Supplementary Material, Hekstra et al.

Note: references are also listed as part of the main text.

S1 Matrix fraction decomposition

We can rewrite the dynamics in eq. 2 as $\mathrm{d}\boldsymbol{x} = \boldsymbol{f}\mathrm{d}t - \mathbf{A}\boldsymbol{x}\mathrm{d}t + \boldsymbol{\eta}\sqrt{\mathrm{d}t}$. We are interested in obtaining an analytical expression for the covariance matrix of $\boldsymbol{x}_s$, $\mathbf{C} = \left\langle \boldsymbol{x}\boldsymbol{x}^\top \right\rangle$ after propagating the dynamics for a time interval Δt. (Note that we will omit the explicit dependence on the system index, s.) To this end, we first derive a differential equation for $\mathbf{C}$ itself.

Since $\boldsymbol{f}$ only affects the average dynamics, we will omit it without loss of generality: it does not affect the dynamics of the covariance matrix. Since $\mathrm{d}\boldsymbol{x} = \boldsymbol{x}(t+\mathrm{d}t) - \boldsymbol{x}(t)$,

$$\boldsymbol{x}(t+\mathrm{d}t) = (\mathbf{I} - \mathbf{A}\mathrm{d}t)\,\boldsymbol{x}(t) + \boldsymbol{\eta}\sqrt{\mathrm{d}t}.$$

We multiply both sides by their transpose and take expectation values, yielding

$$\begin{aligned}\left\langle \boldsymbol{x}\boldsymbol{x}^\top \right\rangle =& (\mathbf{I} - \mathbf{A}\mathrm{d}t)\left\langle \boldsymbol{x}\boldsymbol{x}^\top \right\rangle (\mathbf{I} - \mathbf{A}\mathrm{d}t)^\top + \left\langle \boldsymbol{\eta}\boldsymbol{\eta}^\top \right\rangle \mathrm{d}t + \\ & (\mathbf{I} - \mathbf{A}\mathrm{d}t)\left\langle \boldsymbol{x}\boldsymbol{\eta}^\top \right\rangle \sqrt{\mathrm{d}t} + \left\langle \boldsymbol{\eta}\boldsymbol{x}^\top \right\rangle (\mathbf{I} - \mathbf{A}\mathrm{d}t)^\top \sqrt{\mathrm{d}t},\end{aligned}$$

where we have suppressed the time-dependence of $\left\langle \boldsymbol{x}\boldsymbol{x}^\top \right\rangle$ for clarity. This expression simplifies since cross-terms $\left\langle \boldsymbol{\eta}\boldsymbol{x}^\top \right\rangle = 0$ and terms of order $(\mathrm{d}t)^2$ vanish. Defining $\mathbf{C} \equiv \left\langle \boldsymbol{x}\boldsymbol{x}^\top \right\rangle$ and with $\mathrm{d}\mathbf{C} = \mathbf{C}(t+\mathrm{d}t) - \mathbf{C}(t)$, we arrive at

$$\mathrm{d}\mathbf{C} = -\mathbf{A}\mathbf{C}\mathrm{d}t - \mathbf{C}\mathbf{A}^\top \mathrm{d}t + \mathbf{S}^{\mathrm{d}}\mathrm{d}t. \tag{S1-1}$$

This differential equation can be solved by a method called *matrix fraction decomposition* [8]. The first step is to propose that $\mathbf{C}$ can be written as a "ratio" of two matrices:

$\mathbf{C} = \mathbf{D} \cdot \mathbf{E}^{-1}$. Application of the chain rule of differentiation for matrices yields:

$$\mathrm{d}\mathbf{C} = \mathrm{d}\mathbf{D} \cdot \mathbf{E}^{-1} - \mathbf{D}\mathbf{E}^{-1}(\mathrm{d}\mathbf{E})\mathbf{E}^{-1}\,. \tag{S1-2}$$

We equate equations (S1-1) and (S1-2), divide by $\mathrm{d}t$ and right-multiply by $\mathbf{E}$:

$$\dot{\mathbf{D}} - \mathbf{D}\mathbf{E}^{-1}\dot{\mathbf{E}} = -\mathbf{A}\mathbf{D} - \mathbf{D}\mathbf{E}^{-1}\mathbf{A}^{\top}\mathbf{E} + \mathbf{S}^{\mathrm{d}}\mathbf{E},$$

which is solved by a pair of differential equations:

$$\begin{pmatrix}\dot{\mathbf{D}}\\ \dot{\mathbf{E}}\end{pmatrix} = \begin{pmatrix}-\mathbf{A} & \mathbf{S}^{\mathrm{d}}\\ 0 & \mathbf{A}^{\top}\end{pmatrix}\begin{pmatrix}\mathbf{D}\\ \mathbf{E}\end{pmatrix}, \tag{S1-3}$$

Up to this point, we have left $\mathbf{D}(t)$ and $\mathbf{E}(t)$ unspecified. A convenient choice is to describe the initial condition as $\mathbf{D}(t) = \mathbf{C}(t)$ and $\mathbf{E}(t) = \mathbf{I}$. We hence arrive at:

$$\begin{pmatrix}\mathbf{D}(t+\Delta t)\\ \mathbf{E}(t+\Delta t)\end{pmatrix} = \exp\left[\begin{pmatrix}-\mathbf{A} & \mathbf{S}^{\mathrm{d}}\\ 0 & \mathbf{A}^{\top}\end{pmatrix}\Delta t\right]\begin{pmatrix}\mathbf{C}(t)\\ \mathbf{I}\end{pmatrix}. \tag{S1-4}$$

and $\mathbf{C}(t+\Delta t) = \mathbf{D}(t+\Delta t) \cdot \mathbf{E}(t+\Delta t)^{-1}$. Now, equation 2.7 can be evaluated. $\Gamma_n \equiv \mathrm{Cov}\left(\boldsymbol{x}_{n+1,s}\,|\boldsymbol{x}_{ns}, \boldsymbol{\theta}\right)$, being conditional on the preceding $\boldsymbol{x}_{ns}$, is the solution for $\mathbf{C}(t) = \mathbf{0}$.

Using the vec operator, which converts a matrix into a column vector consisting of its stacked columns, one can write the asymptotic covariance matrix as [5]:

$$\mathrm{vec}(\mathbf{C}(t \to \infty)) = \mathcal{A}^{-1}\mathrm{vec}(\mathbf{S}^{\mathrm{d}}) \tag{S1-5}$$

with $\mathcal{A} = \mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I}$, provided the dynamics are, asymptotically, stationary (that is all eigenvalues of $\mathbf{A}$ have absolute size < 1).

S2 Specification of prior probabilities

We need to specify prior distributions for the trend parameters $\boldsymbol{\phi}$ and the model parameters: the interaction matrix $\mathbf{A}$, the dynamical noise covariance matrix $\mathbf{S}^{\mathrm{d}}$ and the measurement noise variances S_k^{m}. There is a substantial literature on the choice of priors. We base our choices loosely on a few principles commonly used. Many other approaches have been proposed in the literature and it is straightforward to replace the choices made here with alternatives.

First, we propose that model parameters are *a priori independent*. That is, we can determine separate priors for $\mathbf{A}, \mathbf{S}^{\mathrm{d}}$ and S_k^{m} and multiply them. The prior on the trend, may depend on the values of the model parameters.

Secondly, we seek approximately *"non-informative"* priors [2]. Such priors result in a Bayesian likelihood function which is only translated (shifted), not reshaped, for different data sets of the same size. This principle leads to locally uniform priors for so-called "location parameters", such as the mean of a distribution, and for "scale parameters", such as a standard deviation, it leads to prior distributions locally uniform in the logarithm of the parameter.

Finally, priors are often constrained to be *"conjugate" priors*. A prior distribution for a parameter is conjugate if the posterior and the prior distribution have the same functional dependence on that parameter. This constraint is mostly a matter of convenience, although it can be used to express a notion of the amount of information in the prior distribution relative to the data [6]. Additional parameters characterizing prior distributions are known as *hyperparameters*.

Prior distribution on A. The parameters A_{ij} are approximately location parameters (this becomes exact in the limit of weak interactions A_{ij}), suggesting a Gaussian prior distribution, which is nearly a conjugate distribution for weak interactions. From experiments, one cannot meaningfully infer couplings stronger than $1/(\Delta t)_{\min}$, setting a natural scale for the A_{ij}. That is, we take

$$P(\mathbf{A}) = \left(\frac{\Delta t_{\min}}{\sqrt{2\pi}}\right)^{K^2} \exp\left(-\frac{1}{2}\Delta t_{\min}^2 \mathrm{Tr}(\mathbf{A}\mathbf{A}^\top)\right), \tag{S2-1}$$

where K is the number of species.

Prior distribution on S^{m}. Our measurements of local density are based on direct counts of individual organisms. Hence counting statistics place a natural lower limit on the measurement variances (time points with the highest numbers of counts give minimum error estimates of $\sim 1\%$ for *C. reinhardtii* and *E. coli*, and $\sim 3\%$ for *T. thermophila*). We chose a conventional prior distribution on the S_k^{m}, the inverse Wishart distribution, which allows us to impose a soft lower bound on S_k^{m}, and which decays approximately exponentially for large $\log S_k^{\mathrm{m}}$.

$$P(\{S_k^{\mathrm{m}}\}) = \prod_{k=1}^{K} \left(\frac{\psi_k}{2s_k}\right)^{m/2} \frac{e^{-\psi_k/2s_k}}{s_k \Gamma_1\left(\frac{m}{2}\right)} \tag{S2-2}$$

where for simplicity, $s_k = S_k^{\mathrm{m}}$, and Γ_1 represents the Gamma function. We set $m = 1$ and $\psi_k = \frac{2}{3} S_{k,\min}^{\mathrm{m}}$ for each species.

Prior distribution on $\mathbf{S}^{\mathrm{d}}$. Our basic consideration for the diagonal elements $\{S_{kk}^{\mathrm{d}}\}$ is the same as for the S_k^{m}: there is a lower bound on the dynamical noise, as a divergence

$S^{\mathrm{d}}_{kk}T < S^{\mathrm{m}}_{k,\min}$ cannot be meaningfully inferred (hence $S^{\mathrm{d}}_{kk,\min} = S^{\mathrm{m}}_{k,\min}/T$). Indeed, we parametrize $\mathbf{S}^{\mathrm{d}}$ for the purpose of specifying its prior by $\{S^{\mathrm{d}}_{kk}, \rho^{\mathrm{d}}_{kl}\}$, with $k \neq l$. We again use the inverse Wishart distribution to describe the S^{d}_{kk}, with $m = 1$ and $\psi_{kk} = \frac{2}{3}S^{\mathrm{d}}_{kk,\min}$. We simply use a uniform prior on [-1,1] for each ρ^{d}_{kl}. That is,

$$P(\mathbf{S}^{\mathrm{d}}) = P(\{S^{\mathrm{d}}_{kk}, \rho^{\mathrm{d}}_{kl}\}) = \frac{1}{2^{\frac{K(K-1)}{2}}} \prod_{k=1}^{K} \left(\frac{\psi_{kk}}{2s_{kk}}\right)^{m/2} \frac{e^{-\psi_{kk}/2s_{kk}}}{s_{kk}\Gamma_1\left(\frac{m}{2}\right)} \tag{S2-3}$$

with $s_{kk} = S^{\mathrm{d}}_{kk}$. Note that this prior is not entirely proper, since it attaches some weight to matrices $\mathbf{S}^{\mathrm{d}}$ which are not proper covariance matrices. For alternative choices, see e.g. ref. [1].

