

# PNAS



## Supporting Information for

### Inferring coexistence likelihood in changing environments from ecological time-series

Phuong L. Nguyen, Francesco Pomati and Rudolf P. Rohr

Corresponding Author Phuong L. Nguyen.  
E-mail: [linhphuong.nguyen@unifr.ch](mailto:linhphuong.nguyen@unifr.ch)

#### This PDF file includes:

Supporting text  
Figs. S1 to S53  
SI References

## Contents

<b>1</b>	<b>Supporting Material and Methods</b>	<b>3</b>
A	Difference between <i>per capita</i> interaction strength and Jacobian elements . . . . .	3
B	Weighted multilinear multivariate regression . . . . .	3
C	Cross validation . . . . .	5
D	Standard error of the parameters . . . . .	5
E	Computation of coexistence metrics and their confidence intervals . . . . .	5
E.1	Structural niche difference $\Omega$ . . . . .	5
E.2	Resistance angle $\eta_i$ . . . . .	6
E.3	Confidence interval for $\Omega$ and the $\eta_i$ . . . . .	6
F	Synthetic data . . . . .	6
G	Experimental data . . . . .	6
H	Observational data . . . . .	7
<b>2</b>	<b>Supporting Figures for Synthetic Data under Fixed Environment and Chaotic Dynamics</b>	<b>8</b>
<b>3</b>	<b>Supporting Figures for Synthetic Data under Fixed Environment and Cyclic Dynamics</b>	<b>12</b>
<b>4</b>	<b>Supporting Figures for Synthetic Data under Fixed Environment and Fixed Point Dynamics</b>	<b>16</b>
<b>5</b>	<b>Supporting Figures for Synthetic Data under Changing Environment and Chaotic Dynamics</b>	<b>20</b>
<b>6</b>	<b>Supporting Figures for Synthetic Data under Changing Environment and Cyclic Dynamics</b>	<b>27</b>
<b>7</b>	<b>Supporting Figures for Synthetic Data under Changing Environment and Fixed Point Dynamics</b>	<b>34</b>
<b>8</b>	<b>Supporting Figures for Synthetic Data with Extinction</b>	<b>41</b>
<b>9</b>	<b>Supporting Figures for Experimental Data</b>	<b>48</b>
A	Results using data from Blasius et al. 2020 . . . . .	48
B	Results using data from Yoshida et al. 2003 . . . . .	54
<b>10</b>	<b>Supporting Figures for Observational Data</b>	<b>58</b>
A	Population dynamics . . . . .	58
B	Cross-validation results . . . . .	59
C	Inference results – Ecological parameters . . . . .	60
D	Inference results – Coexistence metrics . . . . .	66

## Supporting Information Text

### 1. Supporting Material and Methods

**A. Difference between *per capita* interaction strength and Jacobian elements.** It is essential to not confuse the *per capita* interaction of the Lotka-Volterra model, inferred by the LV-map, with the Jacobian elements, inferred by the S-map. The S-map inference is based on  $n_i(t+1) = b_i(t) + \sum_{j=1}^S J_{ij}(t) \cdot n_j(t)$ , the intercept  $b_i$  carries no biological meaning, and the Jacobian elements can be related to  $r_i(t)$  and  $\alpha_{ij}(t)$  of the LV-map and population abundance  $n_i(t)$  as follows:

$$J_{ij}(t) = n_i(t+1) \cdot \alpha_{ij}(t) \quad i \neq j \quad [1]$$

$$J_{ii}(t) = \exp\left(r_i(t) + \sum_{j=1}^S \alpha_{ij}(t) \cdot n_j(t)\right) + n_i(t+1) \cdot \alpha_{ii}(t). \quad [2]$$

These Jacobian elements indicate the total effect, or “net” interaction, of one population on the growth of another population. Importantly, they are population densities dependent. Thus, a time variation of the Jacobian elements cannot be interpreted as a time variation of the *per capita* interaction (1).

**B. Weighted multilinear multivariate regression.** By using the *per capita* growth rate of the population  $\lambda_i(\mathbf{n}(t), \mathbf{e}(t))$  as the response variable, the LV-map is, in fact, a multivariate (or vector) nonlinear autoregressive process of order 1. Equation 1 in the main text is equivalent to

$$p_i(t+1) = p_i(t) + r_i(t) + \sum_{j=1}^S \alpha_{ij}(t) \cdot e^{p_j(t)}, \quad [3]$$

where  $p_i(t) = \ln n_i(t)$  is the logarithm of population densities. By assuming an additive noise, we obtain the first-order vector autoregressive process,  $p_i(t+1) = p_i(t) + r_i(t) + \sum_{j=1}^S \alpha_{ij}(t) \cdot e^{p_j(t)} + \epsilon_i(t)$ . The noise  $\epsilon_i(t)$  is assumed to follow a Gaussian distribution of mean zeros and variance  $\sigma_i^2$ , and independent of time  $t$ . Note that the variance  $\sigma_i^2$  depends on the population  $i$ , that is, biologically speaking, the amplitude of the random variability is population-dependent. A key point is that with our formulation, we assume additive noise on the logarithm of the population densities, which implies multiplicative noise on the population densities. From a biological perspective, this is a reasonable assumption, as the growing process of reproducing organisms is, by definition, a multiplicative process.

