \mathbf{B}_n^\top \mathbf{\Gamma}_n^{-1} \mathbf{B}_n \\ \beta_n &= \mathbf{\Gamma}_n^{-1} \mathbf{B}_n \end{aligned}$$

model parameters, θ . We discuss the specification of prior distributions in detail in the Supplemental Material [26], Sec. S2.

The full expression for the likelihood function is complicated, but both \mathbf{x} and ϕ appear solely as at most quadratic terms in exponential factors. One can thus integrate over the underlying variables and eliminate them from the likelihood function.

D. Integrating out underlying variables

We can now integrate the likelihood over the variables \mathbf{x} and ϕ , slightly modifying the proposal in Eq. (5) as follows:

$$l_y(\theta; \mu) = \int d\phi \int d\mathbf{x} (P(\mathbf{y}|\mathbf{x}, \theta) P(\mathbf{x}|\phi, \theta) P(\phi|\theta, \mu)). \quad (9)$$

To do so, we put $P(\mathbf{y}|\mathbf{x}, \theta) P(\mathbf{x}|\phi, \theta) P(\phi|\theta, \mu)$ in the form

$$\exp \left\{ -\frac{1}{2} \sum_s (\mathbf{x}_s^\top \mathbf{Q}^{(s)} \mathbf{x}_s - \mathbf{u}_s^\top \mathbf{x}_s - \mathbf{x}_s^\top \mathbf{u}_s + C^{(s)}) - \frac{1}{2} C^{(0)} \right\}, \quad (10)$$

by collecting terms proportional to $x_{ns} x_{n's}$ into $\mathbf{Q}_{nn'}^{(s)}$, proportional to x_{ns} into \mathbf{u}_{ns} , and constant terms into $C^{(s)}$ and $C^{(0)}$, respectively (each defined in Table I). As long as the measurement schedule is the same for all ecosystems, $\mathbf{Q}^{(s)} = \mathbf{Q}$ for all s . One can rewrite the term (10) as

$$\begin{aligned} \exp \left\{ -\frac{1}{2} \sum_s ((\mathbf{x}_s - \mathbf{Q}^{-1} \mathbf{u}_s)^\top \mathbf{Q} (\mathbf{x}_s - \mathbf{Q}^{-1} \mathbf{u}_s) \right. \\ \left. - \mathbf{u}_s^\top \mathbf{Q}^{-1} \mathbf{u}_s + C^{(s)}) - \frac{1}{2} C^{(0)} \right\}. \quad (11) \end{aligned}$$

TABLE II. Integration over ϕ . \mathbf{R} , \mathbf{v} , and C' are defined in the description of Eq. (13). The conventions used are the same as in Table I. Note that $\mathbf{Q}_{nn'}^{-1} = (\mathbf{Q}^{-1})_{nn'}$. Since ϕ is a $K(N-1) \times 1$ vector, \mathbf{v} is also a $K(N-1) \times 1$ vector and \mathbf{D} and \mathbf{R} are $K(N-1) \times K(N-1)$ matrices.

$$\begin{aligned} \mathbf{R}_{nn'} &= \mu^2 \mathbf{D}_{nn'} + \mathbb{1}_{(n<N)} \Delta t_n^2 S \mathbf{\Gamma}_n^{-1} \delta_{nn'} - \mathbb{1}_{(n<N)} \mathbb{1}_{(n'<N)} S \Delta t_n \Delta t_{n'} \times \{ \mathbf{\Gamma}_n^{-1} \mathbf{Q}_{n+1,n'+1}^{-1} \mathbf{\Gamma}_{n'}^{-1} - \mathbf{\Gamma}_n^{-1} \mathbf{Q}_{n+1,n'}^{-1} \beta_{n'}^\top - \beta_n \mathbf{Q}_{n,n'+1}^{-1} \mathbf{\Gamma}_{n'}^{-1} + \beta_n \mathbf{Q}_{nn'}^{-1} \beta_{n'}^\top \} \\ \mathbf{v}_n &= \mu^2 \sum_{n'=1}^{N-1} \mathbf{D}_{nn'} \phi_{n'}^0 + \mathbb{1}_{(n<N)} \Delta t_n \sum_{n',s} (\mathbf{\Gamma}_n^{-1} \mathbf{Q}_{n+1,n'}^{-1} - \beta_n \mathbf{Q}_{nn'}^{-1}) (\mathbf{S}^m)^{-1} \mathbf{y}_{n's} \\ C' &= \mu^2 \sum_{n=1}^{N-1} \sum_{n'=1}^{N-1} \phi_n^0 \mathbf{D}_{nn'} \phi_{n'}^0 + S \sum_{n=1}^{N-1} \log |\mathbf{\Gamma}_n| + SN \log |\mathbf{S}^m| + S \log |\mathbf{Q}| + \sum_{s=1}^S \sum_{n=1}^N \mathbf{y}_{ns}^\top (\mathbf{S}^m)^{-1} \mathbf{y}_{ns} \\ &\quad - \sum_{s=1}^S \sum_{n,n'=1}^N \mathbf{y}_{ns}^\top (\mathbf{S}^m)^{-1} \mathbf{Q}_{nn'}^{-1} (\mathbf{S}^m)^{-1} \mathbf{y}_{n's} - \log |\mu^2 \mathbf{D}| + K(S+1)(N-1) \log(2\pi) \\ \mathcal{L} &= -2 \log P(\mathbf{y}|\theta) = C' + \log |\mathbf{R}| - \mathbf{v}^\top \mathbf{R}^{-1} \mathbf{v} - K(N-1) \log(2\pi) \end{aligned}$$

We will repeatedly encounter such Gaussian integrals. We remind the reader that (for general \mathbf{z} and \mathbf{A}) $\int_{-\infty}^{\infty} d\mathbf{z} e^{-\frac{1}{2} \mathbf{z}^\top \mathbf{A}^{-1} \mathbf{z}} = \sqrt{|\mathbf{A}| (2\pi)^p}$, with $|\mathbf{A}|$ the determinant of \mathbf{A} and p the number of elements in \mathbf{z} . We perform the Gaussian integration over \mathbf{x} in Eq. (9). This leaves an integral over ϕ ,

$$l_y(\theta; \mu) = \int d\phi \exp \left\{ -\frac{1}{2} S (\log |\mathbf{Q}| - KN \log 2\pi) - \frac{1}{2} \left(C^{(0)} + \sum_s (C^{(s)} - \mathbf{u}_s^\top \mathbf{Q}^{-1} \mathbf{u}_s) \right) \right\}. \quad (12)$$

To perform the integration over the ϕ , we note that \mathbf{u}_s , $C^{(s)}$, and $C^{(0)}$ all depend on ϕ (see Table I). One can expand the exponent in Eq. (12) and collect terms proportional to $\phi_n \phi_{n'}$ in $\mathbf{R}_{nn'}$, to ϕ_n in \mathbf{v}_n , and constants into a term C' , completely analogously to the integration over \mathbf{x} (we summarize \mathbf{R} , \mathbf{v} , and C' in Table II). Integration of ϕ is Gaussian as well and yields the following:

$$l_y(\theta; \mu) = \exp \left\{ -\frac{1}{2} (C' - \mathbf{v}^\top \mathbf{R}^{-1} \mathbf{v} + \log |\mathbf{R}| - K(N-1) \log(2\pi)) \right\}. \quad (13)$$

For a Bayesian analysis, we simply multiply the likelihood function in Eq. (13) by the expressions for the prior distributions (see the Supplemental Material [26], Sec. S2, for a detailed description).

E. Likelihood maximization, parameter transformation, and the posterior distribution

We obtained maximum likelihood estimates of the model parameters by numerical optimization of the likelihood func-

tion [Eq. (13)] over the parameters θ . To be precise, we minimized $\mathcal{L}_y(\theta; \mu) \equiv -2 \log l_y(\theta; \mu)$ over a set of *transformed model parameters*, $\vartheta = \mathbf{w}(\theta)$, with \mathbf{w} a parameter transformation (see below). For simplicity, we will sometimes refer to \mathcal{L} itself as the log-likelihood. Minimization was performed using the Nelder-Mead algorithm (MATLAB 6.1, Mathworks, Inc.) and a conjugate gradient algorithm [33], with similar results (cf. Ref. [13]). We recommend examining the likelihood surface to verify a convex shape around the found minimum (as in Fig. S-4). For small data sets it may be helpful to use a stochastic minimization routine such as simulated annealing and/or multiple initial parameter guesses to avoid convergence to local minima.

Parameter transformations play two roles in this work. First, they allow unconstrained minimization. That is, while, for example, S_k^m would need to be constrained to be positive, this is not necessary for $\log S_k^m$. In addition to logarithmic transformation for the S_k^m , we used the so-called log-Cholesky transformation for the matrix \mathbf{S}^d to guarantee the constraint that it be positive definite [34].

The second purpose of parameter transformation is to guarantee a nearly parabolic log-likelihood surface around the likelihood maximum. A suitable choice of $\mathbf{w}(\theta)$ can often yield a nearly quadratic structure of $\mathcal{L}_y(\vartheta; \mu)$ around its minimum, $\hat{\vartheta}$ [35] (for our data, this is illustrated in Fig. S-4). Such a condition facilitates integration over model parameter space and allows maximum likelihood estimates to be as accurate as possible (by the Cramér-Rao inequality [36]). Such a transformation can be done *after* the optimal parameters have been found to make Gaussian approximations in subsequent analysis more accurate. For this purpose, we instead described \mathbf{S}^d by the set $\{\log(S_{kk}^d), S_{kl}^d\}$.

Alternative parameter transformations for this purpose have been proposed in Ref. [35]. More generally, classes of so-called normalizing transformations have been proposed [37], as well as closely related “variance-stabilizing transformations” [37,38].

1. Bayesian approach

In Bayesian statistics one can obtain a full description of the joint posterior probability of the model parameters. To do so, we first sought the minimum of

$$\mathcal{L}_y^B(\vartheta; \mu) \equiv \mathcal{L}_y(\vartheta; \mu) - 2 \log P(\vartheta)$$

as we did for $\mathcal{L}_y(\vartheta; \mu)$. Provided $\mathcal{L}_y^B(\vartheta; \mu)$ is approximately parabolic around its minimum, $\hat{\vartheta}$, one can approximate the unnormalized posterior probability of the parameters around $\hat{\vartheta}$ as

$$P(\mathbf{y}|\vartheta)P(\vartheta) \simeq \exp\left(-\frac{1}{2}\mathcal{L}_y^B(\hat{\vartheta}; \mu) - \frac{1}{2}(\vartheta - \hat{\vartheta})^\top \tilde{\mathbf{H}}(\vartheta - \hat{\vartheta})\right), \quad (14)$$

where $\tilde{\mathbf{H}}$ is the “Hessian matrix” of the log-likelihood function, describing how sharply \mathcal{L}_y^B is peaked around the optimal estimate. That is,

$$\tilde{H}_{ij} = \left. \frac{\partial^2 \mathcal{L}_y^B}{\partial \vartheta_i \partial \vartheta_j} \right|_{\hat{\vartheta}}. \quad (15)$$

In this approximation [Eq. (14)], it is also trivial to normalize the joint posterior distribution by Gaussian integration (which we will do explicitly in Sec. III F). We now obtain the posterior distribution for our original model parameters θ by transforming $P(\theta|\mathbf{y})d\theta = P(\vartheta|\mathbf{y})d\vartheta = P(\mathbf{w}(\theta)|\mathbf{y})d\mathbf{w}(\theta)$ or by constructing a histogram computationally by applying the inverse transformation $\mathbf{w}^{-1}(\vartheta)$ to samples drawn from $P(\vartheta|\mathbf{y})$. The Hessian matrices before and after model parameter transformation are related by

$$\mathbf{H} = \mathbf{J}^\top \tilde{\mathbf{H}} \mathbf{J}, \quad \text{with } J_{ij} = \frac{\partial \vartheta_i}{\partial \theta_j}. \quad (16)$$

We want to make a number of observations about these results. First, provided \mathcal{L}_y^B is nearly parabolic around $\hat{\vartheta}$, $\tilde{\mathbf{H}}^{-1}$ is an excellent approximation of the covariance matrix of the posterior distribution of ϑ , that is, the posterior covariance matrix (Sec. III E). In addition, the expected value of $\tilde{\mathbf{H}}$ is the so-called Fisher information matrix, which places a bound on how well any estimator can perform (see Sec. III G). In practice, we calculate the Hessian matrix using finite difference derivatives of the log-likelihood function, $\mathcal{L}_y(\vartheta; \mu)$ or $\mathcal{L}_y^B(\vartheta; \mu)$. Given the likelihood function in Table II, it is, in principle, possible to calculate the Hessian matrix exactly instead.