Prior distribution on $\boldsymbol{\phi}$. We will motivate our choice of prior distribution for the trend parameters $\boldsymbol{\phi}$ in the limit of weak interactions, $A_{kl} \leq 1/T$, although the prior can be used equally well in the presence of stronger interactions. When interactions are weak, $\boldsymbol{\phi} \approx \boldsymbol{f}$ (see main text) and $\boldsymbol{\phi}$ can directly be considered the derivative of the common trend of the densities, $\boldsymbol{F}(t) = \int_0^t \boldsymbol{f}(t')\mathrm{d}t'$. That is, the expected average dynamics, $\langle \boldsymbol{x}(t) \rangle = \boldsymbol{x}_0 + \boldsymbol{F}(t)$.

The prior distribution on $\boldsymbol{\phi}$ describes in this case how much curvature we expect in the average dynamics. Specifically, our prior distribution assumes that trends displaying more curvature are *a priori* less likely. The curvature of $F(t)$ is approximately given by,

$$\text{curvature} \approx \sum_{m,k} \frac{(\phi_{m+1,k} - \phi_{mk})^2}{\Delta t_{m+1} + \Delta t_m} \tag{S2-4}$$

Where, if each replicate ecosystem was sampled at the same schedule, $t_m = t_n$; otherwise the t_m form the union of all t_{ns}: $\{t_m\} = \bigcup_s \{t_{ns}\}$ as in Fig. S-1. If one has an expected pattern for the trend, $\boldsymbol{\phi}^0 = \{\phi^0_{mk}\}$, eq. (S2-4) can be modified to the curvature around the expected trend:

$$\text{curvature} \approx \sum_{m} \frac{\left((\phi_{m+1,k} - \phi^0_{m+1,k}) - (\phi_{mk} - \phi^0_{mk}\right)^2}{\Delta t_{m+1} + \Delta t_m} \tag{S2-5}$$

In our analysis we set $\boldsymbol{\phi}^0 = \mathbf{0}$. Since $\boldsymbol{\phi}$ is a vector of location parameters, a Gaussian distribution is a suitable, conjugate prior distribution. We thus take as our initial prior simply:

$$P(\boldsymbol{\phi}) \propto \exp\left(-\frac{\mu^2}{2} \sum_{m,k} \frac{\left((\phi_{m+1,k} - \phi^0_{m+1,k}) - (\phi_{mk} - \phi^0_{mk})\right)^2}{\Delta t_{m+1} + \Delta t_m}\right) \tag{S2-6}$$

However, this criterion is by itself insufficient to formulate a proper prior distribution for $\boldsymbol{\phi}$, since it leaves one degree of freedom per species unconstrained. To address this, we multiply eq. (S2-6) by a factor tying each ϕ_{mk} weakly to its expected value, ϕ^0_{mk}, with a standard

deviation $\sigma_0 T/\Delta t_m \geq \sigma_0$ and $\sigma_0 = 0.4$ for each species, based on the size of growth rate fluctuations observed over the data. Putting these components together yields:

$$P(\boldsymbol{\phi}) = \frac{\exp\left(-\frac{\mu^2}{2}(\boldsymbol{\phi}-\boldsymbol{\phi}^0)^\top \mathbf{D}(\boldsymbol{\phi}-\boldsymbol{\phi}^0)\right)}{\sqrt{(2\pi)^{(M-1)K}|\mu^2\mathbf{D}^{-1}|}} \tag{S2-7}$$

with

$$D_{mk,m'k'} = \begin{cases} \frac{\mathbb{1}_{(m>1)}}{\Delta t_{m-1}+\Delta t_m} + \frac{\mathbb{1}_{(m<M-1)}}{\Delta t_m+\Delta t_{m+1}} + \left(\frac{\Delta t_m}{\sigma_0 T}\right)^2 & \text{for} \quad k=k' \wedge m=m' \\ \frac{-1}{\Delta t_m+\Delta t_{m'}} & \text{for} \quad k=k' \wedge m=(m'\pm 1) \\ 0 & \text{for} \quad k\neq k' \vee (|m-m'|>1) \end{cases} \tag{S2-8}$$

Finally, when considering a "unit information prior" on $\boldsymbol{\phi}$, we limit the influence of the prior distribution to the equivalent of a single time series. Doing so induces a dependence of the trend prior distribution on the model parameters (but no direct dependence on the data). Analogous to ref. [6], that is:

$$P(\boldsymbol{\theta},\boldsymbol{\phi}) = P(\boldsymbol{\theta})\cdot P(\boldsymbol{\phi}|\boldsymbol{\theta}) = P(\boldsymbol{\theta})\cdot P(\boldsymbol{\phi}|\mu(\boldsymbol{\theta})) \tag{S2-9}$$

indicating that any dependence of the trend prior distribution on the model parameters is mediated by μ. We explain this approach in the main text.

S3 Linear model for time series with different measurement schedules

The analysis of the (linear) Gompertz model in the main text assumed that all systems were measured at the same schedule. In general, as well as for our data, each replicate system can have its own set of measurement time points $\{t_{ns}\}$. These time points do not need to be evenly spaced (that is, each $\Delta t_{ns} = t_{n+1,s} - t_{ns}$ can be different), and different systems can be measured a different number of times (that is, each system has its own N_s). We describe here inference for this more general case, following the approach in the main text. For a detailed motivation of each step, we refer the reader to the main text.

In addition, we will (i) discuss how to fit different trends to different subsets of the available systems, (ii) describe how to extract the posterior distribution for the underlying "true" dynamics and the common trend, (iii) propose a way to quantify the effective number of degrees of freedom in the trend, and (iv) derive a criterion for when the optimal μ is finite or infinite when optimized directly.

As a reminder, the state space model for the Gompertz model takes the form

$$\dot{\boldsymbol{x}}_s = \boldsymbol{f}(t) - \mathbf{A}\boldsymbol{x}_s(t) + \boldsymbol{\eta}_s \tag{S3-1a}$$

$$\boldsymbol{y}_{ns} = \boldsymbol{x}_s(t_n) + \boldsymbol{\xi}_{ns} \tag{S3-1b}$$

We will use the shorthand $\boldsymbol{x}_{ns}$ for $\boldsymbol{x}_s(t_n)$ from here on and omit explicit mention of the time dependence of $\boldsymbol{f}$ and $\boldsymbol{x}$.

The probability of the data given the model, $P(\boldsymbol{y}|\boldsymbol{\theta},\mu)$ plays a central role in both a likelihood approach and a Bayesian approach. It takes the form

$$P(\boldsymbol{y} \mid \boldsymbol{\theta};\mu) = \int \mathrm{d}\boldsymbol{\phi} \int \mathrm{d}\boldsymbol{x} \left(P\left(\boldsymbol{y}\,|\boldsymbol{x}\right) \cdot P\left(\boldsymbol{x}\,|\boldsymbol{\phi},\boldsymbol{\theta}\right) \cdot P\left(\boldsymbol{\phi}\,|\boldsymbol{\theta},\mu\right)\right). \tag{S3-2}$$

with

$$\begin{aligned} P\left(\boldsymbol{x}\,|\boldsymbol{\phi},\boldsymbol{\theta}\right) &= \prod_{s=1}^{S}\prod_{n=1}^{N_s-1} P(\boldsymbol{x}_{n+1,s}|\boldsymbol{x}_{ns},\boldsymbol{\phi},\boldsymbol{\theta}) \\ &= \exp\Big\{-\frac{1}{2}\sum_{s=1}^{S}\sum_{n=1}^{N_s-1}\left(\boldsymbol{x}_{n+1,s} - \mathbf{B}_{ns}\boldsymbol{x}_{ns} - \boldsymbol{\phi}_{ns}\Delta t_{ns}\right)^{\top}\boldsymbol{\Gamma}_{ns}^{-1}\left(\boldsymbol{x}_{n+1,s} - \mathbf{B}_{ns}\boldsymbol{x}_{ns} - \boldsymbol{\phi}_{ns}\Delta t_{ns}\right) \\ &\quad - \frac{1}{2}K\sum_{s}(N_s-1)\log(2\pi) - \frac{1}{2}\sum_{s=1}^{S}\sum_{n=1}^{N_s-1}\log|\boldsymbol{\Gamma}_{ns}|\Big\} \end{aligned} \tag{S3-3}$$

where $\mathbf{B}_{ns} = \exp\left(-\mathbf{A}\Delta t_{ns}\right)$, $\boldsymbol{\phi}_{ns} = \frac{1}{\Delta t_{ns}}\int_{t_{ns}}^{t_{n+1,s}} \mathrm{e}^{-\mathbf{A}(t_{n+1,s}-t')}\boldsymbol{f}(t')\mathrm{d}t'$ and $\boldsymbol{\Gamma}_{ns}$ is given by eq. 2.7 (with $\boldsymbol{\Gamma}_{ns}^{-1} = (\boldsymbol{\Gamma}_{ns})^{-1}$) and described in more detail in Appendix S1. Measurement error is described by

$$\begin{aligned} P\left(\boldsymbol{y}\,|\boldsymbol{x},\boldsymbol{\theta}\right) &= \prod_{s=1}^{S}\prod_{n=1}^{N_s}\prod_{k=1}^{K} P(y_{kns}|x_{kns},S_k^{\mathrm{m}}) \\ &= \exp\Big\{-\frac{1}{2}\sum_{s=1}^{S}\sum_{n=1}^{N_s}\left(\boldsymbol{y}_{ns} - \boldsymbol{x}_{ns}\right)^{\top}\mathbf{S}^{\mathrm{m}-1}\left(\boldsymbol{y}_{ns} - \boldsymbol{x}_{ns}\right) \\ &\quad - \frac{1}{2}KN_S\log(2\pi) - \frac{1}{2}N_S\log|\mathbf{S}^{\mathrm{m}}|\Big\} \end{aligned} \tag{S3-4}$$

where $N_S = \sum_{s=1}^{S} N_s$, and $\mathbf{S}^{\mathrm{m}}$ is a matrix with the S_k^{m} on its diagonal and 0 otherwise (alternative choices are easily accommodated).