As the values of  $r_i(t)$  and  $\alpha_{ij}(t)$  are likely to vary with different conditions, and thus with time, and that we aim at inferring these dependences, we perform a local linear regression (see chapter 6 of Hastie et al. (2)). A local linear regression is a weighted linear regression at each time point  $t$ . A weighting kernel defines the weight of the variables at the other time points  $l$ , relative to the focal variable at time point  $t$ . Two types of weighting kernel can be defined by considering the distance in the state-space or in time.

The state-space weighting kernel, introduced in the S-map (3, 4) considers the difference between a variable at the time point  $t$  of focus and variables at another time point  $l$ . This weighting kernel is standard in empirical dynamical modelling, and is defined as

$$\omega_s(t, l) = e^{-\theta_s \cdot \frac{\|\mathbf{n}(t) - \mathbf{n}(l)\|}{d}}, \quad [4]$$

where  $\|\mathbf{n}(t) - \mathbf{n}(l)\|$  is the Euclidean distance between the vector of population densities  $\mathbf{n}(t)$  at time point  $t$  and the one at time point  $l$ , and  $d$  is the average Euclidean distance computed across all time-points  $l$ . The parameter  $\theta_s$  determines the strength of the weighting kernel. If  $\theta_s = 0$ , then the weight is put equal for all values of population densities. In contrast, if  $\theta_s > 0$ , more weight is put on similar values of population densities.

In turn, the time weighting kernel considers the time difference, and is defined as

$$\omega_t(t, l) = e^{-\theta_t \cdot \frac{\|t-l\|}{\bar{d}_t}}, \quad [5]$$

where  $\bar{d}_t$  is the average of all the time points in the time-series. Here, if  $\theta_t = 0$ , the weight is put equal for all time-points. In contrast,  $\theta_t > 0$  means that the weight is put more on time points that are closed to each other in time. Note, this formulation of the time weighting kernel is similar to the Laplace kernel used in moving average technics (5). The local regression technic is a generalization of moving average, allowing parameters to change with time.

The weighting parameters  $\theta_s$  and  $\theta_t$  are not directly comparable as they represent different types of weighting. Specifically,  $\theta_s$  determines the weighting in the state-space of population densities, while  $\theta_t$  controls the weighting over time. In the special case where  $\theta_s = 0$  and  $\theta_t = 0$ , the inferred parameters remain constant across time, and the model follows the standard Lotka-Volterra model, i.e., the intrinsic growth rates and per capita interaction strengths are constant with population densities and time. Larger values of  $\theta_s$  imply that only data points, with population densities very similar to a focal time point  $t$ , are

taken into the local linear regression. This makes the fit more localized in state-space, allowing parameters to vary more across state-space. On the other hand, larger values of  $\theta_t$  suggests that data points closer in time, to a focal time  $t$ , are given greater weight. This makes the least squares fit more localized in time, causing parameters to vary more overtime.

The parameters  $r_i(t)$  and  $\alpha_{ij}(t)$ , for each time points  $t$ , are estimated by ordinary least square regression with weights (2, 6). That is, the fitted parameters minimize the weighted residuals sum of squares, which is defined by

$$\text{RSS}(t) = \sum_{i,l} \omega(t,l)^2 \cdot \left( p_i(l+1) - \left( p_i(l) + r_i(t) + \sum_{j=1}^S \alpha_{ij}(t) \cdot e^{p_j(l)} \right) \right)^2, \quad [6]$$

with  $\omega(t,l)$  denotes either the state-space or the time weighting kernel. Note that we have chosen the convention of squaring the weights ( $\omega(t,l)^2$ ) instead of the non-squaring standard convention ( $\omega(t,l)$ ). This choice has no influence on the results, as the square of an exponential equals the exponential of two times the exponents, which results in simply rescaling the  $\theta$  by a factor two.

By putting, back  $n_i(t)$  in the above equation, we obtain

$$\begin{aligned} \text{RSS}(t) &= \sum_{i,l} \omega(t,l)^2 \cdot \left( \ln \left( \frac{n_i(l+1)}{n_i(l)} \right) - r_i(t) - \sum_{j=1}^S \alpha_{ij}(t) \cdot n_j(l) \right)^2 \\ &= \sum_{i,l} \left( \omega(t,l) \cdot \ln \left( \frac{n_i(l+1)}{n_i(l)} \right) - \omega(t,l) \cdot r_i(t) - \sum_{j=1}^S \omega(t,l) \cdot \alpha_{ij}(t) \cdot n_j(l) \right)^2, \end{aligned} \quad [7]$$

which is a weighted linear regression with  $\frac{n_i(l+1)}{n_i(l)}$  as response variables and  $n_j(l)$  as the explanatory variables.