F. Closed ecosystem experiments

These experiments are described in detail in Ref. [1]. Our CESs contained three motile, unicellular organisms: the green alga *Chlamydomonas reinhardtii*, the bacterium *Escherichia coli*, and the ciliate *Tetrahymena thermophila*. Briefly, replicate CESs were constructed by inoculating the three species in 3 ml of a minimal medium with 0.03% proteose peptone (as the only source of carbon and nitrogen) and kept in small (~5-ml) cuvettes. We kept these CESs under constant light and temperature. We noninvasively measured densities in a small subvolume (~10 nl) in each cuvette by *in situ* fluorescence microscopy, acquiring a few thousand images per time point on measurement days. Measured local densities were proportional to the observed average number of individuals per image. We performed two independent experiments, in which densities were measured in sets of eight to nine replicate ecosystems at frequencies of 4–7 times per week (experiment 1) and 2, 1, $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{8}$ times per week (experiment 2) for each set of ecosystems. Sets of systems measured less frequently than once per 2 weeks were not included in the analysis presented here (retaining 35 replicate systems for the analysis presented here).

III. RESULTS

A. Overview

In this section, we outline the challenge posed by the statistical analysis and experimental design of time-series measurements on replicate stochastic systems (Sec. III B). We study the likelihood function derived in Eq. (13) and Tables I and II to address several aspects: First, we characterize the scaling of bias and variance of parameter estimates using simulated time series (Sec. III C). This analysis emphasizes that replication is especially valuable when one needs to

disentangle a common trend and fluctuations. We discuss how one should smoothen the trend: There is a basic tension between smoothing the average dynamics too much (essentially into a straight line) and too little (Sec. III D). Too much smoothing results in mistaking true temporal variations of the average dynamics for fluctuations. Too little results in interpretation of measurement errors as fluctuations. We propose and discuss results for two choices of intermediate smoothing. Measurements typically give different amounts of information on different parameters and parameter combinations (Sec. III E). We show how one can detect poorly constrained parameter combinations. We then address how one can judge the quality of models both in comparative terms (against other models) and in absolute terms (in terms of data fit) in Sec. III F. Doing so, we clarify a remarkably common situation in which parameter estimates and model comparison suggest conflicting interpretations of the data, the Lindley paradox. We then describe how one can improve experimental design with a focus on particular (combinations of) parameters (Sec. III G). Surprisingly, nonstandard experimental designs can in some cases lead to substantial improvement in parameter estimates.

B. The basic problem

Whenever one has noisy time series for one or more instances of a system with stochastic dynamics, one is faced with questions about how to assign any variation in the data: Do variations from one time point to the next represent a change in the common trend, a fluctuation of the dynamics with respect to the common trend, or simply measurement error? For unique time series one cannot answer this question without making strong assumptions about the average dynamics. The situation differs when time series are available for replicate systems, as illustrated in Fig. 1 for our set of replicate closed ecosystems. The measurements, in our notation the y_{kns} , are shown for 2 of 35 systems as circles and squares, respectively.

Crucially, one can estimate both the average dynamics or common trend ($\langle x_{kn} \rangle$) (estimates with associated uncertainty shown as broad bands in Fig. 1) and the fluctuations in each system around the common trend [shown are bands of inferred $x_{ks}(t) \pm 2\sigma(x_{ks}(t))$ tracking the measurement points]. Since we integrate over the underlying variables, we estimate the average dynamics and densities in individual replicates implicitly while estimating the model parameters. Whether implicit or explicit, however, this ability to separate fluctuations around the average dynamics from the average dynamics is crucial for model parameter estimation and the accurate analysis of variability in system dynamics. (We discuss how one can extract the posterior distribution of the underlying variables in detail in the Supplemental Material [26]).

C. Variance and bias of estimates

To validate our statistical method, we examined inference for a number of artificial data sets. In Fig. 2 we show results for artificial data generated by a three-species model following the dynamical and measurement equations used for inference [Eq. (1); for details of the simulations, see the Supplemental Material [26], Sec. S6] and sampled at regular intervals

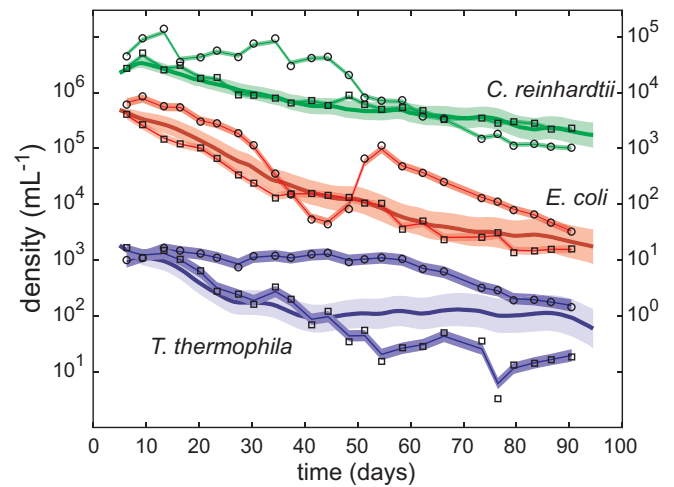


FIG. 1. (Color online) Outline of the statistical problem. Density time series for 2 of 35 replicate closed ecosystems are shown here as circles and squares, respectively (see Ref. [1]). Fitting a model implies that we estimate the true density dynamics per species for each ecosystem (shown as the posterior mode $x_{ks}(t) \pm 2$ posterior standard deviations with the thin bands tracking the data points). In addition, we implicitly estimate the common trend shared between replicate systems, shown here for each species as a broader, less variable band ± 2 posterior standard deviations. (Green) *Chlamydomonas reinhardtii*, (red) *E. coli*, and (blue) *T. thermophila*. Densities of *C. reinhardtii* according to the scale on the right. Data before day 5 were not used in analysis because mean dynamics were fast relative to the time between measurements of different replicates.

$\Delta t = T/N$. Specifically, we examine scaling of the posterior variance of the model parameters (shown here is the largest eigenvalue of the posterior covariance matrix for each group of parameters, see Ref. [39]) and bias (the average difference between the estimated and true value) as a function of the number of data points per species, NS . The number of data points can be varied in a number of ways: One can (i) vary the number of systems S at constant number of time points, N , total duration, T , and measurement interval, Δt (the crosses in Fig. 2); (ii) vary N at constant S and T (that is, Δt varies as well, filled circles); or (iii) vary N at constant S and Δt (that is, T varies as well, open circles).

By and large, the posterior variances of the model parameters scale as $1/NS$ as expected [solid lines, Figs. 2(a)–2(c)]. However, at small NS these variances were in some cases much larger than expected from extrapolation, especially for variable Δt (dashed lines, Figs. 2(a)–2(c)). Close inspection revealed that this happens as the global minimum of \mathcal{L} starts to merge with local minima assigning all the noise in the data to either measurement error alone or to dynamical noise alone (cf. Refs. [14,40]). Cases in which one can no longer separate the different parameters (here the relative contributions of dynamical noise and measurement error at large Δt) are called “identifiability problems” (see Sec. III E). As the difference between the solid and dashed curves in Figs. 2(a)–2(c) suggests, however, one can attain much more precise estimates at the same overall effort by designing the measurement schedule differently. We consider “optimal sampling” in Sec. III G.

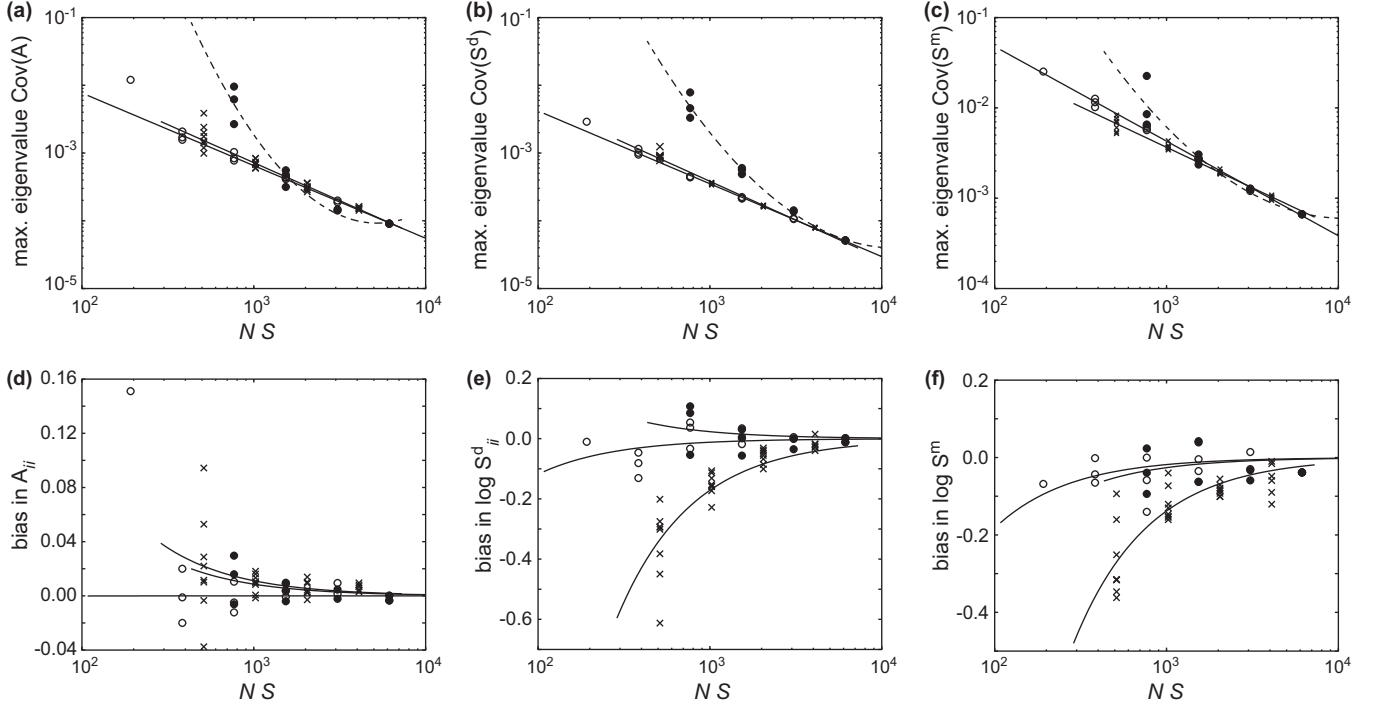


FIG. 2. Validation of the Bayesian inference method for artificial data. Shown are posterior variance [(a), (b), and (c)] and bias [(d), (e), and (f)] as a function of $N \cdot S$, the number of time points per species. (Open circles) For $S = 24$ systems the number of time points, N , was varied at constant measurement interval, Δt (that is, total duration, T , varies, too, for these time series). (Filled circles) For $S = 24$ systems the number of time points was varied at constant total duration (Δt varies, too). (Crosses) At constant N , T , and Δt the number of systems was varied. Different symbols represent different artificial data sets. (a) Scaling of the largest eigenvalue of the posterior covariance matrix of the A_{ij} (the largest-uncertainty direction). [(b) and (c)] Scaling of the largest eigenvalue of the posterior covariance matrix of the $\log S_{ii}^d$ (b) and $\log S_{ii}^m$ (c). Simple power-law fits for each case are shown as solid lines and have exponents, within error, of -1 . For variation of N at constant T (filled circles), linear fits were inadequate. Instead, quadratic fits are shown as dashed curves. Scaling of the observed bias (maximum *a posteriori* estimate=true value) for (d) the diagonal elements of A , (e) $\log S_{ii}^d$, and (f) $\log S_{ii}^m$ (each averaged over species) together with fits proportional to $1/(NS)$. Offsets are not included in any of the fits since they did not significantly differ from 0 for any curve. Results calculated using the unit information prior (see Sec. III D). Note that vertical scales are logarithmic in (a)–(c) and linear in (d)–(f).

Parameter estimates can be biased: methods can systematically under- or overestimate parameter values. To examine bias inherent in our method, we show the average bias in the A_{ij} [Fig. 2(d)], $\log S_{ii}^d$ [Fig. 2(e)] and $\log S_{ii}^m$ [Fig. 2(f)]. These biases vanish as $NS \rightarrow \infty$. Bias in each panel is most pronounced for scaling with variable S (crosses). For the A_{ii} , this is a well-documented effect in the econometrics literature and analytical results are available [41]. Bias of the A_{ii} estimates is expected to scale as $\sim 1/ST$. In essence, the estimated common trend will resemble the data more closely than the true common trend because it is inferred from the same data. Since the A_{ii} measure the time scales at which each time series tends to return to the common trend, each time series will seem to have a stronger tendency to return to the common trend (and thus have a larger A_{ii}) than is true. At smaller values of μ , this effect may be enhanced. The bias in $\log S_{ii}^d$ [Fig. 2(e)] and $\log S_{ii}^m$ [Fig. 2(f)] occurs for much the same reason: the difference between the true and estimated common trends is due to both measurement error and dynamical noise. The difference between the true and the estimated common trend will thus contain some unobserved variance. On average, one thus underestimates the strength of both sources of noise by about a factor $1/S$.