The description of the common trend is a bit more complicated when different systems

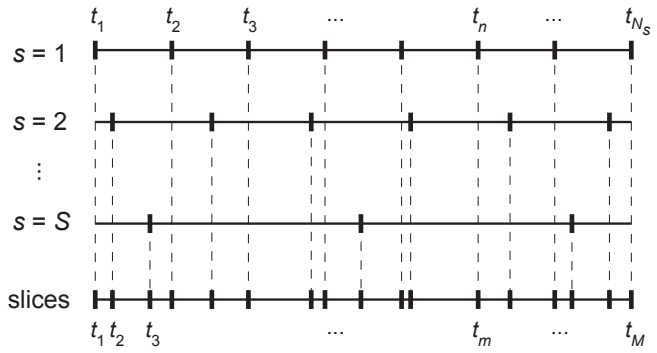

Figure S-1: In the case of different measurement schedules for different systems s, measurement intervals of individual replicates (t_n, t_{n+1}) are subdivided in the analysis by the time points at which other systems are measured $(t_m, m = 1, \ldots, M)$, to allow formulation of a single common trend. Horizontal lines represent measurement schedules for individual ecosystems, with measurement time points indicated by vertical bars. The bottom line represents the union of measurement schedules needed for specification of the trend parameters.

have different measurement schedules. We illustrate this in Figure S-1. The common trend needs to be defined on *all* time points, which we will label t_m, with $m = 1, \ldots, M$. In other words, the t_m are the union of all t_{ns}: $\{t_m\} = \bigcup_s \{t_{ns}\}$. Following eq. (S2-7), and placing all factors in the exponent, we have

$$\begin{aligned} P(\boldsymbol{\phi}|\boldsymbol{\theta};\mu) = \exp\Big\{ &-\frac{\mu^2}{2} \sum_{k,k'}^{K} \sum_{m,m'}^{M-1} (\phi_{mk} - \phi^0_{mk}) D_{mk,m'k'} (\phi_{m'k'} - \phi^0_{m'k'}) \\ &- \frac{K(M-1)}{2} \log(2\pi) + \frac{1}{2} \log |\mu^2 \mathbf{D}| \Big\} \end{aligned} \tag{S3-5}$$

To relate $\boldsymbol{\phi}_n$ and $\boldsymbol{\phi}_m$, we first define

$$\psi^{(s)}_{nm} = e^{-\mathbf{A}(t_{n+1,s} - t_{m+1})} \mathbb{1}_{(m \subset (n,s))} \tag{S3-6}$$

where "$m \subset (n, s)$" means that t_m lies within the interval $[t_n,\ t_{n+1})$, $\mathbb{1}_{()}$ is the indicator function which is 1 when its argument is true and 0 otherwise, and each $\psi^{(s)}_{nm}$ is a $K \times K$

submatrix of $\psi^{(s)}$. Now,

$$\begin{aligned}\boldsymbol{\phi}_{ns}\Delta t_{ns} &= \int_{t_{ns}}^{t_{n+1,s}} \mathrm{e}^{-\mathbf{A}(t_{n+1,s}-t')}\boldsymbol{f}(t')\mathrm{d}t = \sum_{m\subset(n,s)}\int_{t_m}^{t_{m+1}} \mathrm{e}^{-\mathbf{A}(t_{n+1,s}-t')}\boldsymbol{f}(t')\mathrm{d}t \\ &= \sum_{m\subset(n,s)} \mathrm{e}^{-\mathbf{A}(t_{n+1,s}-t_{m+1})}\int_{t_m}^{t_{m+1}} \mathrm{e}^{-\mathbf{A}(t_{m+1}-t')}\boldsymbol{f}(t')\mathrm{d}t \\ &= \sum_{m\subset(n,s)} \psi^{(s)}_{nm}\boldsymbol{\phi}_m\Delta t_m = \sum_{m=1}^{M-1}\psi^{(s)}_{nm}\boldsymbol{\phi}_m\Delta t_m\end{aligned} \tag{S3-7}$$

In order to integrate over $\boldsymbol{x}$, we collect terms of the type $\boldsymbol{x}_{ns}(\cdot)\boldsymbol{x}_{n's}$ into $\mathbf{Q}^{(s)}_{nn'}$, terms of the type $\boldsymbol{x}^\top_{ns}(\cdot)$ into $\boldsymbol{u}_{ns}$, and constant terms into $C^{(s)}$ and $C^{(0)}$. That is,

$$P(\boldsymbol{y}\mid\boldsymbol{\theta};\mu) = \int\mathrm{d}\boldsymbol{\phi}\int\mathrm{d}\boldsymbol{x}\exp\Big\{-\frac{1}{2}\sum_s\big(\boldsymbol{x}_s{}^\top\mathbf{Q}^{(s)}\boldsymbol{x}_s - \boldsymbol{u}_s{}^\top\boldsymbol{x}_s - \boldsymbol{x}_s{}^\top\boldsymbol{u}_s + C^{(s)}\big) - \frac{1}{2}C^{(0)}\Big\}, \tag{S3-8}$$

where

$$\begin{aligned}\mathbf{Q}^{(s)}_{n,n'} &= \delta_{nn'}\mathbb{1}_{(n>1)}\boldsymbol{\Gamma}^{-1}_{n-1,s} - \delta_{n-1,n'}\mathbb{1}_{(n>1)}\beta_{n-1,s} - \delta_{n+1,n'}\mathbb{1}_{(n<N)}\beta^\top_{ns} + \delta_{nn'}\mathbb{1}_{(n<N)}\alpha_{ns} + \delta_{nn'}(\mathbf{S}^{\mathrm{m}})^{-1} \\ \boldsymbol{u}_{ns} &= \mathbb{1}_{(n>1)}\boldsymbol{\Gamma}^{-1}_{n-1,s}\boldsymbol{\phi}_{n-1,s}\Delta t_{n-1,s} - \mathbb{1}_{(n<N)}\beta^\top_{ns}\boldsymbol{\phi}_{ns}\Delta t_{ns} + (\mathbf{S}^{\mathrm{m}})^{-1}\boldsymbol{y}_{ns} \\ C^{(s)} &= \sum_{n=1}^{N-1}\Delta t_{ns}\boldsymbol{\phi}^\top_{ns}\boldsymbol{\Gamma}^{-1}_{ns}\boldsymbol{\phi}_{ns}\Delta t_{ns} + \sum_{n=1}^{N_s}\boldsymbol{y}^\top_{ns}(\mathbf{S}^{\mathrm{m}})^{-1}\boldsymbol{y}_{ns} + \sum_{n=1}^{N_s-1}\log|\boldsymbol{\Gamma}_{ns}| \\ &\quad + N_s\log|\mathbf{S}^{\mathrm{m}}| + K(2N_s-1)\log(2\pi) \\ C^{(0)} &= \mu^2\sum_{m=1}^{M-1}\sum_{m'=1}^{M-1}(\boldsymbol{\phi}_m-\boldsymbol{\phi}^0_m)^\top\mathbf{D}_{mm'}(\boldsymbol{\phi}_{m'}-\boldsymbol{\phi}^0_{m'}) - \log|\mu^2\mathbf{D}| + K(M-1)\log(2\pi) \\ \alpha_{ns} &= \mathbf{B}^\top_{ns}\boldsymbol{\Gamma}^{-1}_{ns}\mathbf{B}_{ns} \quad\text{with}\quad \mathbf{B}_{ns} = \exp(-\mathbf{A}\Delta t_{ns}) \\ \beta_{ns} &= \boldsymbol{\Gamma}^{-1}_{ns}\mathbf{B}_{ns}\end{aligned} \tag{S3-9}$$

where $\mathbf{S}^{\mathrm{m}}$ is a matrix with the S^{m}_k on its diagonal; and each $\mathbf{D}_{mm'}$ is a $K\times K$ submatrix of $\mathbf{D}$. After integration over $\boldsymbol{x}$ as in the main text, we expand, use eq. (S3-7) to replace the $\boldsymbol{\phi}_{ns}$ by $\boldsymbol{\phi}_m$, and collect terms of type $\boldsymbol{\phi}_m(\cdot)\boldsymbol{\phi}_{m'}$ into $\mathbf{R}_{mm'}$, of type $-\boldsymbol{\phi}^\top_m(\cdot)$ into $\boldsymbol{v}_m$, and constants into C'. That is,

$$P(\boldsymbol{y}|\boldsymbol{\theta},\mu) = \int\mathrm{d}\boldsymbol{\phi}\,\exp\Big\{-\frac{1}{2}\left(\boldsymbol{\phi}^\top\mathbf{R}\boldsymbol{\phi} - \boldsymbol{v}^\top\boldsymbol{\phi} - \boldsymbol{\phi}^\top\boldsymbol{v} + C'\right)\Big\} \tag{S3-10}$$

where

$$
\begin{aligned}
\mathbf{R}_{mm'} &= \mu^2 \mathbf{D}_{mm'} + \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \Delta t_m {\psi_{nm}^{(s)}}^\top \mathbf{\Gamma}_{ns}^{-1} \psi_{n,m'}^{(s)} \Delta t_{m'} \\
&\quad - \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \sum_{n'=1}^{N_s-1} \Big\{ \mathbb{1}_{(n<N_s)} \mathbb{1}_{(n'<N_s)} \Delta t_m \psi_{nm}^{(s)} \times \\
&\quad \big\{ \mathbf{\Gamma}_{ns}^{-1} {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{n+1,n'+1} \mathbf{\Gamma}_{n's}^{-1} - \mathbf{\Gamma}_{ns}^{-1} {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{n+1,n'} \beta_{n's}^\top - \beta_{ns} {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{n,n'+1} \mathbf{\Gamma}_{n's}^{-1} + \beta_{ns} {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{nn'} \beta_{n's}^\top \big\} \times \\
&\quad \psi_{n'm'}^{(s)} \Delta t_{m'} \Big\} \\
\boldsymbol{v}_m &= \mu^2 \sum_{m'=1}^{M-1} \mathbf{D}_{mm'} \boldsymbol{\phi}_{m'}^0 + \sum_{n,n',s} \mathbb{1}_{(n<N_s)} \Delta t_m {\psi_{nm}^{(s)}}^\top \big(\mathbf{\Gamma}_n^{-1} {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{n+1,n'} - \beta_n {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{nn'} \big) (\mathbf{S}^{\mathrm{m}})^{-1} \boldsymbol{y}_{n's} \\
C' &= \mu^2 \sum_{m=1}^{M-1} \sum_{m'=1}^{M-1} {\boldsymbol{\phi}_m^0}^\top \mathbf{D}_{mm'} \boldsymbol{\phi}_{m'}^0 + \sum_{s=1}^{S} \sum_{n=1}^{N_s-1} \log |\mathbf{\Gamma}_{ns}| + N_S \log |S^{\mathrm{m}}| + \sum_{s=1}^{S} \log |\mathbf{Q}^{(s)}| \\
&\quad + \sum_{s=1}^{S} \sum_{n=1}^{N} \boldsymbol{y}_{ns}^\top (\mathbf{S}^{\mathrm{m}})^{-1} \boldsymbol{y}_{ns} - \sum_{s=1}^{S} \sum_{n,n'=1}^{N,N} \boldsymbol{y}_{ns}^\top (\mathbf{S}^{\mathrm{m}})^{-1} {\mathbf{Q}^{(\mathbf{s})}}^{-1}_{nn'} (\mathbf{S}^{\mathrm{m}})^{-1} \boldsymbol{y}_{n's} - \log |\mu^2 \mathbf{D}| \\
&\quad + K \sum_s (N_s - 1) \log(2\pi) + K(M-1) \log(2\pi)
\end{aligned}
\tag{S3-11}
$$

where ${\psi_{nm}^{(s)}}^\top = (\psi_{nm}^{(s)})^\top$. The final expression for the log-likelihood is as in the main text:

$$
\mathcal{L} = -2 \log P(\boldsymbol{y}|\boldsymbol{\theta}) = C' + \log |\mathbf{R}| - \boldsymbol{v}^\top \mathbf{R}^{-1} \boldsymbol{v} - K(M-1) \log(2\pi) \tag{S3-12}
$$

In a Bayesian approach, one simply adds the contributions of the prior distributions:

$$
\mathcal{L}^B = \mathcal{L} - 2 \log P(\mathbf{A}, \mathbf{S}^{\mathrm{d}}, \mathbf{S}^{\mathrm{m}}) \tag{S3-13}
$$

with the prior distributions defined in Appendix S2.