It can be further written in a matrix format as,

$$\text{RSS}(t) = \sum_{i,l} \left( \tilde{\mathbf{Y}}(t)_{il} - (\tilde{\mathbf{X}}(t) \cdot \beta(t))_{il} \right)^2, \quad [8]$$

with  $\tilde{\mathbf{Y}}(t)$  and  $\tilde{\mathbf{X}}(t)$ , respectively, as the matrix of weighted response and explanatory variables. Given that

$$\tilde{\mathbf{y}}_j^T = \left( \omega(t,1) \cdot \ln \frac{n_j(2)}{n_j(1)}, \omega(t,2) \cdot \ln \frac{n_j(3)}{n_j(2)}, \dots, \omega(t,T-1) \cdot \ln \frac{n_j(T)}{n_j(T-1)} \right) \quad [9]$$

the column vector  $j$  of matrix  $\tilde{\mathbf{Y}}(t)$ , and

$$\tilde{\mathbf{x}}_i = (\omega(t,i) \cdot 1, \omega(t,i) \cdot n_1(i), \dots, \omega(t,i) \cdot n_s(i)) \quad [10]$$

the row vector  $i$  of matrix  $\tilde{\mathbf{X}}(t)$ . Here,  $\tilde{\mathbf{Y}}(t)$  is of size  $(T-1) \times S$  and  $\tilde{\mathbf{X}}(t)$  is of size  $(T-1) \times (S+1)$ , where  $T$  is the total data points of the time-series. All parameters at time  $t$  are encoded in the parameter matrix  $\beta(t)$ .

From these two matrices, we compute the least square estimation of  $\beta(t)$  by matrix computation:

$$\hat{\beta}(t) = (\tilde{\mathbf{X}}^T(t) \cdot \tilde{\mathbf{X}}(t))^{-1} \cdot \tilde{\mathbf{X}}^T(t) \cdot \tilde{\mathbf{Y}}(t). \quad [11]$$

The matrix  $\hat{\beta}(t)$  contains the estimations of the intrinsic growth rates of all populations as well as all *per capita* interaction strength. For instance, its column  $j$  is given by the column vector  $\hat{\mathbf{b}}_j^T = (\hat{r}_j(t), \hat{\alpha}_{j1}(t), \dots, \hat{\alpha}_{jS}(t))$  and provides the estimations of the intrinsic growth rate of the population  $j$  and the interaction of other populations in the communities on the population  $j$ . Note that  $\hat{\beta}(t)$  is a matrix of size  $(S+1) \times S$ . The mathematical proof can be found in chapters 3.2 and 6 of Hastie et al. (2) and chapter 3.3 of Mardia et al. (7). The strength of the weighting kernel,  $\theta$ , is determined by cross validation. Finally, we provided an estimation of the standard error for each  $\hat{r}_i(t)$  and  $\hat{\alpha}_{ij}(t)$ .

In our manuscript, we have chosen to use ordinary least square regression with weighting kernels. But generalized least square regressions could also be used to incorporate heteroscedasticity or more complicated structure into the residuals (see chapter 4.1 of Fahrmeir et al. (6)).

**C. Cross validation.** The weighting parameters,  $\theta_s$  and  $\theta_t$ , are estimated using cross-validation techniques. As the data points in time-series data are dependent, we cannot use classical cross validation methods that randomly split the training data (or in sample data) and the test data (or out of sample data). Instead, we use the cross validation on rolling origin forecast.

This cross validation technique uses the first- $t$  observation of population abundance ( $\mathbf{n}(l)$ , for  $l = 1, \dots, t$ ) to predict the population abundance  $\hat{\mathbf{n}}(t+1)$  at time step  $t+1$ . This process of prediction is iterated for  $t$  from  $T_s$  to  $T-1$ . The initial time step  $T_s$  is usually chosen as 10% of the total number of time point  $T$ , i.e.,  $T_s = \text{round}(p \cdot T)$  with  $p = 0.1$ . The best  $\theta$  is the one minimizing the root mean sum of error squares (RMSE). The expression of the RMSE is given by

$$\text{RMSE} = \sqrt{\frac{\sum_{t=T_s}^{T-1} (\|\hat{\mathbf{n}}(t+1) - \mathbf{n}(t+1)\|^2)}{T-1-T_s}}. \quad [12]$$

The prediction of  $\hat{\mathbf{n}}(t+1)$  is done as follows. Knowing the abundances from time 1 to time  $t$ , we can estimate  $\hat{\mathbf{r}}(k)$  and  $\hat{\boldsymbol{\alpha}}(k)$  from time 1 to time  $t-1$ . Note that we cannot estimate  $\mathbf{r}(t)$  and  $\boldsymbol{\alpha}(t)$ , as it would require the knowledge of  $\mathbf{n}(t+1)$ . Consequently, we use  $\hat{\mathbf{r}}(t-1)$ ,  $\hat{\boldsymbol{\alpha}}(t-1)$  and  $\mathbf{n}(t)$  to estimate  $\hat{\mathbf{n}}(t+1)$  as follows

$$\hat{n}_i(t+1) = n_i(t) \cdot \exp\left(r_i(t-1) + \sum_{i=1}^S \alpha_{ij}(t-1) \cdot n_j(t)\right). \quad [13]$$

**D. Standard error of the parameters.** The standard error (SE) of the intercept (intrinsic growth rate;  $r_i$ ) and coefficient (inter and intraspecific interaction;  $\alpha_{ij}$ ) follows the statistics of conventional multivariate regression method. The linear regression assumes that the deviation of the response variable  $\tilde{\mathbf{Y}}$  from its predicted values  $\hat{\mathbf{Y}}$  follows a normal distribution, therefore for each column  $i$  of the response matrix  $\tilde{\mathbf{Y}}$  we have