These biases highlight an important advantage of having replicate time series available: The common trend and the dynamics of individual replicate systems can be separated more reliably and, hence, effective interactions determined with less bias. We note that results are similar for Bayesian and maximum likelihood approaches since our priors have little effect on parameter estimation (see the Supplemental Material, [26] Fig. S-4).

D. Parameter estimation and the common trend

1. Choice of smoothing parameter

Often in ecology the common trend is assumed to be a simple exponential function, either by using a constant-coefficient “stochastic exponential” model (in our formulation $\mathbf{A} = \mathbf{0}$ and $\mu \rightarrow \infty$, that is, \mathbf{f} constant; cf. Refs. [9,11,21,23]) or by adding an explicit linear time dependence to $x(t)$ [42,43]. More generally, the constant-coefficient stochastic Gompertz model corresponds to $\mu \rightarrow \infty$ for general \mathbf{A} . Unfortunately, such constant-coefficient models can describe only a few types of average dynamics, and any idiosyncratic variation of the average will be confounded with fluctuations around the average dynamics. Thus constant-coefficient approaches are

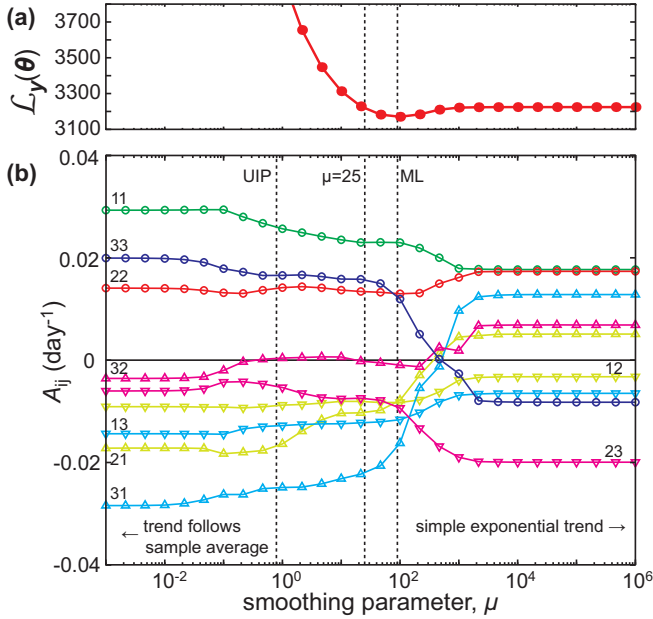


FIG. 3. (Color online) Choice of smoothing parameter. (a) Dependence of the log-likelihood function [$\mathcal{L}_y(\theta) = -2 \log l_y(\theta)$] on the smoothing parameter μ . Dotted lines represent $\mu = 25$ and the maximum likelihood (ML) estimate of μ . (b) Maximum *a posteriori* estimates of interaction coefficients A_{ij} as a function of the smoothing parameter μ . Dotted lines represent μ for the unit information prior (UIP), the ML estimate for μ , and $\mu = 25$, a value used throughout the remainder of the text. Symbols and labels marking each A_{ij} correspond to those in Fig. 4(a). Panels (a) and (b) share the same x axis.

particularly ill suited for time series of replicate ecosystems. For such data the average dynamics can—to some extent—be separated from the dynamics of each individual replicate ecosystem. Thus, one needs models like Eq. (1a), with flexible average dynamics, that is, $\mu < \infty$. We examine two alternative perspectives on the choice of smoothing parameter: treatment as a model parameter and treatment as a Bayesian prior.

From a maximum likelihood perspective, one could treat μ as an additional model parameter characterizing the dynamics of the common trend. For our data set, such straightforward optimization yielded an optimal value of μ , $\mu_{\text{ML}} \simeq 100$ [Fig. 3(a)]. In general, however, such an approach can have surprising consequences. We noticed that for some test data μ kept rising during optimization of the likelihood function. Indeed, even a set of multiple replicate time series displaying some curvature in the common trend can yield $\mu_{\text{ML}} \rightarrow \infty$. We investigate this in detail in the Supplemental Material [26], Sec. S3. For example, for a single species, for $A = 0$ and $S^m = 0$, a critical number of replicates S_c exists below which $\mu_{\text{ML}} \rightarrow \infty$,

$$S_c \approx \frac{NS^d \Delta t}{\sum_{n,n'} \Delta \bar{y}_n \Delta \bar{y}_{n'}} \sim \frac{\text{“noise”}}{\text{“structure”}}, \quad (17)$$

where \bar{y} is the average dynamics over replicate ecosystems and $\Delta \bar{y}_n = \bar{y}_{n+1} - \bar{y}_n$. In the Supplemental Material [26], Sec. S3, we derive a form of Eq. (17) valid for the full model. The result retains the same trade-off between the cost of complexity of the trend description and improvement of the fit to the

data, yielding large, sometimes infinite, μ_{ML} for small, noisy data sets with few features in the average dynamics and small μ_{ML} for large, accurate data sets with detailed features in their average dynamics. Using the result for the full model [Eq. (A3-28) of the Supplemental Material], we estimate that for our experiments one would need at least ~ 12 time points per species (that is, $N \times S$) to obtain a finite μ_{ML} in direct optimization.

From a Bayesian perspective, one can consider μ a “hyperparameter” determining the relative role of the prior distribution of the ϕ_{nk} . Specifically, we test giving the prior distribution a weight equal to that of a single additional time series (a concept known as the “unit information prior” (UIP) [44]). Following Ref. [44], this implies that the eigenvalues, $\mu^2 \lambda_D$, of the matrix $\mu^2 \mathbf{D}$ should on average be S times smaller (and the variance S times larger) than the eigenvalues of the data-derived matrix ($\mathbf{R}' = \mathbf{R} - \mu^2 \mathbf{D}$), $\lambda_{R'}$. (The matrices $\mu^2 \mathbf{D}$ and \mathbf{R} can be considered inverse covariance matrices of the prior and posterior distribution on ϕ , respectively, so \mathbf{R}' is the data-derived part.) This constraint yields

$$\mu_{\text{UIP}}^2 = \frac{\bar{\lambda}_{R'}}{S \bar{\lambda}_D}, \quad (18)$$

where the bars indicate geometric means. In practice, we first calculate \mathbf{R}' and \mathbf{D} and then set μ . For our data, the unit information prior yields $\mu_{\text{UIP}} = 0.8$. Of course, in a Bayesian approach one can introduce other prior “beliefs” about μ instead.

To examine the consequences of each choice of μ , we examine their effects on inference for our experimental data. It is important to keep in mind that, in principle, our method fits all the observed dynamics, not just the average dynamics or the fluctuations around it. At low μ , the inferred common trend is described by a large number of trend variables (here up to 188 variables, $\{\phi_{nk}\}$, per species as $\mu \rightarrow 0$) and effectively traces the sample average dynamics. Since many variables $\{\phi_{nk}\}$ are available to fit the average dynamics, estimates of the (other) model parameters are dominated by the properties of the observed fluctuations *around* the average dynamics. However, some variation among the $\{\phi_{nk}\}$ is likely due to limited sampling and measurement error. As μ increases, this variation is progressively filtered out. This noise-suppression effect should become relevant around μ_{UIP} . The accompanying large decrease in $\mathcal{L}_y(\theta; \mu)$ [Fig. 3(a)] implies that many of the $\{\phi_{nk}\}$ are not well determined at low μ .

As shown in Fig. 3, as we further increase μ up to about the maximum likelihood estimate for μ , μ_{ML} , parameter estimates are fairly insensitive to the precise value of μ . For example, estimates for A_{ij} based on the unit information prior, μ_{UIP} , are quite similar to those at μ_{ML} . The same is true for estimates of the S_i^m and S_{ij}^d (data not shown). However, beyond μ_{ML} few effective degrees of freedom remain in the common trend (see the Supplemental Material [26], Sec. S3 and Fig. S-2, for a quantification). Since the parameters A_{ij} do appear in the description of the common trend (see, for example, the Supplemental Material [26], Sec. S3), a decrease in the number of effective degrees of freedom in the common trend description implies that estimates of the A_{ij} are now also affected by the average dynamics. As a consequence, any systematic deviations between the observed common trend and

any trend the model can describe will introduce idiosyncratic, and possibly large, biases in the A_{ij} . At small μ , on the other hand, the ϕ_{nk} will describe the average dynamics, and the estimates of A_{ij} will be dominated by the properties of the fluctuations around the common trend. If one, thus, is interested in the statistical structure of the fluctuations around the average dynamics, high values of μ should be avoided, even if this is not immediately apparent from the log-likelihood.

This analysis should be disconcerting to anyone analyzing unique time series. Even though our common trend does not display any pronounced features (Fig. 1), the assumption of constant f (or large μ) would lead to incorrect inferences about the statistical structure of fluctuations around the average dynamics. For example, while the change in log-likelihood is modest, three of nine coefficients, involving all species, clearly switch sign as μ is increased beyond μ_{ML} .

2. Estimates of interaction coefficients

Once the smoothing parameter has been set, one can examine the posterior distribution of the model parameters. To do so, we note that the parameter transformations discussed in Sec. II E allow us to normalize the posterior distribution and calculate the posterior covariance and correlation matrices of the inverse parameters.

Of direct ecological interest are the estimates of the effective interaction coefficients A_{ij} . We will analyze these interaction coefficients for $\mu = 25$. As shown in Fig. 4(a), estimates of eight of nine interaction coefficients significantly differ from 0 (the exception, A_{32} , is the effect of *E. coli* on *T. thermophila*).

Besides the statistical validation of our method above, we sought to provide some biological validation for these estimates. That is, do we estimate effective interaction coefficients that are reasonable? To do so, we analyzed an additional set