Multiple trends. In case one wants to use different trends for different sets of systems, one simply defines $\psi_{nm}^{(s,\tau)} = e^{-\mathbf{A}(t_{n+1,s} - t_{m+1})} \mathbb{1}_{(m \subset (n,s))} \mathbb{1}_{(s \subset \mathcal{S}_\tau)}$, with $\mathcal{S}_\tau$ the set of systems described by trend τ. For multiple trends, one also obtains $\mathbf{R}_{mm'}^{(\tau)}$, $\boldsymbol{v}_m^{(\tau)}$ and $C'^{(\tau)}$ for each trend τ by replacing $\psi_{nm}^{(s)}$ by $\psi_{nm}^{(s,\tau)}$ and $\mathbf{D}$ by $\mathbf{D}^{(\tau)}$ in their definitions above. Then $\mathcal{L} = \sum_\tau \mathcal{L}^{(\tau)}$.

Extraction of underlying dynamics and common trend. Once the log-likelihood function, $\mathcal{L}^B$ or $\mathcal{L}$ has been maximized, one has a set of optimal parameters, and, within a Bayesian approach an estimate of the joint posterior distribution of the model parameters.

But, one can also be interested in obtaining estimates of the underlying logarithmic densities for each system, $\boldsymbol{x}_s$, and the common trend, $\langle \boldsymbol{x} \rangle$. We calculate the common trend as the expected average dynamics.

$$\begin{aligned}
\langle \boldsymbol{x}(t_m) \rangle &= \mathrm{e}^{-\mathbf{A}(t_m - t_1)} \cdot \langle \boldsymbol{x}(t_1) \rangle + \int_{t_1}^{t_m} \mathrm{e}^{-\mathbf{A}(t_m - t')} \boldsymbol{f}(t') \mathrm{d}t' \\
&= \mathrm{e}^{-\mathbf{A}(t_m - t_1)} \cdot \langle \boldsymbol{x}(t_1) \rangle + \sum_{m'=1}^{m-1} \int_{t_{m'}}^{t_{m'+1}} \mathrm{e}^{-\mathbf{A}\left(t_m - t_{m'+1} + t_{m'+1} - t'\right)} \boldsymbol{f}(t') \mathrm{d}t' \\
&= \mathrm{e}^{-\mathbf{A}(t_m - t_1)} \cdot \langle \boldsymbol{x}(t_1) \rangle + \sum_{m'=1}^{m-1} \mathrm{e}^{-\mathbf{A}\left(t_m - t_{m'+1}\right)} \int_{t_{m'}}^{t_{m'+1}} \mathrm{e}^{-\mathbf{A}\left(t_{m'+1} - t'\right)} \boldsymbol{f}(t') \mathrm{d}t' \\
&= \mathrm{e}^{-\mathbf{A}(t_m - t_1)} \cdot \langle \boldsymbol{x}(t_1) \rangle + \sum_{m'=1}^{m-1} \mathrm{e}^{-\mathbf{A}\left(t_m - t_{m'+1}\right)} \boldsymbol{\phi}_{m'} \Delta t_{m'} \\
&= \mathrm{e}^{-\mathbf{A}(t_m - t_1)} \cdot \langle \boldsymbol{x}(t_1) \rangle + \sum_{m'=1}^{m-1} \Psi_{mm'} \boldsymbol{\phi}_{m'}
\end{aligned} \tag{S3-14}$$

where $\Psi_{mm'} = \mathrm{e}^{-\mathbf{A}\left(t_m - t_{m'+1}\right)} \Delta t_{m'}$, similar to the above definition of ψ_{nm} (eq. (S3-6)). For the purpose of Figure 2A, we estimated $\langle \boldsymbol{x}(t_1) \rangle$ by linear regression (it plays no role in parameter estimation). Neglecting error in the initial condition, $\boldsymbol{x}(t_1)$, we have as the posterior covariance matrix for the expected average dynamics

$$\begin{aligned}
\mathrm{Cov}(\langle \boldsymbol{x}(t_m) \rangle, \langle \boldsymbol{x}(t_{m'}) \rangle) &= \mathrm{Cov}(\sum_{m''=1}^{m-1} \Psi_{m,m''} \boldsymbol{\phi}_{m''}, \sum_{m'''=1}^{m'-1} \Psi_{m',m'''} \boldsymbol{\phi}_{m'''}) \\
&= \Psi \mathrm{Cov}(\phi, \phi) \Psi^\top = \Psi \mathbf{R}^{-1} \Psi^\top
\end{aligned} \tag{S3-15}$$

In order to obtain the marginal distribution of the underlying densities in individual ecosystems, x_{kns}, one should integrate over the $\boldsymbol{\phi}$ first. We do this by noting that one can put the likelihood function in a Gaussian form for $\boldsymbol{x}$ and $\boldsymbol{\phi}$ at the same time. That is,

$$l \propto \exp\left\{ -\frac{1}{2} \left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \end{pmatrix}^\top \begin{bmatrix} \mathbf{Q}^{(xx)} & \mathbf{Q}^{(x\phi)} \\ \mathbf{Q}^{(x\phi)\top} & \mathbf{Q}^{(\phi\phi)} \end{bmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \end{pmatrix} - \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{u}^{(x)} \\ \boldsymbol{u}^{(\phi)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{u}^{(x)} \\ \boldsymbol{u}^{(\phi)} \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \end{pmatrix} + C \right) \right\}. \tag{S3-16}$$

with

$$\mathbf{Q}^{(xx)} = \begin{bmatrix} \mathbf{Q}^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{(s)} & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Q}^{(S)} \end{bmatrix} \tag{S3-17}$$

where $\mathbf{Q}^{(s)}$ $(s = 1, \ldots, S)$ is defined in eq. (S3-9), and

$$\begin{aligned}
\mathbf{Q}^{(x\phi)}_{ns,m} &= -\mathbb{1}_{(n>1)}\mathbf{\Gamma}^{-1}_{n-1,s}\psi^{(s)}_{n-1,m}\Delta t_m + \mathbb{1}_{(n<N)}\beta^{\top}_{ns}\psi^{(s)}_{n,m}\Delta t_m \\
\mathbf{Q}^{(\phi\phi)}_{m,m'} &= \mu^2\mathbf{D}_{mm'} + \sum_{s=1}^{S}\sum_{n=1}^{N_s-1}\left(\psi^{(s)}_{nm}\right)^{\top}\Delta t_{ns}\mathbf{\Gamma}^{-1}_{ns}\Delta t_{ns}\psi^{(s)}_{n,m'} \\
\boldsymbol{u}^{(x)}_{ns} &= (\mathbf{S}^{\mathrm{m}})^{-1}\boldsymbol{y}_{ns} \\
\boldsymbol{u}^{(\phi)}_{m} &= \mu^2\sum_{m'=1}^{M-1}\mathbf{D}_{mm'}\boldsymbol{\phi}^0_{m'}
\end{aligned} \tag{S3-18}$$

where $\boldsymbol{u}_{ns}$ is defined in eq. (S3-9) as well. The value of C is not important here. Standard results for jointly multivariate normal distributions are that $\boldsymbol{x}$ has as its marginal covariance matrix after integration over $\boldsymbol{\phi}$:

$$\Sigma^{(xx)} = \left(\mathbf{Q}^{(xx)} - \mathbf{Q}^{(x\phi)}{\mathbf{Q}^{(\phi\phi)}}^{-1}{\mathbf{Q}^{(x\phi)}}^{\top}\right)^{-1} \tag{S3-19}$$

and as its (marginal) expectation value:

$$\mu^{(x)}_{ns} = \left(\Sigma^{(xx)}\left(\boldsymbol{u}^{(x)} - \mathbf{Q}^{(x\phi)}{\mathbf{Q}^{(\phi\phi)}}^{-1}\boldsymbol{u}^{(\phi)}\right)\right)_{ns} \tag{S3-20}$$

for each n and s.

Effective number of parameters in the trend. As described above, we obtain a posterior distribution for the common trend characterized (eqns. (S3-14), (S3-15). Here, we examine the number of degrees of freedom in the common trend, $\langle\boldsymbol{x}(t)\rangle$ given by eq. (S3-14), that make a substantial contribution to its description. Naively, each species' trend is described by M parameters (the ϕ_{mk} and an intercept). However, structure in the data and the "smoothing" effect of the prior distribution make most of the trend parameters irrelevant (for example, a straight line can be described just as well by 200 parameters as by 2). More precisely, the estimates of the ϕ_{mk} become strongly correlated at large μ.

For this analysis, we examine the expected power (square size) of fluctuations in the common trend. Such power will arise from measurement error, limited sampling, and true temporal variation of the average dynamics. We ask how many degrees of freedom are needed to describe any power in excess of what is expected from measurement error alone. To do so, we first decompose the fluctuations by projection onto an orthonormal base, say $\mathbf{V} = \left(\boldsymbol{v}^{(1)} \cdots \boldsymbol{v}^{(q)} \cdots \boldsymbol{v}^{(M)}\right)$ with each $\boldsymbol{v}^{(q)}$ a column vector $(q = 1, \ldots, M)$ with M entries. The $\boldsymbol{v}^{(q)}$ satisfy ${\boldsymbol{v}^{(q)}}^{\top}\boldsymbol{v}^{(q')} = \delta_{qq'}$. A well-known example of such decomposition of fluctuations is the discrete Fourier transformation, for which $v^{(q)}_m = \exp(-2\pi i\frac{qm}{M})$ (for constant Δt). We will examine fluctuations in $\langle\boldsymbol{x}(t_m)\rangle - \overline{\langle\boldsymbol{x}(t_m)\rangle}$ where the bar indicates an

additional average over time (we will add this degree of freedom back again at the end). For simplicity, we present the analysis for a single species; the joint analysis is a trivial modification. We will hence suppress the species index, k in most places.