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i. \quad [14]$$

The vector  $\boldsymbol{\beta}_i$  corresponds to the column  $i$  of the matrix of parameters  $\boldsymbol{\beta}$ . The residuals  $\boldsymbol{\epsilon}_i$  are assumed to be independent and identically distributed and follow a normal distribution with mean zero and variance  $\sigma_i^2$ . The residual variance  $\sigma_i^2$  can be estimated by

$$\sigma_i^2 = \frac{1}{T-2-S} \sum_j^{T-1} (y_{ij} - \hat{y}_{ij})^2 \quad [15]$$

Then the estimation  $\hat{\boldsymbol{\beta}}_i$  of the parameters  $\boldsymbol{\beta}_i$  follows a normal distribution of mean  $\boldsymbol{\beta}$  and variance-covariance matrix given by

$$\sigma_i^2 \cdot (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}. \quad [16]$$

And thus, the square roots of the diagonal elements of this variance-covariance matrix are the standard error of  $\hat{\boldsymbol{\beta}}_i$ . See chapter 3.2 of (2) and chapter 3.3 of (7). The computation of the standard errors is done at each time point  $t$ .

## E. Computation of coexistence metrics and their confidence intervals.

**E.1. Structural niche difference  $\Omega$ .** In dimension  $S$ , i.e., with  $S$  species, following Ribando (8), the solid angle of the cone of feasibility can be computed by the following multiple integral (9):

$$\Omega = \frac{2^S |\det(\boldsymbol{\alpha})|}{\pi^{n/2}} \int \dots \int_{\mathbf{R}_{\geq 0}^S} e^{-\mathbf{x}^T \boldsymbol{\alpha}^T \boldsymbol{\alpha} \mathbf{x}} d\mathbf{x}. \quad [17]$$

This can be viewed as the cumulative function of a multivariate normal distribution of mean 0 and variance-covariance matrix  $(\boldsymbol{\alpha}^T \boldsymbol{\alpha})^{-1}$ . As, there is no close formula for its evolution, we used the quasi Monte-Carlo method developed by Genz (10) and implemented in R (11). Although this works for any dimension, for  $S = 2$  and  $S = 3$ , there exist close formal. For  $S = 2$ , we can simply compute  $\Omega$  based on the scalar product of the two columns vector in the interaction matrix:

$$\Omega = \frac{1}{2\pi} \arccos \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}. \quad [18]$$

For  $S = 3$ , the exact formula is given by (12)

$$\Omega = \frac{1}{2\pi} \arctan \frac{|\det \boldsymbol{\alpha}|}{\|\mathbf{v}_1\| \|\mathbf{v}_2\| \|\mathbf{v}_3\| + \|\mathbf{v}_1\| \|\mathbf{v}_2 \cdot \mathbf{v}_3\| + \|\mathbf{v}_2\| \|\mathbf{v}_1 \cdot \mathbf{v}_3\| + \|\mathbf{v}_3\| \|\mathbf{v}_1 \cdot \mathbf{v}_2\|}. \quad [19]$$

Note that we standardized  $\Omega$  such that  $\Omega = 1$  if the cone were to be the full sphere. Second, as the value of  $\Omega$ , usually, scales as the power of the number of species  $S$ , we reported, and we recommend to report, the logarithm of it.

**E.2. Resistance angle  $\eta_i$ .** The resistance angle of species  $i$  is the smallest angle  $\eta_i$  between the vector of intrinsic growth rate and the border of the feasibility domain at which this species becomes extinct (13). This border is on the opposite of the vector  $-\mathbf{v}_i$ , that is, the hyperplane generated by the vectors  $-\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_S$ . This angle is given by the angle made by the orthogonal projection of the intrinsic growth rate vector on that hyperplane and the intrinsic growth rate vector (13). Then the sign and the resistance angle, positive when  $r_i$  is inside feasibility cone and negative otherwise, is determined by projecting  $\mathbf{r}$  on the direction orthogonal to the above hyperplane, and then determining if this projection points toward  $-\mathbf{v}_i$  (in which case  $r_i$  is inside the feasibility cone) or the opposite.

**E.3. Confidence interval for  $\Omega$  and the  $\eta_i$ .** The confidence intervals for  $\Omega$  and the  $\eta_i$  are estimated by parametric bootstrap (14). Structural niche difference  $\Omega$  and resistance angles  $\eta_i$  are determined by the intrinsic growth rates vector  $\mathbf{r}$  and the matrix of *per capita* interaction strengths  $\alpha$ , which are inferred from the abundance time-series by the LV-map. As explained in the previous section, the inferred parameters follow a multivariate normal distribution, i.e., for each species  $i$ , the vector

$$\begin{bmatrix} r_i \\ \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{iS} \end{bmatrix} \sim \mathcal{N} \left( \text{mean} = \begin{bmatrix} \hat{r}_i \\ \hat{\alpha}_{i1} \\ \hat{\alpha}_{i2} \\ \vdots \\ \hat{\alpha}_{iS} \end{bmatrix}, \text{variance-covariance} = \sigma_i^2 \cdot (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \right). \quad [20]$$

Therefore, we can parametrically resample  $\mathbf{r}$  vectors and  $\alpha$  matrices, and for each resampling, compute the corresponding  $\Omega$  and  $\eta_i$ . This generates the bootstrapped distribution of  $\Omega$  and  $\eta_i$ , from which we can estimate their 95%, by computing as lower bound the 2.5% quantile and as upper bound the 97.5% quantile.