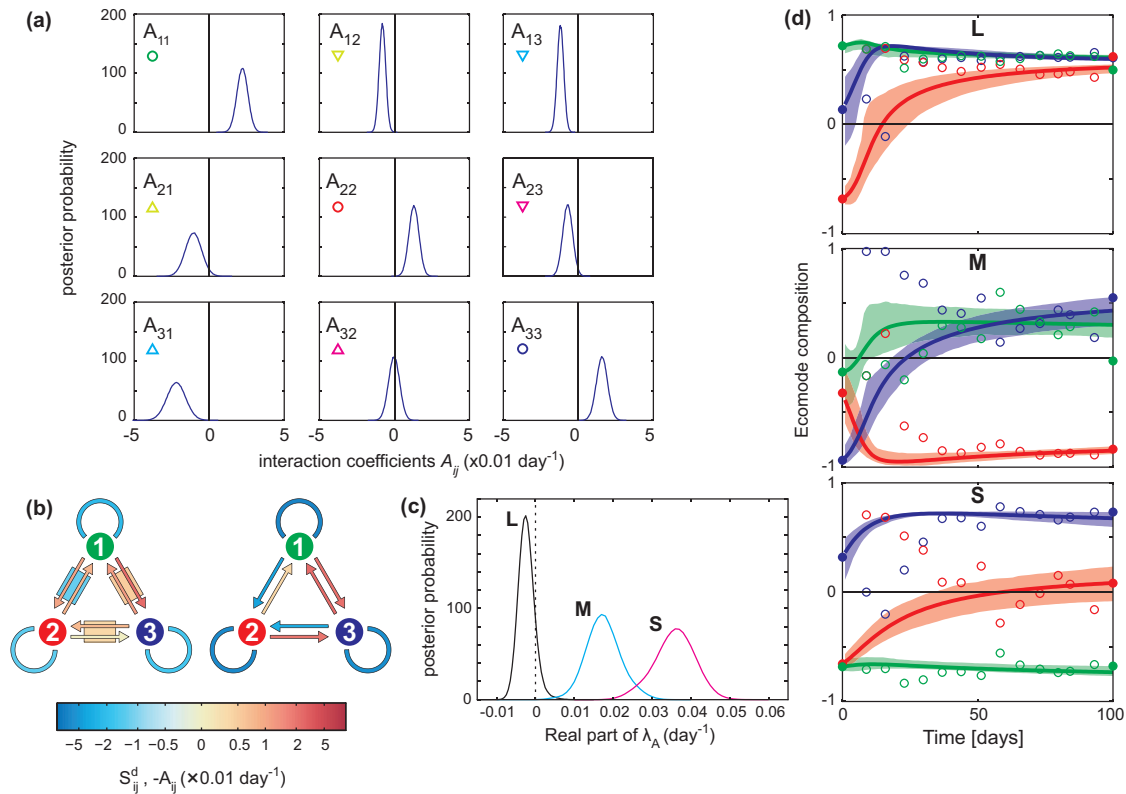


FIG. 4. (Color online) Inference for data obtained for replicate closed microbial ecosystems. Throughout, we indicate *C. reinhardtii* by 1 and the color green (light gray), *E. coli* by 2 and red (gray), and *T. thermophila* by 3 and blue (dark gray). (a) Marginal posterior distributions for the interaction coefficients A_{ij} . (b) Schematic display of estimated interactions between the three species (maximum *a posteriori* estimates). Arrows: $-A_{ij}$. Undirected “blocks” between species: S_{ij}^d (off-diagonal elements only). Left diagram: Estimates for our three-species data set. Right diagram: Estimates for “coarse” one- and two-species data. We use $-A_{ij}$, since $-A_{ij}$ appear in the state equation [Eq. (2)]. That is, $-A_{ij} < 0$ indicates a negative effect of species j on the density of species i . In both cases all self-interactions are negative and all interspecies interactions are positive, except for the undirected term due to correlations in dynamical noise between *C. reinhardtii* and *E. coli* in the left diagram and effects of *C. reinhardtii* and *T. thermophila* on *E. coli* in the right diagram. (c) Posterior distributions for the three eigenvalues of A , sorted by size. Imaginary components were small or nonexistent. Labels L, M, and S are indicated to match panel d. (d) Observed dynamics (open circles) of the eigenvectors of C_S (“ecomodes”), compared to the dynamics of these ecomodes predicted by the model (for x), shown as posterior mean and 90% confidence interval for the mean. Ecomodes are labeled according to the corresponding eigenvalues of C_S as L (large), M (middle), and S (small), respectively. Predictions were obtained by drawing from the joint posterior distribution of (A, S^d) and propagating the dynamics of C_S according to Supplemental Material [26], Sec. S1. Filled circles at $t = 0$ indicate the eigenvectors of S^d , to which C_S should reduce as $t \rightarrow 0$. At $t = 100$ days, filled circles indicate the eigenvectors of A . Comparison of panels (c) and (d) shows, for example, that the largest ecomode (d) is similar to the eigenvector of the matrix A (c) with the smallest eigenvalue. Observed ecomodes are poorly defined over the first 20 days.

of “coarse” measurements on duplicate one- and two-species closed ecosystems described in the Supplemental Material [26], Sec. S4. In these experiments, ecosystems were sampled manually after gentle mixing every 7–10 days for 9 weeks and densities determined by haemocytometry (*T. thermophila* and *C. reinhardtii*) and plate counts (*E. coli*).

The level of sampling and replication of this coarse data set precludes separate identification of the two sources of noise, measurement error, and dynamical noise. A schematic comparison of the obtained effective interaction coefficients is shown in Fig. 4(b). For most interactions, we obtain the same sign for both types of experiment, but the estimated effective interactions tend to be stronger in the coarse measurements. This could be a result of increased bias in the analysis of this small data set. For two of the interactions, the effects of both *T. thermophila* and *C. reinhardtii* on *E. coli*, the estimated sign differs between the two types of experiments. We do note, however, that a corresponding discrepancy can be seen between the effective interactions in the coarse one- and two-species experiments and a coarse three-species experiment conducted at the same time (see also the Supplemental Material [26], Sec. S4 and Fig. S-3). This suggests that *E. coli* consistently achieves higher densities in the presence of both other species than expected from the effective interactions with either species alone, suggesting a possible three-species interaction term. We detail these observations in the Supplemental Material [26], Sec. S4. It is noteworthy that the interactions obtained from fluctuations in a three-species system resemble those found in response to very strong perturbations (that is, the presence or absence of one or more species).

3. Fluctuations in closed ecosystems

The availability of the joint posterior distribution of the model parameters allows us to ask more intricate questions, too. For example, one can determine the posterior probability distribution for each of the eigenvalues of \mathbf{A} , λ_A . The posterior probability distribution of the real parts of the λ_A , sorted by size, is shown in Fig. 4(c). From a biological point of view, these results are remarkable: The smallest eigenvalue of \mathbf{A} is indistinguishable from 0 (suggesting random-walk dynamics along its corresponding eigenvector [1]), even though the growth of each species, directly, is self-limiting ($A_{ii} > 0$). That is, mutually “stimulating” interspecies interactions appear to largely cancel out intraspecies density dependence at the measured densities and time scales. We caution that the obtained interaction coefficients are effective ones and can reflect correlated changes in fitness during evolutionary change or spatial interactions as much as, e.g., predator-prey interactions or direct resource competition.

We further illustrate the utility of having the joint posterior distribution of model parameters by addressing how species-species interactions structure the variation arising between replicate ecosystems. This question was raised by our previous work [1], in which we studied the structure of variability in populations in replicate closed ecosystem experiments. To examine the variability across the set of replicate ecosystems (S), we estimated the logarithmic density correlation matrix

\mathbf{C}_S as follows:

$$\mathbf{C}_{Skl}(t_n) = \frac{\text{Cov}(y_{kn}, y_{ln})}{\sigma(y_{kn})\sigma(y_{ln})}. \quad (19)$$

That is, this matrix describes how different realizations of replicate ecosystems at a particular time point are distributed in the space of species densities ($\log n_1, \log n_2, \dots$). We observed, first, that the eigenvectors and eigenvalues of \mathbf{C}_S were distinct and stabilized after about 20 days in a reproducible way [see also Fig. 4(d), open circles]. Because of these properties, we call these eigenvectors “ecomodes.”

Second, we observed that these ecomodes were similar to the eigenvectors of the “species interaction matrix” [1,19], which measures the structure of short-term (weekly) fluctuations in species densities within individual ecosystems, suggesting that species interactions determine the long-term structure of variation between replicates. We obtained this species interaction matrix by simple linear regression of measured deviations of the logarithmic densities from the replicate average 1 week apart.

Within the Gompertz model, one can examine the relationship between species interactions and ecomodes in more detail. Specifically, there are *two* sources of correlation between species densities in the model: the effective interaction coefficients A_{ij} and the dynamical noise covariances S_{ij}^d . Since we have an analytical expression for the dynamics of the logarithmic density covariance matrix [Eq. (7) and more detail in the Supplemental Material [26], Sec. S1], we can repeatedly sample the joint posterior distribution for \mathbf{A} and \mathbf{S}^d and obtain a distribution of predicted ecomode dynamics (Fig. 4(d), compared with estimates for the data obtained in Ref. [1]). The ecomodes are labeled L (large), M (medium), and S (small) in order of their corresponding eigenvalues (L: ~ 2 , M: ~ 0.7 , and S: ~ 0.3).

We compare the predicted ecomode dynamics (lines with 90% posterior confidence interval) with observations from Ref. [1] [Fig. 4(d), open circles]. To determine which source of species-species correlations dominates ecomode composition within our model, we display the eigenvectors of the density correlation matrix corresponding to \mathbf{S}^d alone as filled circles on the left y axis ($t = 0$), and the eigenvectors of \mathbf{A} as filled circles on the right-hand y axis ($t = 100$ days). This display is motivated by the observation that dynamical noise exerts its effects instantaneously, while the A_{ij} act over time. To establish the correspondence between eigenvectors of \mathbf{A} and ecomodes, we use the labels L, M, and S, connecting the eigenvectors of \mathbf{A} shown in Fig. 4(d) at $t = 100$ days to the eigenvalues of \mathbf{A} [Fig. 4(c)]. Even though the relation between \mathbf{A} and \mathbf{C}_S is in general quite complex [19], the ecomodes resemble the eigenvectors of \mathbf{A} quite closely after a few weeks, despite the additional interspecies correlations present in the dynamical noise. In particular, the L ecomode is connected with the eigenvector of \mathbf{A} with a nearly 0 eigenvalue.

We conclude that after a few weeks the structure of density covariation across systems is dominated by the effective species interactions and not by correlations within the dynamical noise. Indeed, we see that the largest density fluctuations occur along a collective dimension of the ecosystem which nearly coincides with the eigenvector of \mathbf{A} with nearly zero

eigenvalue. Moreover, this occurs despite negative density dependence for each individual species (the $A_{ii} > 0$). This situation is reminiscent of the phenomenon of co-integration of macroeconomic time series [45].

E. Identifiability of model parameters

It may not be possible to estimate every parameter combination accurately from the data. In essence, this problem arises when different θ give rise to similar $P(\mathbf{Y}|\theta)$, with \mathbf{Y} the general stochastic variable of which the data, \mathbf{y} , are considered a realization. In such a case, the model is said to be not (or not completely) *identifiable*. Identifiability can be global or local in parameter space [for example, in a regression $y = (a/b)x + \eta$, a and b cannot be estimated independently anywhere in parameter space, while in a regression $y = ax/(x + b) + \eta$, a and b cannot be estimated independently if $b \gg x$ for the available measurements].

Using our data set, we study identifiability empirically using the posterior correlation matrix of the model parameters [Fig. 5(a)]. (Note that the posterior correlation matrix is estimated using the Hessian matrix [Eq. (15)] as an approximation of the inverse posterior covariance matrix.) For example, the observed posterior correlation between the estimates of S_{11}^d and S_1^m is -0.60 [highlighted in Fig. 5(a)]. We illustrate this using a contour plot of the joint posterior distribution of S_{11}^d and S_1^m [Fig. 5(b)]. Shown are also the eigenvectors of the posterior covariance matrix of these two model parameters [Fig. 5(b), blue arrows]. Along the direction nearly parallel to the diagonal, along which the sum of the two noises varies, the posterior variance is about $6\times$ smaller than along the eigenvector orthogonal to it, along which their

difference varies. That is, the sum is better constrained than the difference. (We note that such effects imply as well that marginal variances of $\log S_{11}^d$ and $\log S_1^m$ are both about twice as large as their conditional variances with all other model parameters constrained at their optimal values.)

In Fig. 5(c) we show the eigenvalue spectrum for the entire posterior correlation matrix (red dots). The larger the eigenvalue, λ , the harder the parameter combination described by the corresponding eigenvector is to estimate. We show the composition of the eigenvectors with the largest eigenvalues in Fig. 5(d). The dominant effect displayed by each eigenvector is a noise identifiability problem for each species. It is often hard to distinguish between dynamical noise and measurement error in models containing both [9]. For example, for a simple Gompertz model (in our notation $K = 1$, $S = 1$, $\mu \rightarrow \infty$), Knappe [40] showed that for large A (strong “density dependence”) it is hard to separate the relative contributions of S^d and S^m . In a single species model, the total noise in dynamics (asymptotically) equals $S^d/A + S^m$. Indeed, we see that the estimates of S_{ii}^d , A_{ii} , and S_i^m are correlated for each species [Fig. 5(d)]. For example, as shown in $\mathbf{v}^{(c)}$, for the same data a larger estimate of S_{22}^d would yield a larger estimate of A_{22} and a smaller estimate for S_2^m . An important consequence is that, if one assumes any of these three parameters to be zero, one introduces significant bias in the estimates of the other two parameters. There is significant additional structure in the eigenvectors shown in Fig. 5(d). In particular, since the densities of *C. reinhardtii* and *T. thermophila* tend to covary in our data set, there is some difficulty telling the effects of the two species apart [that is, a larger estimate of A_{11} implies a smaller estimate of A_{13} and vice versa in $\mathbf{v}^{(a)}$, and likewise for A_{31} and A_{33} in $\mathbf{v}^{(b)}$, Fig. 5(d)].