For the common trend, its projection on the q^{th} vector in the orthonormal base is:

$$X_q = \sum_m v_m^{(q)} \left(c_m + \sum_{m'} \Psi_{mm'} (\phi_{m'} - \langle \phi_{m'} \rangle + \langle \phi_{m'} \rangle - \overline{\phi}) \right) \tag{S3-21}$$

with $c_m = \Big((\exp(-\mathbf{A}(t_m - t_1)) - \overline{\exp(-\mathbf{A}(t_m - t_1))}) \boldsymbol{x}(t_1) \Big)_k$. Corresponding to each decomposition of the fluctuations is a decomposition of power in the fluctuations (in the case of Fourier transforms, this is the power spectrum). The power in the fluctuations, averaged over the posterior distribution, is now given by

$$\begin{aligned} \left\langle X_q X_q^* \right\rangle = \sum_{m,m'''} v_m^{(q)} \Bigg(& c_m c_{m'''} + \sum_{m',m''} \Psi_{mm'} \mathrm{Cov}(\phi_{m'}, \phi_{m''}) \Psi^\top_{m''m'''} \\ & + \sum_{m',m''} \Psi_{mm'} \left(\langle \phi_{m'} \rangle - \overline{\phi} \right) \left(\langle \phi_{m''} \rangle - \overline{\phi} \right) \Psi^\top_{m''m'''} \Bigg) v_{m'''}^{(q)} \end{aligned} \tag{S3-22}$$

where * indicates the complex conjugate if necessary, and Ψ as in eq. (S3-14). We define

$$\mathcal{M}_{mm'''} = c_m c_{m'''} + \sum_{m',m''} \Psi_{mm'} \mathrm{Cov}(\phi_{m'}, \phi_{m''}) \Psi^\top_{m''m'''} + \sum_{m',m''} \Psi_{mm'} \left(\langle \phi_{m'} \rangle - \overline{\phi} \right) \left(\langle \phi_{m''} \rangle - \overline{\phi} \right) \Psi^\top_{m''m'''} \tag{S3-23}$$

The power in fluctuations of a particular mode is hence $\sum_{m,m'''} v_m^{(q)} \mathcal{M}_{mm'''} v_{m'''}^{(q)}$ for each vector $\boldsymbol{v}^{(q)}$ in our basis. (Note that for our data the second term in eq. (S3-23) dominates the analysis. One could thus perform the analogous analysis of the ϕ_{mk} with nearly identical results.) Let the $\boldsymbol{v}^{(q)}$ be arranged in a predescribed order from the most "trend-like" (i.e. slowly varying) to the most "noise-like" (e.g. most variable with m). We consider partial sums $C_q = \sum_{q'=q+1}^{M} \left\langle X_q X_q^* \right\rangle$, that is, over the most noise-like modes. We are interested in how many modes (the $\boldsymbol{v}^{(q)}$) need to be included in order to accommodate the expected amount of power in the common trend due to measurement noise, $\varepsilon = \mathrm{Tr}(\mathbf{S}^{\mathrm{m}}) \sum_m 1/N_m$, where N_m is the number of replicate systems for which a measurement was made at time t_m. We define a number q_c such that $C_{q_c} < \varepsilon < C_{q_c+1}$. Now the effective number of degrees of freedom is $1 + q_c(k)$ for an analysis per species.

We have performed this analysis for a number of choices of orthonormal basis as a function of the smoothing parameter, μ. The result for perhaps the simplest choice, using the eigenvectors of $\mathcal{M}$ is shown in Figure S-2A. In Figure S-2B we show a decomposition using as a basis the eigenvectors of $\mathbf{D}$. Because of the structure of $\mathbf{D}$, this can be thought of

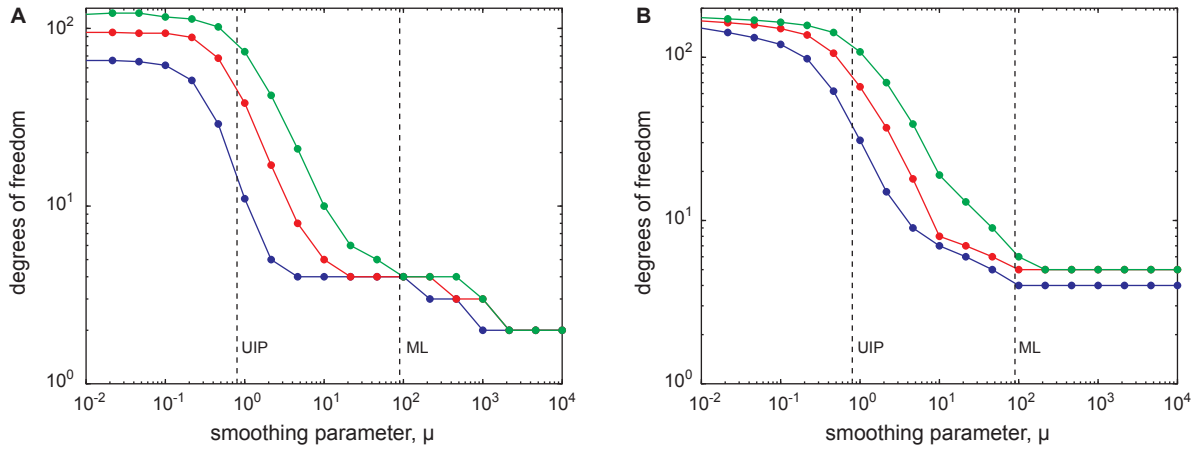

Figure S-2: Scaling of the number of effective degrees of freedom in the description of the common trend with the smoothing parameter, μ for *C. reinhardtii* (green); *E. coli* (red); *T. thermophila* (blue) using (A) the eigenvectors of $\mathcal{M}$ and (B) the eigenvectors of $\mathbf{D}$. UIP: unit information prior; ML: maximum likelihood estimate of μ.

as a discrete Fourier transform accommodating unevenly spaced measurements.

Optimal smoothing parameter, μ. In the main text we assert that if one optimizes μ together with the model parameters (rather than keeping it fixed or using a unit information prior), the optimal value, μ^* can be infinite. To examine this more closely, assume, for simplicity, that the other parameters are known. Following Table 2, we have

$$\begin{aligned}\mathcal{L}_{\boldsymbol{y}}(\mu,\boldsymbol{\theta}) &= -\log|\mu^2\mathbf{D}| + \log|\mathbf{R}| - \boldsymbol{v}^\top\mathbf{R}^{-1}\boldsymbol{v} + \dots \\ &= -\log|\mu^2\mathbf{D}| + \log|\mathbf{R}' + \mu^2\mathbf{D}| - \boldsymbol{v}^\top(\mathbf{R}' + \mu^2\mathbf{D})^{-1}\boldsymbol{v} + \dots\end{aligned} \tag{S3-24}$$

where $\mathbf{R}' = \mathbf{R} - \mu^2\mathbf{D}$ is the prior-independent part of $\mathbf{R}$. We will look for optimal μ assuming it is large (in which case large-μ assumptions are self-consistent). In this limit, we can use the following two approximations:

$$\begin{aligned}\log|\mathbf{R}' + \mu^2\mathbf{D}| &= \log|\mu^2\mathbf{D}(\mathbf{I} + \tfrac{1}{\mu^2}\mathbf{D}^{-1}\mathbf{R}')| \\ &= \log|\mu^2\mathbf{D}| + \log|(\mathbf{I} + \tfrac{1}{\mu^2}\mathbf{D}^{-1}\mathbf{R}')| \\ &\approx \log|\mu^2\mathbf{D}| + 1 + \tfrac{1}{\mu^2}\mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}')\end{aligned} \tag{S3-25}$$

and

$$\begin{aligned}(\mathbf{R}' + \mu^2\mathbf{D})^{-1} &= \left(\mu^2\mathbf{D}(\mathbf{I} + \tfrac{1}{\mu^2}\mathbf{D}^{-1}\mathbf{R}')\right)^{-1} \\ &= \left(\mathbf{I} + \tfrac{1}{\mu^2}\mathbf{D}^{-1}\mathbf{R}'\right)^{-1}(\mu^2\mathbf{D})^{-1} \\ &\approx \left(\mathbf{I} - \tfrac{1}{\mu^2}\mathbf{D}^{-1}\mathbf{R}'\right)\tfrac{1}{\mu^2}\mathbf{D}^{-1}\end{aligned} \tag{S3-26}$$

Introducing $\tau = \frac{1}{\mu^2}$, we obtain an expression for the log-likelihood function which is a simple quadratic function of τ:

$$\begin{aligned}\mathcal{L}_{\boldsymbol{y}}(\mu, \boldsymbol{\theta}) &= \tau \mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}') - \boldsymbol{v}^\top \left\{ \left(\mathbf{I} - \tau \mathbf{D}^{-1}\mathbf{R}'\right) \tau \mathbf{D}^{-1} \right\} \boldsymbol{v} \\ &= \tau^2 (\boldsymbol{v}^\top \mathbf{D}^{-1}\mathbf{R}'\mathbf{D}^{-1}\boldsymbol{v}) + \tau \left(\mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}') - \boldsymbol{v}^\top \mathbf{D}^{-1}\boldsymbol{v} \right) + C\end{aligned} \tag{S3-27}$$

with C collecting terms independent of τ (or μ). The optimum, μ^*, is finite, provided $\tau^* > 0$. That is,

$$\tau^* = \frac{\boldsymbol{v}^\top \mathbf{D}^{-1}\boldsymbol{v} - \mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}')}{2\boldsymbol{v}^\top \mathbf{D}^{-1}\mathbf{R}'\mathbf{D}^{-1}\boldsymbol{v}} > 0 \tag{S3-28}$$

That is, $\boldsymbol{v}^\top \mathbf{D}^{-1}\boldsymbol{v} > \mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}')$ (the denominator in eq. (S3-28) is necessarily positive). This condition is more likely to be fulfilled when more replicate systems are available since $\boldsymbol{v}^\top \mathbf{D}^{-1}\boldsymbol{v}$ grows quadratically in the number of systems while $\mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}')$ grows only linearly in S. As a result, one can estimate $S_c/S = \mathrm{Tr}(\mathbf{D}^{-1}\mathbf{R}')/\boldsymbol{v}^\top \mathbf{D}^{-1}\boldsymbol{v}$. If the number of measurement points per time series can be varied too, one can read $(NS)_c/(NS)$ rather than S_c/S.

We note that for the example descibed in the main text (§3.3, of a single species at constant Δt, $\mathbf{D}^{-1}/\mathrm{Tr}(\mathbf{D}^{-1}) \approx \mathbf{1}$), with $\mathbf{1}$ a square matrix of ones.

Within the model, the inferred value of S_c over its true value follows from eq. 3.8.

S4 Analysis of preliminary "coarse" experiments

Duplicate single-species and two-species ecosystems were prepared as for three-species ecosystems [3], except that we used clear screw-thread vials (Fisherbrand, nominal volume: 4 mL, total internal volume: 4.9 mL) with screw caps with teflon-lined silicon septum. At each time point, $\sim 100\,\mu$L was sampled by syringe through the septum after gentle homogenization by repeated inversion (unlike the three-species data which were obtained by fluorescence selective plane illumination microscopy [3]). *T. thermophila* and *C. reinhardtii* counts were determined by haemocytometry, *E. coli* densities by plating on Luria-Bertani agar.