**F. Synthetic data.** Synthetic data is simulated from a discrete time Lotka-Volterra model with three competing population and environmental noise. The equation is given by:

$$n_i(t+1) = n_i(t) \cdot \exp \left( r_i + \sum_{j=1}^3 \alpha_{ij} \cdot n_j(t) + \epsilon_i(t) \right) \quad i = 1, 2, 3. \quad [21]$$

The parameters  $r_i$  and  $\alpha_{ij}$  are the intrinsic growth rate and the *per capita* interaction strengths, respectively. The random variables  $\epsilon_i(t)$  represent the environmental noise, which are drawn independently at random from a centered normal distribution and of standard deviation proportional to the  $r_i$ , i.e.,  $\epsilon_i(t) \sim \mathcal{N}(0, r_i \cdot \sigma)$ . The parameter  $\sigma$  determines the overall environmental noise level. We use all points that are generated by the simulation. Note that this equation is a time-discrete version with environmental noise of the Lotka-Volterra model, which can also be equivalent to a multispecies Ricker model with any type of interaction and environmental noise (15).

**G. Experimental data.** The time-series experimental data from Blasius et al. (16) were obtained from open-source data shared by the authors on figshare (<https://doi.org/10.6084/m9.figshare.10045976.v1>). These experiments were conducted at the clonal level of algae. Data from nine experiments (C1–C9) were used in our analysis.

- Experiments C1–C4: Conducted under identical conditions with a constant nutrient influx of 80  $\mu\text{mol/l}$  and an outflux rate of 0.55/day, categorized as low-nutrient conditions.
- Experiment C5: Maintained a constant nutrient influx of 80  $\mu\text{mol/l}$  but had an increased outflux rate of 0.66/day, categorized as high outflux conditions.
- Experiments C6 and C7: Alternated nutrient influx between 160  $\mu\text{mol/l}$  and 0  $\mu\text{mol/l}$  every eight days, categorized as forced-nutrient conditions.
- Experiments C8 and C9: Had a high constant nutrient influx of 160  $\mu\text{mol/l}$ , categorized as high-nutrient conditions.

Note that in experiments C1–C7, *Monoraphidium minutum* was used, whereas *Chlorella vulgaris* was used in experiments C8 and C9.

We interpolate the missing data points. In particular, seven missing points for experiment C1, two missing points for experiments C2 and C3, twelve missing points for experiment C5, one missing point for experiments C6 and C7. In addition, we replace the zero density of rotifers by the minimum value of rotifer densities of the corresponding experiment and divide it by 8. Finally, to standardize time-series length, we only kept data from the first 139 days across all experiments, except for C5 where the time-series is too short (85 days).

The time-series experimental data from Yoshida et al. (17) were obtained using PlotDigitizer app <https://plotdigitizer.com/app>. Screenshots of figures from the paper were uploaded to the PlotDigitizer platform, and the data points were manually extracted. In this experiment, only *C. vulgaris* was used.

**H. Observational data.** Plankton high-frequency data were collected from Lake Greifensee, Switzerland, by a dual-magnification dark-field imaging microscope (18). Pelagic plankton images in the size range between  $\sim 10 \mu\text{m}$  and  $\sim 1 \text{ cm}$  were collected at 3 m depth for 10 minutes every hour, and abundances (as regions of interest per second, ROI/s) were aggregated (summed) per day. For this study, we used data collected between March 2019 and June 2023. We classified taxa using a deep-learning classifier (19) (the code for the classification can be found in <https://github.com/kspruthviraj/Plankiformer>), and focused on five aggregated guilds of phytoplankton: Cyanobacteria, Green algae, Diatoms, Crypsophytes, and Cryptophytes. We interpolate the missing data, including 50 points out of 1558 points for Green algae, 47 points out of 1558 points for the other four guilds.

## 2. Supporting Figures for Synthetic Data under Fixed Environment and Chaotic Dynamics

For the chaotic dynamics, we used the following parameters:

$$\mathbf{r} = \begin{bmatrix} 2.7 \\ 3.24 \\ 4.59 \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} -4.05 & -0.621 & -0.27 \\ -0.81 & -2.7 & -0.81 \\ -1.08 & -1.755 & -2.70 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.005 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.02 \cdot r_i^2)$ .

On each figure:

- Top panel: population dynamics.
- Left column of below panels: results with the state-space weighing kernel.
- Right row of below panels: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate the true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