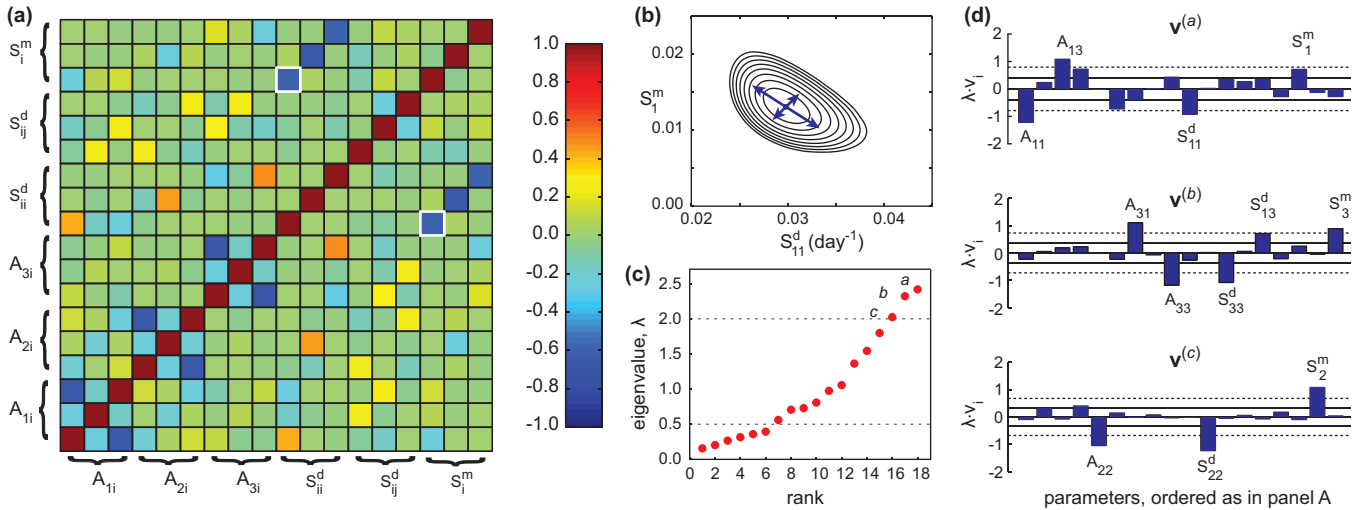


FIG. 5. (Color online) Analysis of posterior correlations between parameters. (a) The posterior correlation matrix of the linear model for our experimental data contains some significantly nonzero elements (for the scale, see the colorbar; substantial posterior correlations between interaction coefficients are negative, between measurement error and dynamical noise negative, and between dynamical noise and interaction coefficients positive). (b) Illustration of posterior correlation: There is some correlation between the estimates of dynamical and measurement noise as illustrated for *C. reinhardtii*: The total noise can be estimated much better than the size of either dynamical or measurement noise separately. The other parameters were constrained to their optimal values. (c) Eigenvalues of the posterior correlation matrix ranked by magnitude. The three largest eigenvalues are labeled a , b , and c , respectively. (d) Composition of the three eigenvectors of the posterior correlation matrix [see the corresponding labels in (c)]. Solid and dotted lines indicate one and two σ levels, respectively, for uncorrelated Gaussian random variables with the same number of observations. Some of the most pronounced components of each eigenvector are labeled. In all panels we indicate *C. reinhardtii* by 1, *E. coli* by 2, and *T. thermophila* by 3 in subscripts.

How does one judge the statistical significance of these results? In principle, since a closed form expression for the likelihood function is available, one can calculate the posterior correlation matrix exactly and even derive its entire distribution for any set of model parameter values. We will use such an explicit calculation for the elements of the expected Hessian matrix for a simpler model when we examine the possibility of improvement of experimental designs in Sec. III G. For our data, statistical fluctuations are expected to be relatively small, since the number of observations is much larger than the number of entries in the posterior correlation matrix. For smaller data sets, we recommend study of the Hessian matrix instead in order to avoid numerical errors from matrix inversion.

At this point, we see room for further application of techniques from statistical physics in two directions. First, it seems reasonable that methods from random matrix theory [46] can be extended to posterior correlation matrices for complex models. That is, it should be possible to judge the significance of modes of the posterior correlation matrix by more general considerations of number of replicates, time points, and so on. Second, the posterior correlation matrix is closely related to matrices used to study system sensitivity to parameter perturbations in systems biology and other fields [47]. Recent statistical physics approaches to such matrices suggest such matrices may fall into certain universality classes [48], based on their eigenvalue spectra.

We make three more observations. First, replication plays an important role. An empirical analysis [as in Figs. 5(b)–5(d)] generally requires at least a number of observations comparable to the number of elements in the posterior correlation matrix [49]. In our case, keeping K, N constant, this means one would need at least $S > p^2 / KN \approx 4.5$ replicates, with $p = 18$ the number of model parameters. More generally, for the Gompertz model $p = \frac{3}{2}K(K + 1)$ and one needs at least $S > \frac{9}{4}K(K + 1)^2 / N$.

Second, as long as posterior correlation does not equal ± 1 , large eigenvalues of the posterior correlation matrix only mean that one needs more data than in the absence of posterior correlation to obtain good parameter estimates. For example, the posterior correlation between S_{11}^d and S_1^m is -0.60 . It would be easier to estimate these parameters without this posterior correlation, but we obtain good estimates for both (relative errors of 6% and 15% for S_{11}^d and S_1^m , respectively) and even for their ratio (relative error: 19%). Finally, as we will show below, thoughtful experimental design can substantially alleviate identifiability issues.

F. Model selection

Once one has estimated model parameters, it is natural to ask how good the model really is. We will consider two ways of addressing this. First, we will compare the quality of one model relative to another, a task known as “model selection.” Second, we will look for a way to address how well an individual model fits the data. Our principle aim in this section is to stimulate a more critical attitude towards model selection criteria in ecology.

1. Model selection criteria

To judge the relative merits of models, we calculate Bayes factors [50] or, more precisely, their logarithms. For a model \mathcal{M} and data \mathbf{y} , Bayes’s theorem yields $P(\mathcal{M}|\mathbf{y})P(\mathbf{y}) = P(\mathbf{y}|\mathcal{M})P(\mathcal{M})$, so one can compare the relative probabilities of two models, \mathcal{M}_1 and \mathcal{M}_2 , being the models generating the data, as follows:

$$\frac{P(\mathcal{M}_1|\mathbf{y})}{P(\mathcal{M}_2|\mathbf{y})} = \frac{P(\mathbf{y}|\mathcal{M}_1)}{P(\mathbf{y}|\mathcal{M}_2)} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \quad (20)$$

For now, we will assume that models are *a priori* equally likely, although alternative assumptions can also be considered (and, as we will see, this choice can have nontrivial consequences). The ratio of $P(\mathbf{y}|\mathcal{M}_1)$ and $P(\mathbf{y}|\mathcal{M}_2)$ is known as the *Bayes factor*. Hence, we calculate for each model $P(\mathbf{y}) = P(\mathbf{y}|\mathcal{M})$ as follows:

$$\begin{aligned} P(\mathbf{y}|\mathcal{M}) &= \int d\theta P(\mathbf{y}|\theta)P(\theta) \\ &\simeq \int d\theta P(\mathbf{y}|\hat{\theta})P(\hat{\theta}) \\ &\quad \times \exp\left(-\frac{1}{2}(\theta - \hat{\theta})^\top \mathbf{H}(\theta - \hat{\theta})\right) \\ &= P(\mathbf{y}|\hat{\theta})P(\hat{\theta})\sqrt{(2\pi)^p |\mathbf{H}^{-1}|}, \end{aligned} \quad (21)$$

where, for each model, $\theta = \theta_{\mathcal{M}}$; $\hat{\theta}$ is the optimal parameter vector and where \mathbf{H} is the Hessian matrix at $\hat{\theta}$ as in Eq. (15). The approximation made here is that the likelihood function is essentially Gaussian (or the log-likelihood function parabolic) around $\hat{\theta}$. Often, one can make this an excellent approximation by finding a suitable transformation $\vartheta = \mathbf{w}(\theta)$ after $\hat{\theta}$ has been found [35], as described in Sec. II E. In our case, simply using $\log S_i^m$ and $\log S_{ii}^d$ proves adequate, leaving the other parameters untransformed (Supplemental Material [26], Fig. S-4).

We can compare our analysis to the Bayesian information criterion (BIC) [51]. The BIC is commonly used as a model selection criterion in ecology. The BIC in our case is as follows:

$$\text{BIC} = -2 \log P(\mathbf{y}|\hat{\theta}) + p \log(KNS). \quad (22)$$

For comparison, our criterion [Eq. (21)] can be rewritten by taking logarithms [52],

$$\begin{aligned} -2 \log P(\mathbf{y}) &= -2 \log P(\mathbf{y}|\hat{\theta}) - 2 \log P(\hat{\theta}) \\ &\quad - p \log(2\pi) + \log |\mathbf{H}| \\ &= -2 \log P(\mathbf{y}|\hat{\theta}) - 2 \log P(\hat{\theta}) \\ &\quad + p \log(KNS) + \log |\mathbf{h}| \\ &\text{with } \mathbf{h} = \frac{2\pi \mathbf{H}}{KNS}, \end{aligned} \quad (23)$$

where we defined a normalized Hessian matrix, \mathbf{h} , since the size of elements of \mathbf{H} will grow asymptotically linearly with the number of data points, KNS (provided the model is identifiable).

We observe two differences between Eqs. (23) and (22): The BIC neglects the prior information on model parameters, and the detailed structure of \mathbf{h} , both of which do not scale with the number of data points. Indeed, the BIC, like other such criteria,

TABLE III. Comparison of different models under two treatments of the trend prior ($\mu = 25$, top rows) and using a “unit information prior” (μ_{UIP} , bottom rows). Each model is described in the text. p is the number of parameters in each model. The other quantities refer to Eqs. (23) and (22). Smallest values among alternative models are indicated as bold text.

Trend prior, model		p	$-2 \log P(\mathbf{y} \hat{\boldsymbol{\theta}})$	$-2 \log P(\hat{\boldsymbol{\theta}})$	$-2 \log P(\mathbf{y})$	BIC
$\mu = 25$	Full model	18	3216.8	49.6	3351.2	3358.0
	Diagonal \mathbf{S}^d	15	3275.1	30.8	3360.6	3392.7
	Diagonal \mathbf{A}	12	3271.5	33.4	3332.6	3365.6
	Uncoupled	9	3338.8	14.4	3351.1	3409.4
	Random walk	9	3280.7	25.2	3303.2	3351.3
μ_{UIP}	Full model	18	4081.7	45.1	4208.0	4222.9
	Diagonal \mathbf{S}^d	15	4130.5	27.7	4211.0	4248.2
	Diagonal \mathbf{A}	12	4139.6	29.0	4193.0	4233.7
	Uncoupled	9	4193.5	11.3	4200.7	4264.1
	Random walk	9	4146.1	20.9	4160.8	4216.7

was derived for the limit of a large number of data points. Unfortunately, ecology rarely operates in this limit. Posterior correlations between the model parameters (see above) further limit the range of validity of such asymptotic approximations. For completeness, we note that for transformed parameters $\boldsymbol{\vartheta} = \mathbf{w}(\boldsymbol{\theta})$, we obtain

$$\begin{aligned}
 -2 \log P(\mathbf{y}) &= -2 \log P(\mathbf{y}|\hat{\boldsymbol{\vartheta}}) - 2 \log P_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}) \\
 &\quad - p \log(2\pi) + \log |\tilde{\mathbf{H}}| \\
 \text{with } P_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}|\mathbf{J}) &= P_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \quad \text{and} \quad \mathbf{J}^{\top} \tilde{\mathbf{H}} \mathbf{J} = \mathbf{H} \\
 \text{where } J_{ij} &= \frac{\partial \vartheta_i}{\partial \theta_j}.
 \end{aligned} \tag{24}$$

To illustrate the use of Bayes factors, we examine a number of “submodels.” As a reminder, the full model is given by $\dot{\mathbf{x}}_s = \mathbf{f}(t) - \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$ and $y_{kns} = x_{ks}(t_n) + \xi_{kns}$, with $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{S}^d dt)$ and $\xi_{kns} \sim \mathcal{N}(0, S_k^m)$. We considered leaving out each of the two sources of density correlations between species: The second row of Table III (“diagonal \mathbf{S}^d ”) shows the effects of setting the off-diagonal elements of the dynamical noise matrix \mathbf{S}^d to 0, while the third row (“diagonal \mathbf{A} ”) shows the effect of setting the off-diagonal elements in the effective interaction matrix to 0. In the fourth row (“uncoupled”), the dynamics of each species are entirely independent (zero off-diagonal elements in both \mathbf{A} and \mathbf{S}^d). Finally, in the fifth row (“random walk”), we explore a simple random-walk description ($\mathbf{A} = \mathbf{0}$), retaining the off-diagonal elements in \mathbf{S}^d .

In each case, the full model fits the data the best at the optimal parameters [as measured by $-2 \log P(\mathbf{y}|\hat{\boldsymbol{\theta}})$, to which we will return below]. However, both measured by our “Bayes factor criterion,” $-2 \log P(\mathbf{y})$, and by the BIC, the random-walk description is superior as an efficient description of the data, that is, once we average over the posterior distribution of the model parameters. This result is consistent with our previous work [1]. In that work, we showed that the random-walk-like fluctuations of logarithmic density along the largest ecomode (see Sec. III D) dominate the density fluctuations of each of the species.

Even though the Bayes factor and BIC agree in selecting the random-walk model as the best model, they do not agree on the ranking of the other models. Indeed, the full and uncoupled models do equally well, as measured by Bayes factors, but

their BICs differ by >50 , which constitutes a large difference in evidence [50]. Such a difference is comparable to the differences in BIC scores between the different models. *This means that the formulation of the prior and the structure of the normalized Hessian cannot be ignored in model selection, even for our data set, which is large by any ecological standard.* [We note that the ranking of the models by Bayes factors is robust to the choice of smoothing parameter (Table IV, bottom rows)].

The fact that both the “diagonal \mathbf{S}^d ” and “diagonal \mathbf{A} ” models fit the data much better than the uncoupled model suggests that off-diagonal elements in \mathbf{S}^d and \mathbf{A} contain partially redundant information about the correlations between species. This observation is related to the identifiability issues described in Sec. III E. Consistent with the potential for redundancy in the statistical description, we note that the estimated effective interaction matrix is fairly symmetric [Fig. 4(b)] so A_{ij} and S_{ij}^d could readily absorb some of the effects of each other’s omission.