We will use the following notation: $y_{kn} = \log N_{kn}$ is the logarithmic density. For simplicity, we will omit the system index s. Instead, every sum over time points is, implicity, also a sum over replicate systems (here $S = 2$ in each case). To distinguish the different species combinations, we will indicate them with superscripts. That is, $y_{in}^{[ij]}$ is the logarithmic density of species i at time point n in a two-species system containing species i and j.

The advantage of having one- and two-species data is that we can extract the trend for isolated species, and then estimate the interaction coefficients from two-species data. That

is, we will model the single-species dynamics as:

$$\frac{\mathrm{d}y_i^{[i]}}{\mathrm{d}t} = f_i(t) - A_{ii}y_i^{[i]}(t) \tag{S4-1}$$

We use the single species data to determine the trend parameters, i.e.,

$$f_i(t) = \frac{\mathrm{d}y_i^{[i]}}{\mathrm{d}t} + A_{ii}y_i^{[i]}(t) \tag{S4-2}$$

The available data are fairly coarse, and we take a pragmatic approach. First, the data do not permit a meaningful separation of measurement and dynamical noise. We will use the single-species data to predict the two-species data, as a function of the unknown A_{ij}

$$\begin{aligned}\frac{\mathrm{d}x_i^{[ij]}}{\mathrm{d}t} &= f_i(t) - A_{ii}y_i^{[ij]}(t) - A_{ij}y_j^{[ij]}(t) \\ &= \frac{\mathrm{d}y_i^{[i]}}{\mathrm{d}t} - A_{ii}(y_i^{[ij]}(t) - y_i^{[i]}(t)) - A_{ij}y_j^{[ij]}(t)\end{aligned}, \tag{S4-3}$$

where $x_i^{[ij]}$ are the predicted dynamics of i in a two-species i, j system. After replacing derivatives by finite differences,

$$\begin{aligned}\frac{\Delta x_{in}^{[ij]}}{\Delta t_n} &\simeq f_i(t_n) - A_{ii}y_{in}^{[ij]} - A_{ij}y_{jn}^{[ij]} \\ &= \frac{\Delta y_{in}^{[i]}}{\Delta t_n} - A_{ii}(y_{in}^{[ij]} - y_{in}^{[i]}) - A_{ij}y_{jn}^{[ij]}\end{aligned}, \tag{S4-4}$$

we simultaneously estimate all interaction coefficients by minimizing (with respect to each interaction coefficient) the sums of squares (SS_i) of deviations between predicted and observed increments for each species, i,

$$SS_i = \sum_n \left(\frac{\Delta y_{in}^{[ij]}}{\Delta t_n} - \frac{\Delta x_{in}^{[ij]}}{\Delta t_n}\right)^2 + \sum_n \left(\frac{\Delta y_{in}^{[ik]}}{\Delta t_n} - \frac{\Delta x_{in}^{[ik]}}{\Delta t_n}\right)^2 \tag{S4-5}$$

That is, we get a system of linear equations $\frac{\partial SS_i}{\partial A_{ij}} = 0$ which is readily solved.

In the same experiment, we also obtained duplicate time series for the full three-species system. We use those to examine whether the estimated A_{ij} also adequately describe the three-species dynamics. That is, our predicted three-species dynamics are:

$$\begin{aligned}\frac{\Delta x_i^{[ijk]}}{\Delta t} &= f_i(t) - A_{ii}x_i^{[ijk]}(t) - A_{ij}x_j^{[ijk]}(t) - A_{ik}x_k^{[ijk]}(t) \\ &= \frac{\Delta y_i^{[i]}}{\Delta t} - A_{ii}(x_i^{[ijk]}(t) - y_i^{[i]}(t)) - A_{ij}x_j^{[ijk]}(t) - A_{ik}x_k^{[ijk]}(t)\end{aligned} \tag{S4-6}$$

where the $x_i^{[ijk]}(t)$ are obtained by propagation of eq. (S4-6).

The results of fitting the Gompertz model to the data are shown in Figure S-3.

S5 Optimal sampling schedule

We seek to calculate the Fisher Information matrix for the single-species model with no interactions described in the text. In this case,

$$\begin{aligned}\mathcal{L} &= -2\log P(\boldsymbol{y}|\boldsymbol{\theta}) = \log|\mathbf{Q}| - N\log(2\pi) + C - \boldsymbol{u}^\top\mathbf{Q}^{-1}\boldsymbol{u} \\ &= \log|\mathbf{Q}| + \boldsymbol{y}^\top\big(\frac{\mathbf{I}}{S^{\mathrm{m}}} - \frac{\mathbf{Q}^{-1}}{(S^{\mathrm{m}})^2}\big)\boldsymbol{y} + \sum_n \log(S^{\mathrm{d}}\Delta t_n) + N\log(S^{\mathrm{m}})\end{aligned} \tag{S5-1}$$

As before, we note that the last two terms ($\sum_n \log(S^{\mathrm{d}}\Delta t_n) + N\log(S^{\mathrm{m}})$) yield 0 second derivatives in $\log(S^{\mathrm{d}})$ and $\log(S^{\mathrm{m}})$. In addition, $\mathbf{Q}$ can be written as $\mathbf{Q} = \frac{\tilde{\mathbf{Q}}}{S^{\mathrm{d}}} + \frac{\mathbf{I}}{S^{\mathrm{m}}}$, where $\tilde{\mathbf{Q}}$ depends on the measurement schedule but not on model parameters.

To limit the amount of algebra and highlight the problem structure, we introduce the following two matrices

$$\begin{aligned}\mathbf{G} &= S^{\mathrm{d}}\mathbf{Q} = \tilde{\mathbf{Q}} + \gamma\mathbf{I} \\ \mathbf{F} &= S^{\mathrm{m}}\big(\frac{\mathbf{I}}{S^{\mathrm{m}}} - \frac{\mathbf{Q}^{-1}}{(S^{\mathrm{m}})^2}\big) = \mathbf{I} - \gamma\mathbf{G}^{-1}\end{aligned} \tag{S5-2}$$

so

$$\mathcal{L} = \log|\mathbf{G}| + \frac{1}{S^{\mathrm{m}}}\left\langle \boldsymbol{y}^\top\mathbf{F}\boldsymbol{y}\right\rangle \tag{S5-3}$$

where both $\mathbf{F}$ and $\mathbf{G}$ are functions of $\gamma = S^{\mathrm{d}}/S^{\mathrm{m}}$, but do not otherwise depend on S^{d} or S^{m}. We omitted additional terms linear in $\log S^{\mathrm{m}}$ and $\log S^{\mathrm{d}}$. The Fisher Information matrix is:

$$\mathcal{I}_{ij} = \left\langle \frac{\partial^2\mathcal{L}}{\partial\theta_i\partial\theta_j}\right\rangle \tag{S5-4}$$

where the expectation value is over possible realizations of the data; θ_i is in our case $\log S^{\mathrm{d}}$ or $\log S^{\mathrm{m}}$ since we are interested in relative errors. The expectation value is, eventually, evaluated at the true value of the parameters, $\log S^{\mathrm{d}*}$ or $\log S^{\mathrm{m}*}$. Integration and differentiation in eq. (S5-4) can be interchanged under mild conditions, that is,

$$\begin{aligned}\mathcal{I}_{ij} &= \frac{\partial^2\left\langle\mathcal{L}\right\rangle}{\partial\theta_i\partial\theta_j} \\ &= \frac{\partial^2}{\partial\theta_i\partial\theta_j}\left\{\log|\mathbf{G}| + \frac{1}{S^{\mathrm{m}}}\left\langle\boldsymbol{y}^\top\mathbf{F}\boldsymbol{y}\right\rangle\right\} \\ &= \frac{\partial^2}{\partial\theta_i\partial\theta_j}\left\{\log|\mathbf{G}| + \frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}}\mathrm{Tr}(\mathbf{F}\mathbf{F}^{*-1})\right\}\end{aligned} \tag{S5-5}$$

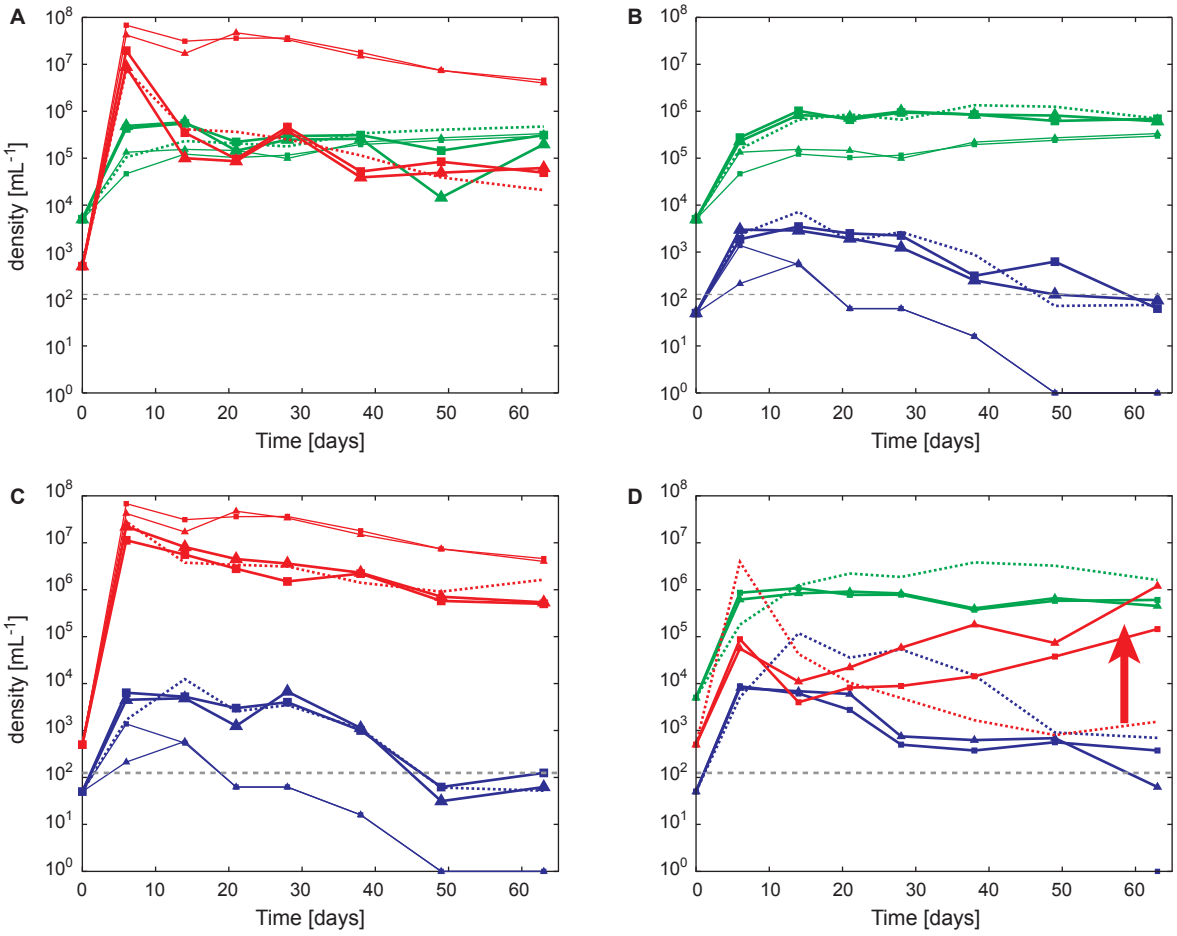

Figure S-3: (**A**, **B** and **C**) One and two-species data for *C. reinhardtii* (green), *E. coli* (red) and *T. thermophila* (blue). Observations for single-species systems are shown as thin solid lines, observations for two-species systems as thick solid lines. Dotted lines: predicted dynamics in two-species cultures as in eq. (S4-4). (**D**) Observed dynamics in duplicate three-species ecosystems in the same experiment. Predictions based on single and two-species systems as in eq. (S4-6) shown as dotted lines. The red arrow indicates the effect mentioned in the main text: *E. coli* does better in the three-species systems than expected based on the one and two-species data. Gray dashed lines: approximate detection limit for haemocytometry for *C. reinhardtii* and *T. thermophila*.