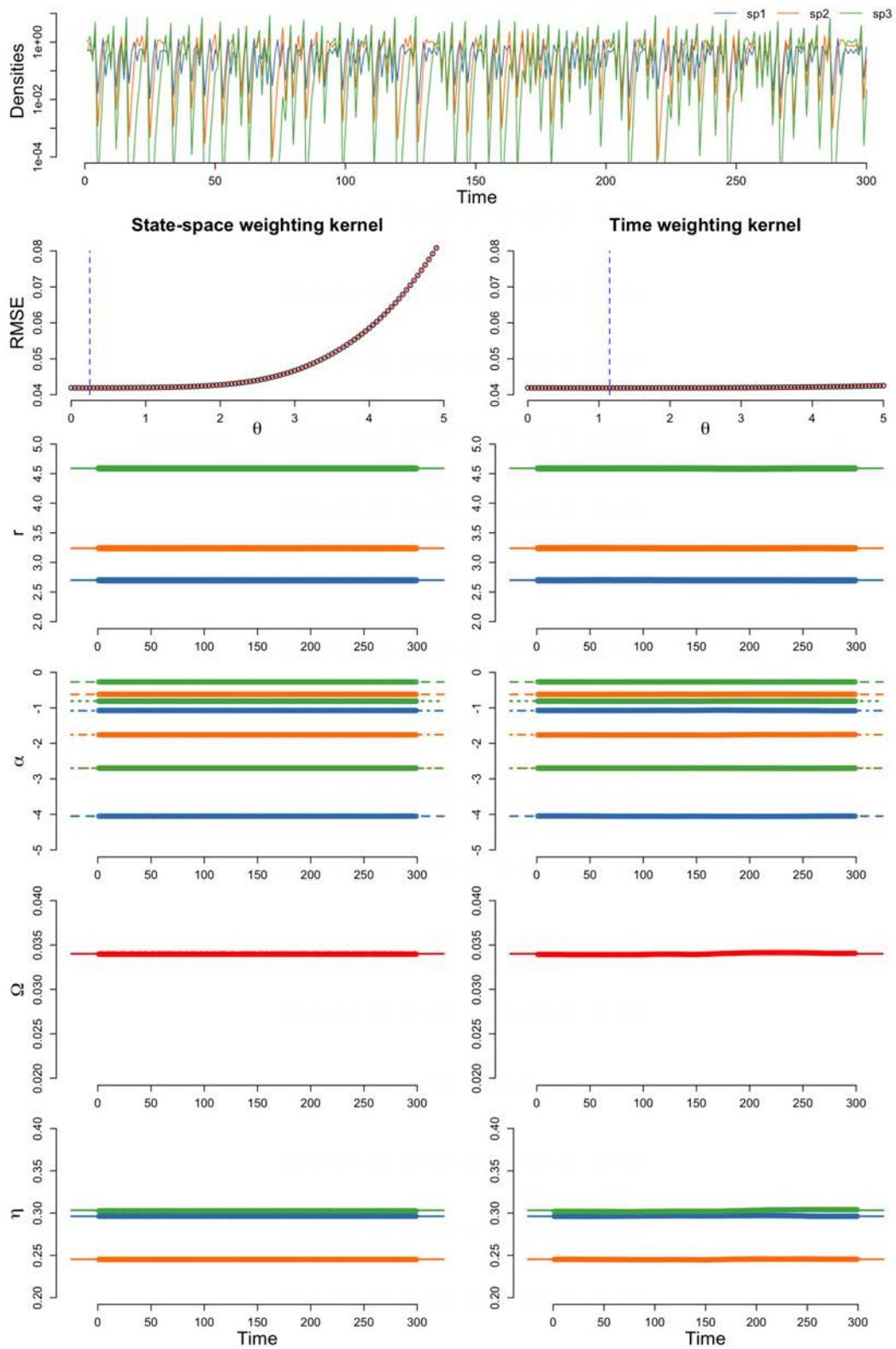


Fig. S1. Fixed environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.005$ .