2. Lindley paradox

There is something surprising about the performance of the random-walk model ($\mathbf{A} = \mathbf{0}$): above (Sec. III D) we found that eight of nine elements of \mathbf{A} differed significantly from 0, yet the random-walk model provides a seemingly better description of the data than the full model, judging by the Bayes factors. This inconsistency is known as Lindley’s paradox [53] and occurs often in ecology [54]. It occurs if one assigns finite prior probability to a finite set of discrete parameter values. On the one hand, our prior probability density on the A_{ij} assigns 0 probability to any precise value of A_{ij} [e.g., $\lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{+\epsilon} P(A_{ij}) dA_{ij} = 0$]. On the other hand, in comparing the full model to the random-walk model using $-2 \log P(\mathbf{y})$, we assigned the model with all $A_{ij} = 0$ and the full model equal prior probability.

These choices together imply an overall prior distribution on \mathbf{A} as illustrated in Fig. 6(a) for a single interaction coefficient. The question is whether it is natural to assign finite prior probability to the case of effectively noninteracting species. This is as much a biological consideration as a statistical one. If one wants to compare “noninteracting species” and “interacting species” as *a priori* equally plausible

TABLE IV. Constants used in Eq. (25) to derive the expected distribution of the likelihood function. \mathbf{V} is an auxiliary matrix used to simplify notation.

$$\begin{aligned} \mathbf{M}_{ns, \bar{n}\bar{s}} &= (\mathbf{S}^m)^{-1} \delta_{s\bar{s}} \delta_{n\bar{n}} - (\mathbf{S}^m)^{-1} \mathbf{Q}_{n\bar{n}}^{-1} (\mathbf{S}^m)^{-1} \delta_{s\bar{s}} - \sum_{n', \bar{n}'} \mathbf{V}_{nn'}^{(s)\top} \mathbf{R}_{n'\bar{n}'}^{-1} \mathbf{V}_{\bar{n}\bar{n}'}^{(s)} \\ \mathbf{h}_{ns} &= \mu^2 \sum_{n', \bar{n}, \bar{n}'} \mathbf{V}_{nn'}^{(s)\top} \mathbf{R}_{n'\bar{n}}^{-1} \mathbf{D}_{\bar{n}\bar{n}'} \phi_{\bar{n}'}^0 \\ C' &= \sum_{n=1}^{N-1} \sum_{n'=1}^{N-1} \phi_n^{0\top} (\mu^2 \mathbf{D}_{nn'} - \mu^4 (\mathbf{D}\mathbf{R}^{-1}\mathbf{D})_{nn'}) \phi_{n'}^0 + S \sum_{n=1}^{N-1} \log |\Gamma_n| + SN \log |S^m| \\ &\quad + \log |\mathbf{R}| + S \log |\mathbf{Q}| - \log |\mu^2 \mathbf{D}| + K(S+1)(N-1) \log(2\pi) \\ \mathbf{V}_{nn'}^{(s)} &= \mathbb{1}_{(n < N)} \Delta t_n (\Gamma_n^{-1} \mathbf{Q}_{n+1, n'}^{-1} - \beta_n \mathbf{Q}_{nn'}^{-1}) (\mathbf{S}^m)^{-1} \end{aligned}$$

alternative models, the random-walk model turns out to be the best for our data. One can, alternatively, take, for example, $P(\{|A_{ij}| < 1/T \text{ for all } i, j\})$ as a rough prior probability of the random-walk model. In other words, one can ask whether the interactions are effectively weak on the time scale of the experiment. The resulting implied overall prior distribution on \mathbf{A} is illustrated in Fig. 6(b) (again for a single interaction coefficient). This choice would increase $-2 \log P(\mathbf{y})$ for the random-walk model by about +94, making the full model the better one. In other words, the answer provided by Bayes factors will—correctly—depend on which question one precisely asks. If data analysis is to provide any clarity, questions should be asked precisely and prior beliefs stated explicitly and consistently.

3. Absolute model quality

As a second, absolute criterion for model quality, we ask whether the data look like typical “output” of the model [55]. For example, the best model identified by a model selection criterion, as in Eq. (23), may still not fit the data very well. Alternatively, we could have overfitted the data, making the data unusually probable under the model. To judge this, we ask *how the probability of the data compares with the probability of realizations of the model with the same measurement schedule*. While we can ask this question for unique time series as well, it seems particularly well suited for a set of replicate time series.

As a starting point, we reinterpret the likelihood function, properly normalized, [Eq. (13) and Table II] as a probability distribution over \mathbf{Y} , the space of possible data sets. That is, $P(\mathbf{Y}|\boldsymbol{\theta}) = l_{\mathbf{y}}(\boldsymbol{\theta}; \mu)$. In general, a linear model with Gaussian dynamical and measurement error will produce a log-likelihood function of the type

$$\begin{aligned} \mathcal{L}_{\mathbf{y}}(\boldsymbol{\theta}) &= -2 \log l_{\mathbf{y}}(\boldsymbol{\theta}; \mu) \\ &= \mathbf{y}^\top \mathbf{M} \mathbf{y} - \mathbf{y}^\top \mathbf{h} - \mathbf{h}^\top \mathbf{y} + C \\ &= (\mathbf{y} - \mathbf{M}^{-1} \mathbf{h})^\top \mathbf{M} (\mathbf{y} - \mathbf{M}^{-1} \mathbf{h}) \\ &\quad - \mathbf{h}^\top \mathbf{M}^{-1} \mathbf{h} + C \end{aligned} \quad (25)$$

$$\text{hence, } P(\mathbf{Y}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{M}^{-1} \mathbf{h}, \mathbf{M}^{-1}),$$

where \mathcal{N} indicates a multivariate normal distribution. One readily obtains C , \mathbf{M} , and \mathbf{h} by rearranging terms in the likelihood function (see Table II), yielding Table IV. We now compare the observed value of the log-likelihood function to the distribution of log-likelihood function values one would observe for data sets generated by the fitted model. Defining $\mathcal{L}_0 = C - \mathbf{h}^\top \mathbf{M}^{-1} \mathbf{h}$, we see that $\mathcal{L}_{\mathbf{y}}(\hat{\boldsymbol{\theta}}) - \mathcal{L}_0 = (\mathbf{Y} - \mathbf{M}^{-1} \mathbf{h})^\top \mathbf{M} (\mathbf{Y} - \mathbf{M}^{-1} \mathbf{h})$ has a χ^2 distribution with KNS degrees of freedom for realizations \mathbf{Y} of the state space model. For our data, we calculate the log-likelihood $\mathcal{L}_{\mathbf{y}}(\hat{\boldsymbol{\theta}})$ at the estimated model parameters. Surprisingly, the observed data are quite typical for the model at $\hat{\boldsymbol{\theta}}$, meaning that the fit of the

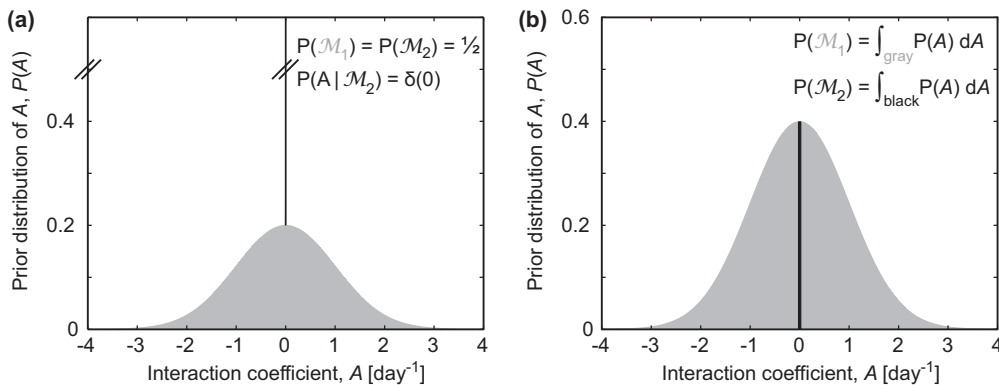


FIG. 6. The implied overall prior distribution and model comparison. For simplicity, we illustrate implied prior distributions on \mathbf{A} for a single interaction coefficient A . The overall prior distribution is $P(A) = P(A|\mathcal{M}_1)P(\mathcal{M}_1) + P(A|\mathcal{M}_2)P(\mathcal{M}_2)$. In both panels, the contribution from the prior distribution of the full model, \mathcal{M}_1 , is shown in gray; the contribution from the random-walk model, \mathcal{M}_2 with $A = 0$, is shown in black. (a) Comparison of the two models using equal *a priori* weights ($P(\mathcal{M}_1) = P(\mathcal{M}_2)$) and a precise prior on $A = 0$. The implied overall prior has an infinite peak (delta function) at $A = 0$ with integrated weight $\frac{1}{2}$. (b) One can instead first specify the overall prior distribution on A and then derive model probabilities from it. For example, in this illustration, we chose $P(\mathcal{M}_2) = \int_{-\epsilon}^{+\epsilon} P(A) dA$ and $P(\mathcal{M}_1) = 1 - P(\mathcal{M}_2)$, with ϵ dependent on what can be resolved in the experiment ($\epsilon = 1/T$).

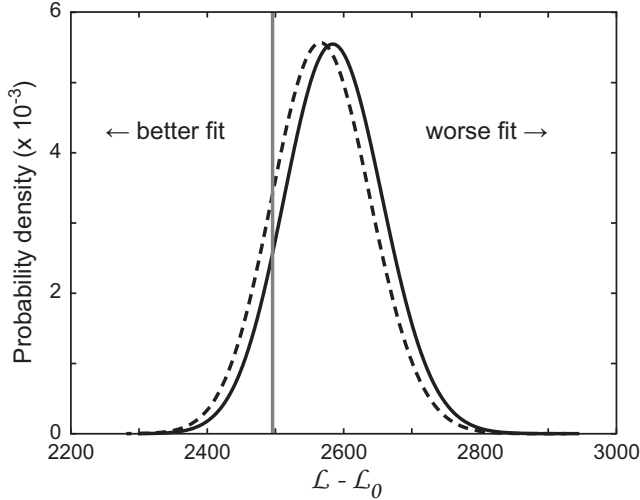


FIG. 7. Absolute model quality. The fit of the model to the closed ecosystem data, measured by comparing the observed value of the log-likelihood (gray vertical line) to the distribution of log-likelihood values expected for data generated by the Gompertz model at the inferred parameters (black). The latter is a χ^2 distribution with KNS degrees of freedom. Correction of the number of degrees of freedom by the number of parameters yields the dashed curve.

model to the data is as good as expected were the data actually generated by the model (Fig. 7).

We note that this criterion for model quality has a structure similar to so-called “goodness-of-fit” quantities, like Pearson’s χ^2 test statistic, as well as the deviance information criterion [56]. An obvious concern is that $\hat{\theta}$ is itself estimated from the data rather than the true θ , leading to an artificially good fit. Like for conventional asymptotic goodness-of-fit tests, one can (approximately) correct for this bias by adjusting the number of degrees of freedom of the χ^2 distribution to $KNS - p$ [this is based on a second-order Taylor expansion of $\mathcal{L}_Y(\hat{\theta})$ around $\mathcal{L}_Y(\theta)$ and taking expectation values]. The results for both the unadjusted and adjusted distribution of the log-likelihood are given in Fig. 7. In either case, the observed data fall well within the main part of the distribution and are thus similar to typical output of the model itself [$P(\mathcal{L} < \mathcal{L}_{\text{obs}}) = 0.11$ (uncorrected) or 0.16 (corrected)].

G. Optimal sampling

Can a judicious choice of sampling effort reduce parameter estimation and identification problems studied in Secs. III C–III E? Questions of optimal experimental design and optimal sampling have a long history in a variety of scientific fields [57,58]. Perhaps most pertinent to ecology is recent work in the study of HIV disease dynamics, in which one needs to choose which factors to measure and at what schedule [59], subject to cost and time constraints.

In our study of identifiability of model parameters above, we saw that difficulties in identifiability arise fundamentally from the presence of two sources of noise, error in the measurements and dynamical noise. We will illustrate the utility of ideas for optimal sampling by restricting ourselves to this elementary identification problem, the simultaneous estimation of measurement error and dynamical noise variances.

Central to most studies of optimal design is the so-called Fisher information matrix, $\mathcal{I} = \mathbb{E}(\mathbf{H})$, that is, the expected value of the Hessian matrix [36]. By the Cramér-Rao inequality, subject to some regularity conditions, \mathcal{I}^{-1} places a bound on the best expected covariance matrix for unbiased estimators [36] and thus tells us how much one can learn *at most* from a particular experiment about each parameter or parameter combination.