where $\mathbf{F}^*$ is evaluated at the true parameter values $S^{\mathrm{m}*}$ and $S^{\mathrm{d}*}$. We have used that the expectation $\left\langle \boldsymbol{y}^\top \mathbf{F} \boldsymbol{y} \right\rangle = \boldsymbol{m}^\top \mathbf{F} \boldsymbol{m} + \mathrm{Tr}(\mathbf{F}\boldsymbol{\Sigma})$ for a multivariate Gaussian over $\boldsymbol{y}$ with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$ ([7]; here $\boldsymbol{m} = 0$ and $\boldsymbol{\Sigma} = S^{\mathrm{m}*}\mathbf{F}^{*-1}$).

Finite N and T. For finite N and T we will first calculate the derivatives in eq. (S5-5) and then calculate the resulting terms (which will have the form of matrix traces) for each element of $\mathcal{I}$. We will need the following relations, stated here for a general parameter θ and matrix $\mathbf{Q}$ [7]:

$$\partial \log \theta = \frac{\partial \theta}{\theta} \Leftrightarrow \frac{\partial}{\partial \log \theta} = \theta \frac{\partial}{\partial \theta} \tag{S5-6}$$

$$\frac{\partial \log |\mathbf{Q}|}{\partial \theta} = \mathrm{Tr}(\mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial \theta}) \tag{S5-7}$$

$$\frac{\partial \mathbf{Q}^{-1}}{\partial a} = -\mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial a} \mathbf{Q}^{-1} \tag{S5-8}$$

Since we have reparametrized the likelihood function in terms of S^{m} and γ (by using $\mathbf{F}$ and $\mathbf{G}$, we need to reformulate the partial derivatives using the chain rule:

$$\begin{aligned} \frac{\partial}{\partial \log S^{\mathrm{m}}}\Big|_{(S^{\mathrm{m}}, S^{\mathrm{d}})} &= S^{\mathrm{m}} \frac{\partial}{\partial S^{\mathrm{m}}} + S^{\mathrm{m}} \frac{\partial \gamma}{\partial S^{\mathrm{m}}} \frac{\partial}{\partial \gamma} = S^{\mathrm{m}} \frac{\partial}{\partial S^{\mathrm{m}}} - \gamma \frac{\partial}{\partial \gamma} \\ \frac{\partial}{\partial \log S^{\mathrm{d}}}\Big|_{(S^{\mathrm{m}}, S^{\mathrm{d}})} &= S^{\mathrm{d}} \frac{\partial}{\partial S^{\mathrm{d}}} + S^{\mathrm{d}} \frac{\partial \gamma}{\partial S^{\mathrm{d}}} \frac{\partial}{\partial \gamma} = \gamma \frac{\partial}{\partial \gamma} \end{aligned} \tag{S5-9}$$

where the partial derivatives on the right apply only to explicit dependencies. We note that first derivatives of matrices $\mathbf{F}$ and $\mathbf{G}$ are 0 with respect to S^{m}. The first derivative of $\mathbf{G}$ with respect to γ is the identity matrix, $\mathbf{I}$. Furthermore, $\frac{\partial \mathbf{F}}{\partial \gamma} = -\mathbf{G}^{-1}\mathbf{F}$. The first derivatives of the expected log-likelihood are now:

$$\begin{aligned} \frac{\partial \langle \mathcal{L} \rangle}{\partial \log S^{\mathrm{m}}} &= -\gamma \mathrm{Tr}(\mathbf{G}^{-1}) - \frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \mathrm{Tr}(\mathbf{F}\mathbf{F}^{*-1}) + \frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma \mathrm{Tr}(\mathbf{G}^{-1}\mathbf{F}\mathbf{F}^{*-1}) \\ \frac{\partial \langle \mathcal{L} \rangle}{\partial \log S^{\mathrm{d}}} &= \gamma \mathrm{Tr}(\mathbf{G}^{-1}) - \frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma \mathrm{Tr}(\mathbf{G}^{-1}\mathbf{F}\mathbf{F}^{*-1}) \end{aligned} \tag{S5-10}$$

and the second derivatives:

$$\begin{aligned}
\frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \log {S^{\mathrm{m}}}^2} =& \gamma \mathrm{Tr}(\mathbf{G}^{-1}) - \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}) + \frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \mathrm{Tr}(\mathbf{F}\mathbf{F}^{*-1}) - 3\frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma \mathrm{Tr}(\mathbf{G}^{-1}\mathbf{F}\mathbf{F}^{*-1}) \\
&+ 2\frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}\mathbf{F}\mathbf{F}^{*-1}) \\
\frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \log S^{\mathrm{m}} \partial \log S^{\mathrm{d}}} =& -\gamma \mathrm{Tr}(\mathbf{G}^{-1}) + \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}) + 2\frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma \mathrm{Tr}(\mathbf{G}^{-1}\mathbf{F}\mathbf{F}^{*-1}) \\
&- 2\frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}\mathbf{F}\mathbf{F}^{*-1}) \\
\frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \log {S^{\mathrm{d}}}^2} =& \gamma \mathrm{Tr}(\mathbf{G}^{-1}) - \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}) - \frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma \mathrm{Tr}(\mathbf{G}^{-1}\mathbf{F}\mathbf{F}^{*-1}) \\
&+ 2\frac{S^{\mathrm{m}*}}{S^{\mathrm{m}}} \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}\mathbf{F}\mathbf{F}^{*-1})
\end{aligned} \tag{S5-11}$$

These derivatives simplify when evaluated at the true parameter values ($\mathbf{F} = \mathbf{F}^*$ and $S^{\mathrm{m}} = S^{\mathrm{m}*}$):

$$\begin{aligned}
\frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \log {S^{\mathrm{m}}}^2} =& N - 2\gamma \mathrm{Tr}(\mathbf{G}^{-1}) + \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}) = \mathrm{Tr}(\mathbf{F}^2) \\
\frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \log S^{\mathrm{m}} \partial \log S^{\mathrm{d}}} =& \gamma \mathrm{Tr}(\mathbf{G}^{-1}) - \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2}) = \gamma \mathrm{Tr}(\mathbf{G}^{-1}\mathbf{F}) \\
\frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \log {S^{\mathrm{d}}}^2} =& \gamma^2 \mathrm{Tr}(\mathbf{G}^{-2})
\end{aligned} \tag{S5-12}$$

Since each trace is the sum over eigenvalues of its argument, we can rewrite this. To do this, we first denote the eigenvalues of $\tilde{\mathbf{Q}}$ by $\lambda_n^{(\tilde{\mathbf{Q}})}$, with $n = 1, \ldots, N$. For the eigenvalues of $\mathbf{G}$, $\mathbf{G}^{-1}$ and $\mathbf{F}$, we have

$$\begin{aligned}
\lambda_n^{(\mathbf{G})} =& \lambda_n^{(\tilde{\mathbf{Q}})} + \gamma \\
\lambda_n^{(\mathbf{G}^{-1})} =& \frac{1}{\lambda_n^{(\tilde{\mathbf{Q}})} + \gamma} \\
\lambda_n^{(\mathbf{F})} =& 1 - \frac{\gamma}{\lambda_n^{(\tilde{\mathbf{Q}})} + \gamma} = \frac{\lambda_n^{(\tilde{\mathbf{Q}})}}{\lambda_n^{(\tilde{\mathbf{Q}})} + \gamma}
\end{aligned} \tag{S5-13}$$

Rewriting the traces in eq. (S5-12) using the eigenvalues in eq. (S5-13), we arrive at the Fisher information matrix for $\log S^{\mathrm{d}}$ and $\log S^{\mathrm{m}}$:

$$\mathcal{I} = \sum_{n=1}^{N} \frac{1}{(\lambda_n + \gamma)^2} \begin{pmatrix} {\lambda_n}^2 & \gamma\lambda_n \\ \gamma\lambda_n & \gamma^2 \end{pmatrix} \tag{S5-14}$$

where $\lambda_n = \lambda_n^{(\tilde{\mathbf{Q}})}$. These results apply to any measurement schedule.

Asymptotic results. In the limit of infinite N and T, we can also obtain the Fisher Information matrix provided the measurement schedule is periodic. We make use of the periodicity of the measurement intervals, Δt_ℓ, to diagonalize $\tilde{\mathbf{Q}}$ and determine its eigenvalues. Technically speaking, the problem is identical to the determination of an electronic wave function in a periodic potential on a one-dimensional lattice, a well-studied problem in quantum condensed matter physics. For a periodic measurement schedule, we can associate with each integer $j \in [1; N]$ the external block index $k \in [0, \frac{N}{L} - 1]$ and the internal block index $\ell \in [1; L]$ such that $j = kL + \ell$. That is, $\Delta t_{\ell+L} = \Delta t_\ell$ for all ℓ (see Figure 8C for an illustration).