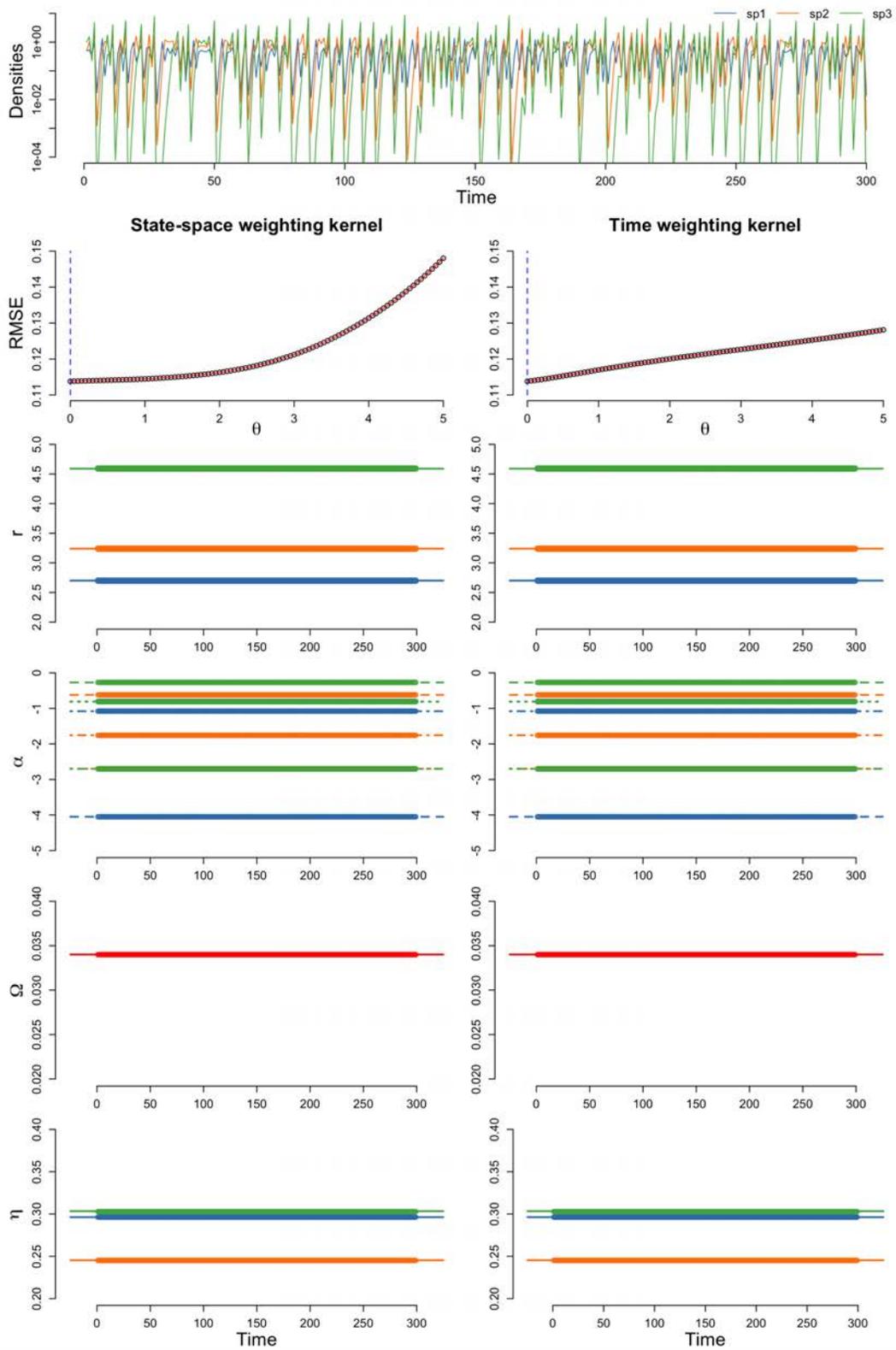


Fig. S2. Fixed environment and chaotic dynamics, with intermediate environmental noise  $\epsilon = 0.01$ .

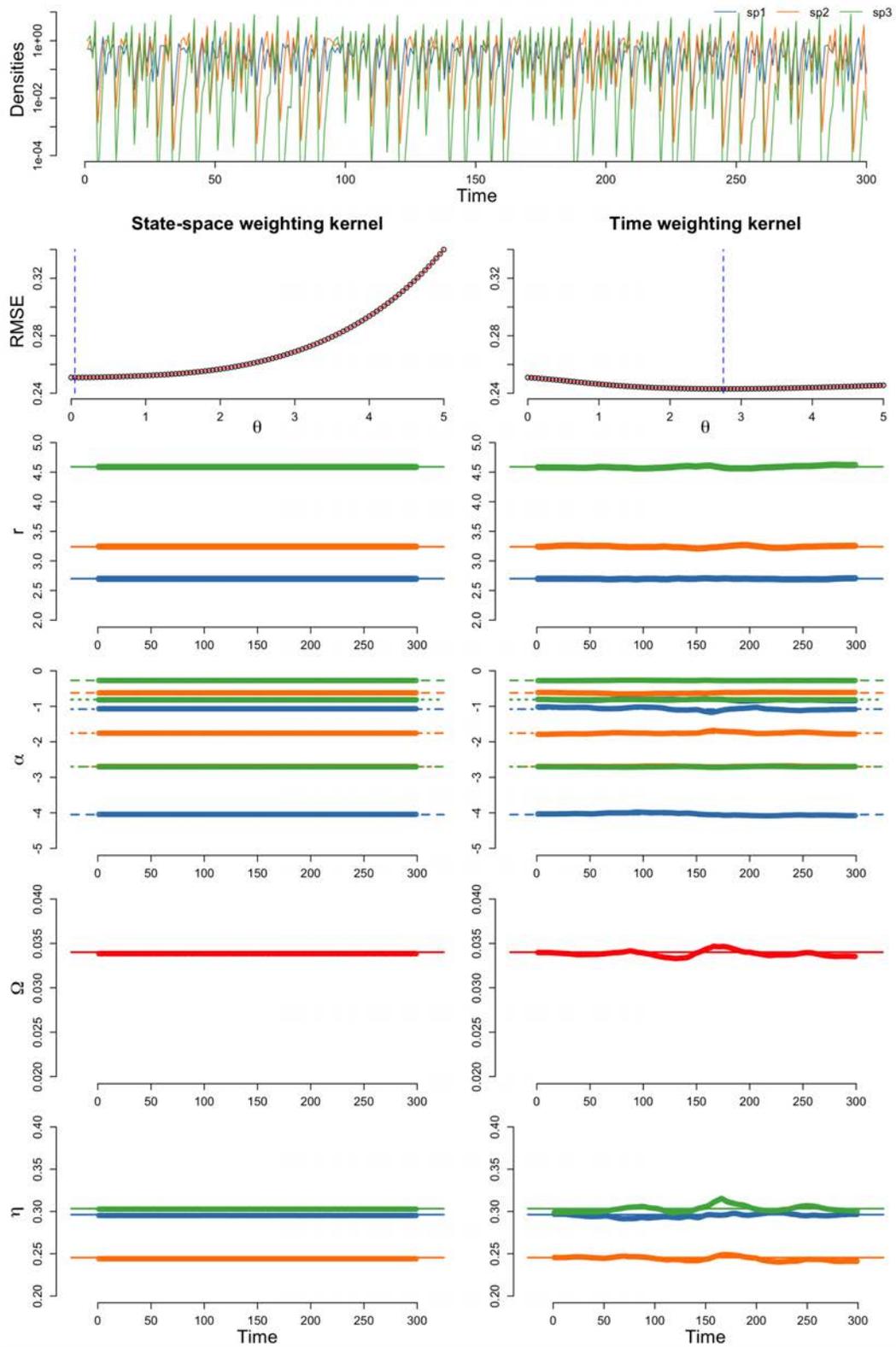


Fig. S3. Fixed environment and chaotic dynamics, with high environmental noise  $\epsilon = 0.02$ .