Here we will determine \mathcal{I} for $\log S^m$ and $\log S^d$ rather than S^m and S^d , since expected standard deviations ($\sqrt{(\mathcal{I}^{-1})_{ii}}$) can be interpreted as relative errors [as $\sigma(\log x) \approx \sigma(x)/x$ provided $\sigma(x) \ll x$]. We then look at the smallest eigenvector of \mathcal{I} , which acts as a bound on the *precision* (inverse variance) of the combination of $\log S^m$ and $\log S^d$ which is the *hardest to estimate*. To calculate \mathcal{I} , we first observe that the Hessian matrix can be calculated analytically when an analytical expression is available for the likelihood function. As in Sec. III F, we then reinterpret the likelihood function as the probability distribution for data given the model parameters and, hence, calculate both the distribution of \mathbf{H} under the model and its expectation value, \mathcal{I} .

For simplicity, we examine the case of a single species ($K = 1$), in a single ecosystem ($S = 1$), for which both intraspecies interactions (A_{ii}) and the common trend are known to be zero ($\phi = 0$). Since more measurements (N) and longer experimental duration (T) obviously improve estimation, we keep N and T constant. In this case, the likelihood function is given by the integrand of Eq. (12). As $A = 0$, $\Gamma_n = S^d \Delta t_n$, $B = 1$, and $C^{(0)} = 0$, the log-likelihood function simplifies to

$$\begin{aligned} \mathcal{L} &= -2 \log P(\mathbf{y}|\theta) \\ &= \log |\mathbf{Q}| - N \log(2\pi) + C - \mathbf{u}^\top \mathbf{Q}^{-1} \mathbf{u} \\ &= \log |\mathbf{Q}| + \mathbf{y}^\top \left(\frac{\mathbf{I}}{S^m} - \frac{\mathbf{Q}^{-1}}{(S^m)^2} \right) \mathbf{y} \\ &\quad + \sum_n \log(S^d \Delta t_n) + N \log(S^m), \end{aligned} \quad (26)$$

where we omitted parameter- and data-independent constants. Only the first two terms can contribute to \mathcal{I} , since the remaining terms are linear in $\log S^m$ and $\log S^d$ and have zero second derivatives. We calculate the second derivatives of $\log |\mathbf{Q}|$ and \mathbf{Q}^{-1} in the Supplemental Material [26], Sec. S5, but note that this is a simple task, since \mathbf{Q} can be written as $\mathbf{Q} = \frac{\tilde{\mathbf{Q}}}{S^d} + \frac{\mathbf{I}}{S^m}$, where $\tilde{\mathbf{Q}}$ depends on the measurement schedule but not on model parameters and \mathbf{I} is the identity matrix. We thus can study \mathcal{I} directly for different measurement schedules (through $\tilde{\mathbf{Q}}$) and values of S^d and S^m . Indeed, only the ratio of S^d and S^m matters, which we summarize in the dimensionless number $\gamma = S^d \bar{\Delta t} / S^m$, with $\bar{\Delta t} = T/N$ the average measurement interval. The resulting expression for the Fisher information matrix for $\log S^m$ and $\log S^d$ is

$$\mathcal{I} = \sum_{n=1}^N \frac{1}{(\lambda_n + \gamma)^2} \begin{pmatrix} \lambda_n^2 & \gamma \lambda_n \\ \gamma \lambda_n & \gamma^2 \end{pmatrix}, \quad (27)$$

where the λ_n are the eigenvalues of $\tilde{\mathbf{Q}}$. Calculation of \mathcal{I} for many random measurement schedules (not shown) suggested that optimal measurement schedules are periodic: optimal measurement schedules appeared to consist of bunches of

close-together measurements of various multiplicity, with the optimal multiplicity dependent on the value of γ . These “repeated sampling” schedules are reminiscent of the measurement approach proposed by Dennis *et al.* [60], based on disaggregation of data in the North American Breeding Bird Survey. (Here we should note some confusing terminology in the ecological literature. Dennis *et al.* [60] refer to such repeated sampling of an individual ecosystem as “replicated sampling,” but this should not be confused with the use of replicate ecosystems, and we will instead refer to the former as repeated sampling.) Knappe *et al.* [61] extended approximate maximum likelihood inference for such sampling schemes. Our formalism can handle repeated sampling without modification. Repeated sampling has been proposed in other scientific fields as well [58].

Since general optimization of measurement schedules is a highly nontrivial task, we restrict ourselves further to periodic schedules. Within these schedules, measurements occur in bunches of the same multiplicity, L , separated by small intervals ε . For simplicity, we measure time in units of the average measurement interval (that is, $\overline{\Delta t} = 1$) without loss of generality. We illustrate such measurement schedules in the diagram in Fig. 8(c). Within each bunch, measurements are numbered $1, \dots, \ell$.

We illustrate the achievable gains in precision [according to Eq. (27)] as a function of L and ε for $N = 120$ and $\gamma = 0.1$ [dotted lines, Fig. 8(a)] and $\gamma = 3$ [Fig. 8(b)]. In each case, the reference is measurement at equally spaced time points. For both values of γ , substantial gains in precision relative to the equally spaced reference can be achieved by sampling repeatedly, although different multiplicities of the measurements are optimal for different values of γ . For comparison, single species estimates of γ for our data are, for the most densely sampled time series, about 1.1 for *C. reinhardtii*, 2.1 for *E. coli*, and 2.9 for *T. thermophila*.

Figures 8(a) and 8(b) also shows that the optimal spacing within a bunch is often as large as possible (that is, it reduces to equal spacing) or as small as possible ($\varepsilon^* = 0$). However, for larger values of L and γ , intermediate optimal ε are seen [Fig. 8(c), thickened line segments].

For measurement schedules with many time points, it may become inefficient to calculate the optimal schedule this way. Instead, one can also calculate \mathcal{I} in the limit of long time series ($T, N \rightarrow \infty$, while T/N constant). In this case, the periodicity of the measurement schedule can be exploited to easily calculate \mathcal{I} . Since $\tilde{\mathbf{Q}}$ is periodic, we can use a tool from condensed matter physics, Bloch’s theorem [62], to reduce the complexity of finding its eigenvectors and eigenvalues from an

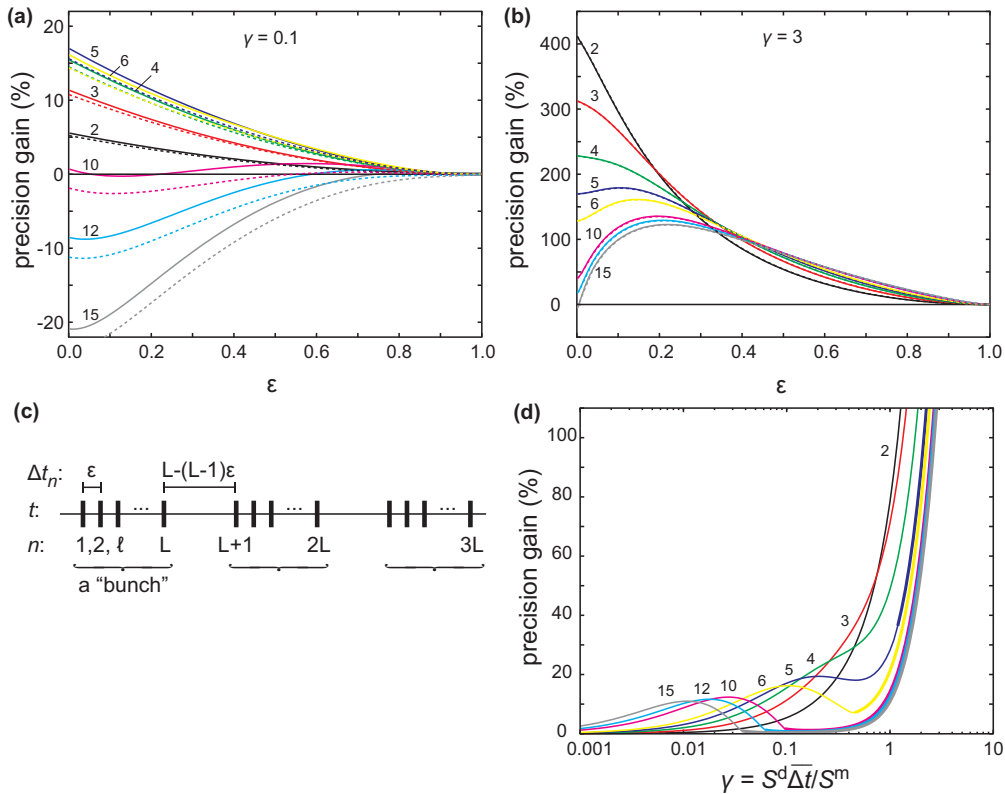


FIG. 8. (Color online) Optimal experimental design. Expected gain in precision (for the least precise eigenvector of the Fisher information matrix, \mathcal{I}) for different measurement schedules relative to equally spaced measurements. All curves are for the same average measurement rate, $\overline{\Delta t} = T/N$. Results are a function of the ratio between dynamical and measurement noise, $\gamma = S^d \overline{\Delta t} / S^m$. [(a) and (b)] Gain in precision as a function of the spacing between measurements, ε , in bunches of different multiplicity, L , for $\gamma = 0.1$ (a) and $\gamma = 3$ (b) [L labeled on curves; dotted: $N = 120$, solid: asymptotic result; the two nearly coincide in panel (b)]. (c) Illustration of the “repeated sampling” schedule discussed in the text. Thick vertical bars represent measurement times. (d) Gain in precision at optimal ε^* as a function of γ for each multiplicity L (as labeled on each curve). Along each curve, segments at which ε^* intermediate ($0 < \varepsilon^* < 1$) are indicated by thickening of the curve (within the plot area this occurs only for $L \geq 5$ at large γ).

$NL \times NL$ problem to an $L \times L$ problem (see the Supplemental Material [26], Sec. S5 for details).

As the number N of measurements grows, we find that $\mathcal{I} = N\mathcal{I}_0$, with \mathcal{I}_0 the Fisher information matrix per measurement point as follows:

$$\mathcal{I}_0 = \frac{1}{\pi L} \sum_{\ell=1}^L \int_0^\pi \frac{d\varphi}{(\gamma + \lambda_\ell(\varphi))^2} \begin{pmatrix} \lambda_\ell(\varphi)^2 & \gamma \lambda_\ell(\varphi) \\ \gamma \lambda_\ell(\varphi) & \gamma^2 \end{pmatrix}. \quad (28)$$

This is analogous to Eq. (27). Here, $\lambda_\ell(\varphi)$, with $\ell = 1, \dots, L$, denotes the eigenvalues of the $L \times L$ matrix $\tilde{Q}(\varphi)$, which is an $L \times L$ submatrix of \tilde{Q} modified at its corner elements, as follows:

$$[\tilde{Q}(\varphi)]_{\ell,\ell'} = \begin{cases} \frac{1}{\Delta t_\ell} + \frac{1}{\Delta t_{\ell+1}} & \text{if } \ell = \ell' \leq L-1 \\ \frac{1}{\Delta t_L} + \frac{1}{\Delta t_1} & \text{if } \ell = \ell' = L \\ -\frac{1}{\Delta t_\ell} & \text{if } \ell = \ell' + 1 \\ -\frac{1}{\Delta t_{\ell+1}} & \text{if } \ell = \ell' - 1 \end{cases} - \delta_{\ell,L} \delta_{\ell',1} \frac{e^{-i\varphi}}{\Delta t_1} - \delta_{\ell,1} \delta_{\ell',L} \frac{e^{i\varphi}}{\Delta t_1}, \quad (29)$$

where $i^2 = -1$. The proposed periodic repeated sampling measurement schedules have $\Delta t_\ell = \varepsilon$ for $\ell = 1, \dots, L-1$ and $\Delta t_L = L - (L-1)\varepsilon$. Differences in precision gain between the asymptotic result and the finite time series results tend to be small even for modest N [Figs. 8(a) and 8(b), dotted versus solid lines].

Using our asymptotic results, we show in Fig. 8(d) the gain in precision for different sampling multiplicities (in each case at their optimal $\varepsilon = \varepsilon^*$) as a function of γ . Obviously, one typically does not know γ precisely at the start of an experiment. Nevertheless, it is clear that low multiplicities ($L = 2-4$) yield robust gains in precision over a large range of γ , mostly for $\varepsilon^* \approx 0$. Temporal autocorrelation in the measurement errors can make the optimal ε larger than 0 and decrease the achievable gain in precision.

Finally, we note that, since the full likelihood function is available for multiple interacting species, replicate systems, and arbitrary common trend (Tables I and II), one can derive and optimize the eigenvalues of the Fisher information, \mathcal{I} , for more complex dynamics and measurement schedules. For example, while in the above case the Fisher information is simply additive over replicate systems, we expect this to be no longer the case if the common trend is unknown. As a consequence, it might be optimal, at constant total effort, to sample different systems in different ways. We also studied the case of single species dynamics with $S_{11}^d > 0$, $A_{11} > 0$, and $S_1^m = 0$ in a single replicate with known average dynamics. In this case, our analysis suggested little room for improvement over equally spaced measurements, provided $A_{11} \ll 1/\Delta t$ (results not shown).