Bloch's theorem (after Felix Bloch) stipulates that the eigenvectors, $\boldsymbol{v}$, of $\tilde{\mathbf{Q}}$ are of the form $v_{(k,\ell)} = V_\ell \, \exp(ik\varphi)$, where φ is a real-valued phase to be determined later, provided the L–dimensional vector, $\boldsymbol{V}$ with components V_ℓ is an eigenvector of $\tilde{Q}(\varphi)$, a submatrix of $\tilde{\mathbf{Q}}$ modified at the corner elements:

$$\left[\tilde{Q}(\varphi)\right]_{\ell,\ell'} = \begin{cases} \frac{1}{\Delta t_\ell} + \frac{1}{\Delta t_{\ell+1}} & \text{if} \quad \ell = \ell' \leq L-1 \\ \frac{1}{\Delta t_L} + \frac{1}{\Delta t_1} & \text{if} \quad \ell = \ell' = L \\ -\frac{1}{\Delta t_\ell} & \text{if} \quad \ell = \ell' + 1 \\ -\frac{1}{\Delta t_{\ell+1}} & \text{if} \quad \ell = \ell' - 1 \end{cases} - \delta_{\ell,L}\delta_{\ell',1}\frac{e^{-i\varphi}}{\Delta t_1} - \delta_{\ell,1}\delta_{\ell',L}\frac{e^{i\varphi}}{\Delta t_1} \qquad \text{(S5-15)}$$

Let λ be an eigenvalue of $\tilde{Q}(\varphi)$; λ is real since $\tilde{Q}(\varphi)$ is a Hermitian matrix. Let us note $V_\ell = \rho_\ell \, \exp(i\,\alpha_\ell)$ where ρ_ℓ and α_ℓ are real-valued numbers. As the complex conjugate of $v_{k,\ell}$ is an eigenvector of $\tilde{\mathbf{Q}}(-\varphi)$ with the same eigenvalue λ. We therefore look for eigenvectors of $\mathcal{Q}$ equal to $w_{(k,\ell)} = A_+ \, v_{k,\ell} + A_- \, v^*_{k,\ell}$. The boundary conditions are $v_{-1,L} = v_{\frac{N}{L},1} = 0$, which gives

$$\varphi_m = \frac{\pi\, m + \alpha_L - \alpha_1}{\frac{N}{L} + 1}, \quad A_+ = \exp[i(\varphi - \alpha_L)], \; A_- = -\frac{1}{A_+} \; . \qquad \text{(S5-16)}$$

Parameter m can take any integer value between 0 and $\frac{N}{L}$. Finally, we find $w_{(k,\ell)} = \rho_\ell \, \sin[(k+1)\varphi_m + \alpha_\ell - \alpha_L]$.

The above reasoning holds for any of the L eigenvalues λ of $Q(\varphi)$. We therefore denote the eigenvalues by $\lambda_\ell(\varphi)$. In the $N \to \infty$ limit, the phase φ_m can take any value comprised between 0 and π. We may thus replace the sum over discrete m with a continuous integral over φ. As a result, the asymptotic expression analogous to eq. (S5-14) is:

$$\mathcal{I}_0 = \frac{1}{L}\sum_{\ell=1}^{L} \frac{1}{\pi}\int_0^\pi \frac{d\varphi}{(\gamma + \lambda_\ell(\varphi))^2} \begin{pmatrix} \lambda_\ell(\varphi)^2 & \gamma\,\lambda_\ell(\varphi) \\ \gamma\,\lambda_\ell(\varphi) & \gamma^2 \end{pmatrix} \; . \qquad \text{(S5-17)}$$

Remarks We conclude with a few remarks:

(1) As in the main text, we have set the average measurement interval, $\overline{\Delta t} = T/N$ to 1, that is, it defines the unit of time. More generally, one should read $S^{\mathrm{d}}\overline{\Delta t}$, rather than S^{d}.

(2) Matrices $\mathbf{Q}$, $\tilde{\mathbf{Q}}$, $\mathbf{F}$ and $\mathbf{G}$ commute. They have the same eigenvectors but different eigenvalues (eq. (S5-13)).

(3) In the known absence of a trend, the benefit of having replicate systems is simply $\mathcal{I} = \sum_s \mathcal{I}_s$.

S6 Description of artificial data

We verified convergence of the proposed method on a number of artificial data sets. For the results presented in Figure 2, we generated artificial data by simulation according to the Gompertz model presented in the main text with the following parameters.

$$\mathbf{A} = \begin{pmatrix} 0.1593 & -0.0623 & 0.0163 \\ -0.1326 & 0.0684 & -0.0049 \\ -0.0135 & 0.0305 & 0.0213 \end{pmatrix} \tag{S6-1}$$

and

$$\mathbf{S}^{\mathrm{d}} = \begin{pmatrix} 0.012 & 0 & 0 \\ 0 & 0.024 & 0 \\ 0 & 0 & 0.0083 \end{pmatrix} \tag{S6-2}$$

and $S_k^{\mathrm{m}} = 0.001$ for all species. Simulations spanned $T = 256$ units of time on $N_{grid} = 8192$ number of grid points. With a simulation time step $\mathrm{d}t = T/N_{grid}$, the update equation reads [4]:

$$\boldsymbol{x}(t + \mathrm{d}t) = \boldsymbol{x}(t) - \mathrm{d}t\mathbf{A}\boldsymbol{x} + \sqrt{\mathbf{S}^{\mathrm{d}}}\boldsymbol{\zeta}\sqrt{\mathrm{d}t} \tag{S6-3}$$

where $\sqrt{\mathbf{S}^{\mathrm{d}}}$ has the same eigenvectors as $\mathbf{S}^{\mathrm{d}}$, but its eigenvalues are the square roots of those of $\mathbf{S}^{\mathrm{d}}$. $\boldsymbol{\zeta}$ has a standard Gaussian distribution ($\mathcal{N}(0, \mathbf{I})$). The simulation was started from its equilibrium density and trends added at the end to verify that estimates of the interaction coefficients were not affected by the presence of a trend. Specifically we added $0.5\sin(0.4\pi t_n/T)$ to $x_1(t_n)$ and $-0.35\cos(0.9\pi t_n/T)$ to $x_3(t_n)$, and no trend added for x_2.

S7 Supplementary figures

Figure S-4 (*next page*). **Likelihood profiles and contribution of prior.** Log-likelihood function ($-2\log P(\boldsymbol{y}|\boldsymbol{\theta})$) in blue and contribution of the prior ($-2\log P(\boldsymbol{\theta})$) in black for our experimental data, both with their minimum subtracted, as a function of each model parameter (indicated in top-left corners; $\rho_{ij}^d = S_{ij}^{\mathrm{d}}/\sqrt{S_{ii}^{\mathrm{d}}S_{jj}^{\mathrm{d}}}$). Shown in red: quadratic fit to the log-likelihood function.

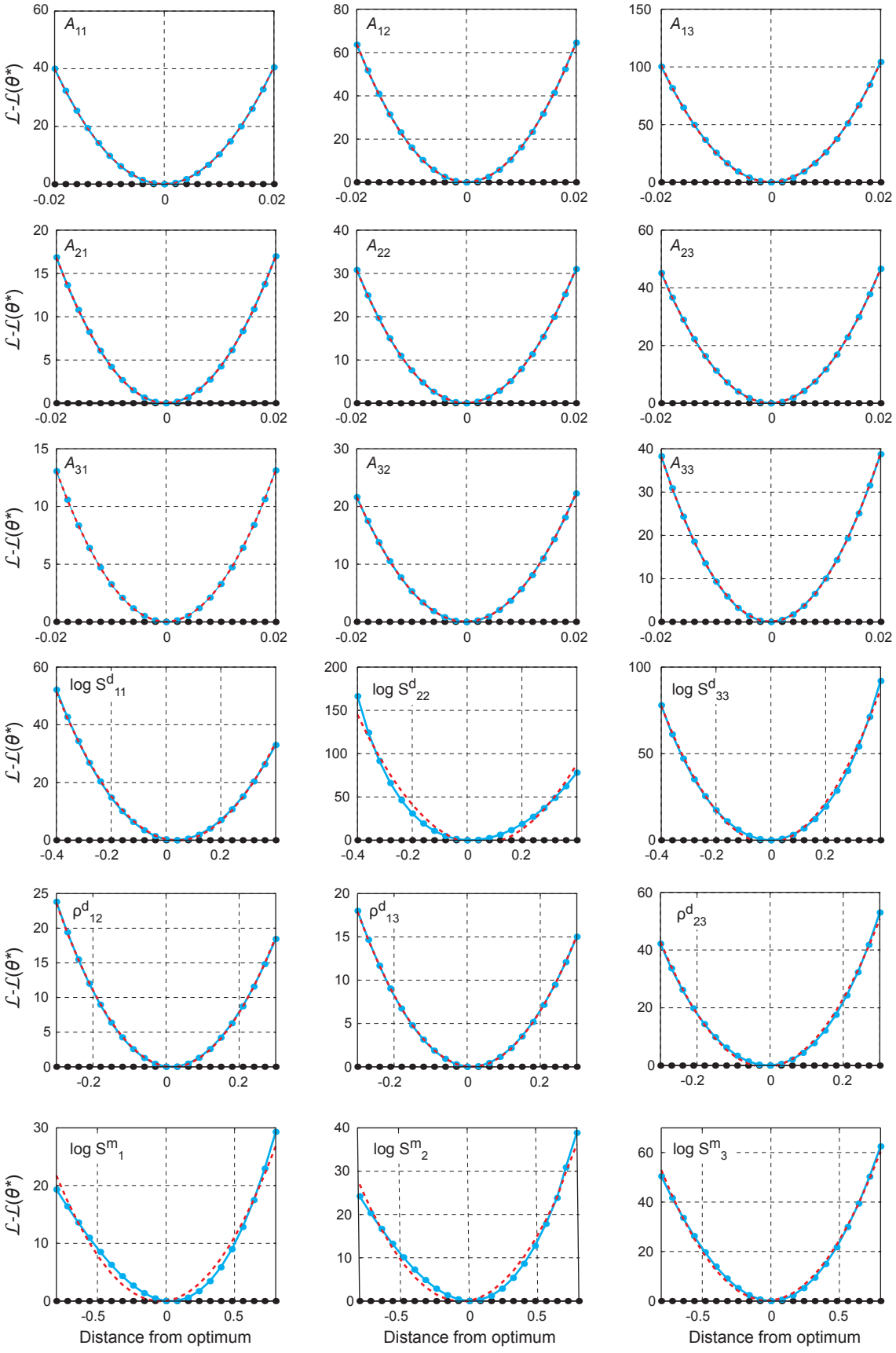
A_{11}
A_{12}
A_{13}
A_{21}
A_{22}
A_{23}
A_{31}
A_{32}
A_{33}
log S^d_{11}
log S^d_{22}
log S^d_{33}
ρ^d_{12}
ρ^d_{13}
ρ^d_{23}
log S^m_1
log S^m_2
log S^m_3
$\mathcal{L}$-$\mathcal{L}(\theta^*)$
Distance from optimum

Figure S-4:

References

[1] J. Barnard, R. McCulloch, and X. L. Meng, *Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage*, Statistica Sinica **10** (2000), 1281–1311.

[2] G. P. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*, 1st edition ed., Addison-Wesley Publishing Company, Reading, MA, 1973.

[3] D. R. Hekstra and S. Leibler, *Contingency and statistical laws in replicate microbial closed ecosystems*, Cell **149** (2012), no. 5, 1164–1173.

[4] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Review **43** (2001), 525–546.

[5] A. R. Ives, B. Dennis, K. L. Cottingham, and S. R. Carpenter, *Estimating community stability and ecological interactions from time-series data*, Ecological Monographs **73** (2003), no. 2, 301–330.

[6] R. E. Kass and L. Wasserman, *A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion*, Journal of the American Statistical Association **90** (1995), 928–934.

[7] K. B. Petersen and M. S. Pedersen, *The matrix cookbook*, available at http://www2.imm.dtu.dk/pubdb/p.php?3274, 2008.

[8] S. Särkkä, *Recursive bayesian inference on stochastic differential equations*, Ph.D. thesis, Helsinki University of Technology Laboratory of Computational Engineering, 2006.