IV. DISCUSSION AND CONCLUSIONS

Study of the stochastic nature of any complex system requires accounting for possible idiosyncratic trends in the data. For ecological systems such trends are clearly pervasive. Replication of ecosystems allows the exciting possibility that one can actually study variability in ecological dynamics, but one also needs flexible but simple models capable of

accommodating an idiosyncratic trend. Indeed, for our experimental data on a set of replicate closed ecosystems, we confirm this notion: Flexibility in the common trend described by the model is necessary to obtain good estimates of effective species interactions characterizing the variability in dynamics [Fig. 4(a)].

Specifically, we use a linear, Gaussian, continuous-time Gompertz model to illustrate the challenges one faces fitting models to replicate ecological time series. We emphasize that this is a phenomenological model and its parameters are effective ones. Changes in model structure, for example, in the functional form of density dependence or the time dependence of parameters, could lead to different estimates of *all* parameters. Without doubt, alternative model structures may be more appropriate for some other ecosystems. Conceivably, too, different approaches may be required, e.g., for strongly cyclical dynamics in which the main mode of variability might be random variation in the phase of the cycles or dynamical noise is intermittent.

A central difficulty in fitting replicate time series is the description of the common trend. Here we chose a segment-by-segment description with time dependence in the trend variables f and a smoothing parameter μ to limit the overall complexity of its description. There are several reasonable choices of the smoothing parameter, and we advise exploring inference over a range of μ . As shown in Fig. 4(a), the common choice, constant coefficients ($\mu \rightarrow \infty$), can lead to significant error, and we have proposed ways to make a better choice. We anticipate that more sophisticated choices of smoothing parameter, and perhaps a description of the common trend altogether, will be developed.

While we believe that much can be learned from a simple, linear model, we briefly mention here a number of modifications and extensions of our work.

(1) One can make alternative choices for prior distributions and still apply the methods here, provided the overall probability of the data can still be integrated over x and ϕ .

(2) We note that our analysis places no restrictions on measurement schedule. Different replicates can be sampled at different time points, and time intervals do not need to be identical. Indeed, the analysis of our data was based on a more general form of our algorithm than presented in the main text, accommodating arbitrary measurement schedules as explained in the Supplemental Material [26], Sec. S3, and avoiding “missing data” problems.

(3) One can extend our approach to the comparison of sets of replicates each with their own trend. We discuss briefly how to do this in the Supplemental Material [26].

(4) We have described our approach for linear stochastic differential equations since stochastic differential equations (rather than difference equations) have received little attention in the ecological literature and were suitable for our data. It is straightforward to apply our methods to linear difference equations instead. One simply replaces Eqs. (6) and (7) with the new dynamical model.

(5) It is also straightforward to include delays (e.g., ARMA models, Refs. [10,24]), life cycle (stage) or age structure. Such models can, in principle, alleviate the assumption of a Markov property for the dynamics. Not surprisingly, however, such models suffer from significant identifiability problems and

complicated, multimodal likelihood surfaces [63], even for a single species.

(6) It is trivial to include covariates with known time dependence into the analysis, as long as they are additive, or linear in the logarithmic densities (cf. Ref. [19]).

(7) One can separate dynamical noise into a “global” component affecting all replicate systems and per-system components. This would simply lead to cross-terms $Q^{(s,s')}$ between systems but not affect the overall mathematical structure.

(8) We conjecture that one can approximately accommodate some nonlinear models in the present framework. For example, formulated in terms of logarithmic densities, the Lotka-Volterra model contains a nonlinear term e^x , which is quite readily linearized. Likewise, replacing log-normal measurement error with Gaussian measurement error simply means multiplication of the error terms by e^{-x} . One can also approximate alternative stochastic components using linear combinations of Gaussian distributions (“mixtures”). In any case, from a statistical perspective, linear Gaussian models provide a good starting point, in comparison to which nonlinear contributions can be tested for their significance.

(9) It would be interesting to compare our methodology further to methods from econometrics based on vector autoregressive models.

We want to make a few closing remarks. First, rarely is one interested in a list of parameter values per se but rather in system properties like the robustness to perturbations, the sources, size and time scale of fluctuations, or prediction of extinction dynamics. As we have shown, inference of the full joint posterior distribution of the model parameters is essential to this end. We anticipate that future population dynamics experiments may well differ markedly from past practice. They may include an *a priori* or “online” (i.e., as measurements come in) analysis of the identifiability of the system properties of interest. This analysis can steer the choice of initial conditions, system perturbations, and measurement schedule [64], enabling more efficient focus on the ecosystem properties or processes under study. More fundamentally, even “simple” laboratory ecosystems display a rich interplay between physical, metabolic, phenotypic, spatial, and population dynamics. This multitude of relevant variables and associated length and time scales raises the question whether differential equations, or, more generally, conventional population dynamics models, can at all be applied to data in a meaningful manner. Perhaps entirely different statistical and conceptual tools are necessary. At the same time, we believe that replicable model systems are necessary for ecology. Such model systems may not match the complexity present in natural ecosystems. They could, however, be our only hope of obtaining the conditions and data necessary for making fundamental progress.

-
- [1] D. R. Hekstra and S. Leibler, *Cell* **149**, 1164 (2012).
- [2] C. E. Folsome and J. A. Hanson, in *Ecosystem Theory and Applications*, edited by N. Polunin (John Wiley & Sons, New York, 1986).
- [3] F. B. Taub, *Ann. Rev. Ecol. Syst.* **5**, 139 (1974).
- [4] G. F. Fussmann, S. P. Ellner, K. W. Shertzer, and N. G. Hairston, Jr., *Science* **290**, 1358 (2000).
- [5] L. Becks, S. P. Ellner, L. E. Jones, and N. G. Hairston, Jr., *Ecol. Lett.* **13**, 989 (2010).
- [6] A. G. Fredrickson, *Annu. Rev. Microbiol.* **31**, 63 (1977).
- [7] J. J. Faith, N. P. McNulty, F. E. Rey, and J. I. Gordon, *Science* **333**, 101 (2011).
- [8] H. A. Gordon and L. Pesti, *Bacteriol. Rev.* **35**, 390 (1971).
- [9] R. P. Freckleton, A. R. Watkinson, R. E. Green, and W. J. Sutherland, *J. Anim. Ecol.* **75**, 837 (2006).
- [10] P. Brockwell and R. Davis, *Introduction to Time Series and Forecasting*, Springer Texts in Statistics (Springer-Verlag, New York, 2002).
- [11] J. S. Clark and O. N. Bjørnstad, *Ecology* **85**, 3140 (2004).
- [12] G. M. Wang, *Ecol. Inform.* **4**, 69 (2009).
- [13] P. De Valpine and A. Hastings, *Ecolog. Monogr.* **72**, 57 (2002).
- [14] B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples, *Ecol. Monogr.* **76**, 323 (2006).
- [15] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter, *Stat. Comput.* **10**, 325 (2000).
- [16] Z. Ghahramani, *Lect. Notes Comput. Sci.* **1387**, 168 (1998).
- [17] R. E. Kalman, *Transact. ASME J. Basic Eng. (Series D)* **82**, 35 (1960).
- [18] H. E. Rauch, F. Tung, and C. T. Striebel, *AIAA J.* **3**, 1445 (1965).
- [19] A. R. Ives, B. Dennis, K. L. Cottingham, and S. R. Carpenter, *Ecol. Monogr.* **73**, 301 (2003).
- [20] L. Polansky, P. de Valpine, J. O. Lloyd-Smith, and W. M. Getz, *Ecology* **90**, 2313 (2009).
- [21] J. M. Fryxell, I. M. Smith, and D. H. Lynn, *Oikos* **111**, 143 (2005).
- [22] J. Y. Humbert, L. S. Mills, J. S. Horne, and B. Dennis, *Oikos* **118**, 1940 (2009).
- [23] B. Dennis, P. L. Munholland, and J. M. Scott, *Ecol. Monogr.* **61**, 115 (1991).
- [24] N. L. Ziebarth, K. C. Abbott, and A. R. Ives, *Ecol. Lett.* **13**, 21 (2010).
- [25] T. J. Russell, D. W. Lwetoijera, B. G. J. Knols, W. Takken, G. F. Killeen, and H. M. Ferguson, *Proc. R. Soc.* **278**, 3142 (2011).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.88.062714> for detailed mathematical derivations referred to in the text, treatment of the general case of multiple time series with different, arbitrary measurement schedules, a full specification of prior probabilities, and a supplementary figure.
- [27] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, and K. Van't Riet, *Appl. Environ. Microbiol.* **56**, 1875 (1990).
- [28] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (Elsevier, Amsterdam, 1992).
- [29] A. R. Ives, *Ecol. Monogr.* **65**, 217 (1995).
- [30] D. T. Gillespie, *Am. J. Phys.* **64**, 225 (1996).

- [31] S. Särkkä, Ph.D. thesis, Helsinki University of Technology Laboratory of Computational Engineering, 2006.
- [32] C. H. Reinsch, *Numer. Math.* **10**, 177 (1967).
- [33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in.*, 2nd ed. (Cambridge University Press, Cambridge, UK, 1992).
- [34] J. C. Pinheiro and J. M. Bates, *Stat. Comput.* **6**, 289 (1996).
- [35] D. J. C. MacKay, *Mach. Learn.* **33**, 77 (1998).
- [36] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory* (Prentice Hall, Upper Saddle River, NJ, 1993).
- [37] B. Efron, *Ann. Stat.* **10**, 323 (1982).
- [38] A. Foi, *Optimization of Variance-Stabilizing Transformations* (2009), available online at <http://www.cs.tut.fi/~foi/optvst/>.
- [39] S. R. Lele, B. Dennis, and F. Lutscher, *Ecol. Lett.* **10**, 551 (2007).
- [40] J. Knappe, *Ecology* **89**, 2994 (2008).
- [41] P. Shaman and R. A. Stine, *J. Am. Stat. Assoc.* **83**, 842 (1988).
- [42] E. Paradis, S. R. Baillie, W. J. Sutherland, and R. D. Gregory, *Ecology* **81**, 2112 (2000).
- [43] M. Kéry, R. M. Dorazio, L. Soldaat, A. Van Strien, A. Zuiderwijk, and J. A. Royle, *J. Appl. Ecol.* **46**, 1163 (2009).
- [44] R. E. Kass and L. Wasserman, *J. Am. Stat. Assoc.* **90**, 928 (1995).
- [45] R. F. Engle and C. W. J. Granger, *Econometrica* **55**, 251 (1987).
- [46] V. A. Marčenko and L. A. Pastur, *Math. USSR-Sbornik* **1**, 457 (1967).
- [47] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, *PLoS Comput. Biol.* **3**, 1871 (2007).
- [48] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, *Science* **342**, 604 (2013).
- [49] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, *Int. J. Theor. Appl. Finan.* **3**, 391 (2000).
- [50] R. E. Kass and A. E. Raftery, *J. Am. Stat. Assoc.* **90**, 773 (1995).
- [51] G. Schwarz, *Ann. Stat.* **6**, 461 (1978).
- [52] C. H. Wiggins and I. Nemenman, *Exp. Mech.* **43**, 361 (2003).
- [53] D. V. Lindley, *Biometrika* **44**, 187 (1957).
- [54] E. D. Green and K. J. Elgersma, (2010), 95th ESA meeting PS 48-179, <http://eco.confex.com/eco/2010/webprogram/Paper25724.html>.
- [55] T. Cover and J. Thomas, *Elements of Information Theory* (John Wiley & Sons, Inc., New York, 1991).
- [56] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van der Linde, *J. R. Stat. Soc. Ser.* **64**, 583 (2002).
- [57] S. D. Silvey, *Optimal Design: An Introduction to the Theory for Parameter Estimation* (Chapman & Hall, London, 1980).
- [58] L. Pronzato, *Automatica* **44**, 303 (2008).
- [59] J. Guedj, R. Thiébaud, and D. Commenges, *Bull. Math. Biol.* **69**, 2493 (2007).
- [60] B. Dennis, J. M. Ponciano, and M. L. Taper, *Ecology* **91**, 610 (2010).
- [61] J. Knappe, P. Besbeas, and P. De Valpine, *Ecology* **94**, 2097 (2013).
- [62] F. Bloch, *Zeitschr. Phys.* **52**, 555 (1929).
- [63] A. R. Ives, K. C. Abbott, and N. L. Ziebarth, *Ecology* **91**, 858 (2010).
- [64] L. Paninski, J. Pillow, and J. Lewi, *Prog. Brain Res.* **165**, 493 (2007).