### 3. Supporting Figures for Synthetic Data under Fixed Environment and Cyclic Dynamics

For the cyclic dynamics, we used the following parameters:

$$\mathbf{r} = \begin{bmatrix} 3.8 \\ 0.3 \\ 2.4 \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} -0.2 & -0.4 & -0.8 \\ -0.011 & -0.033 & -0.067 \\ -0.2 & -0.2 & -0.6 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.005 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.02 \cdot r_i^2)$ .

On each figure:

- Top panel: population dynamics.
- Left column of below panels: results with the state-space weighing kernel.
- Right row of below panels: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

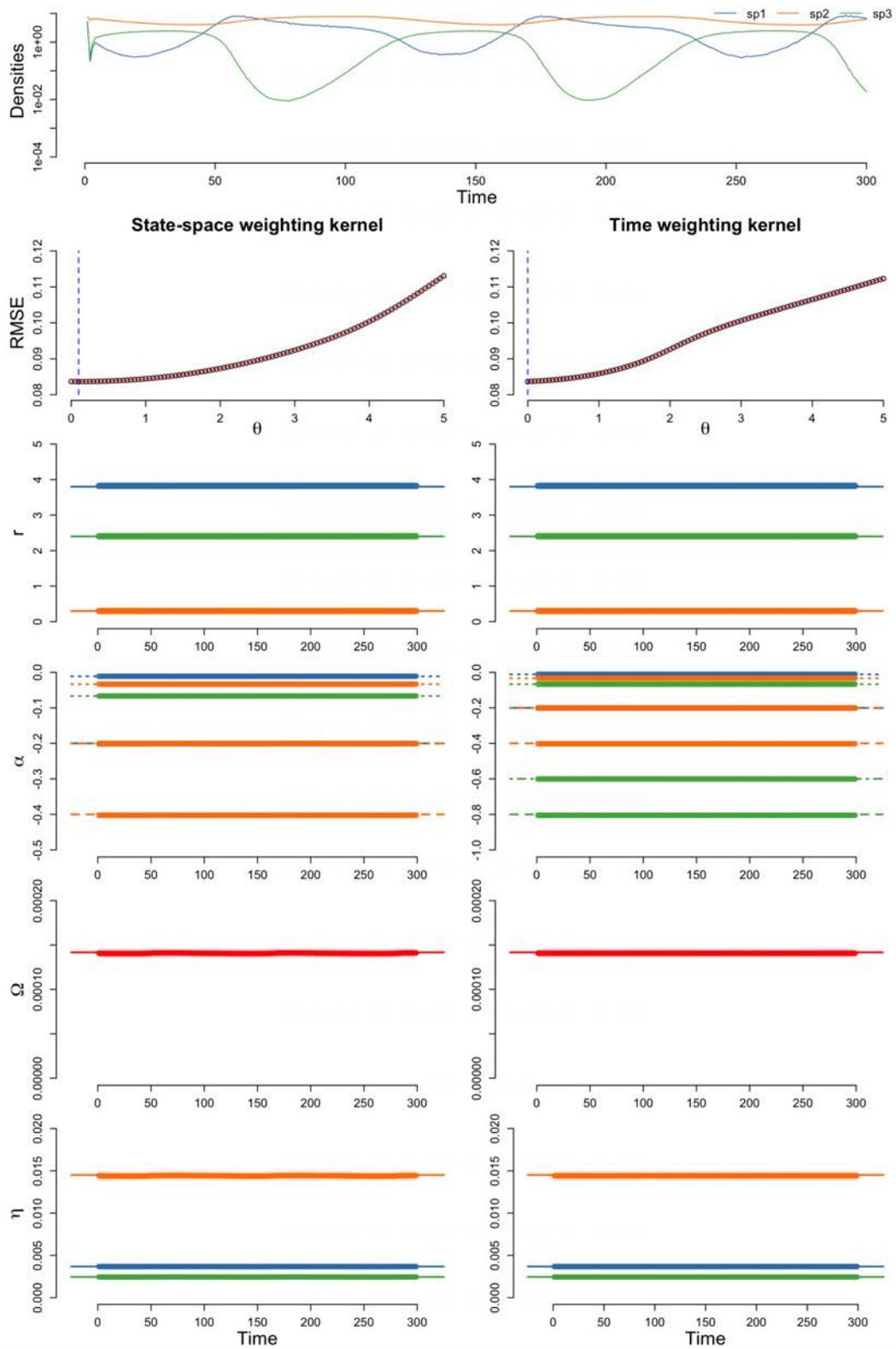


Fig. S4. Fixed environment and cyclic dynamics, with small environmental noise  $\epsilon = 0.005$ .

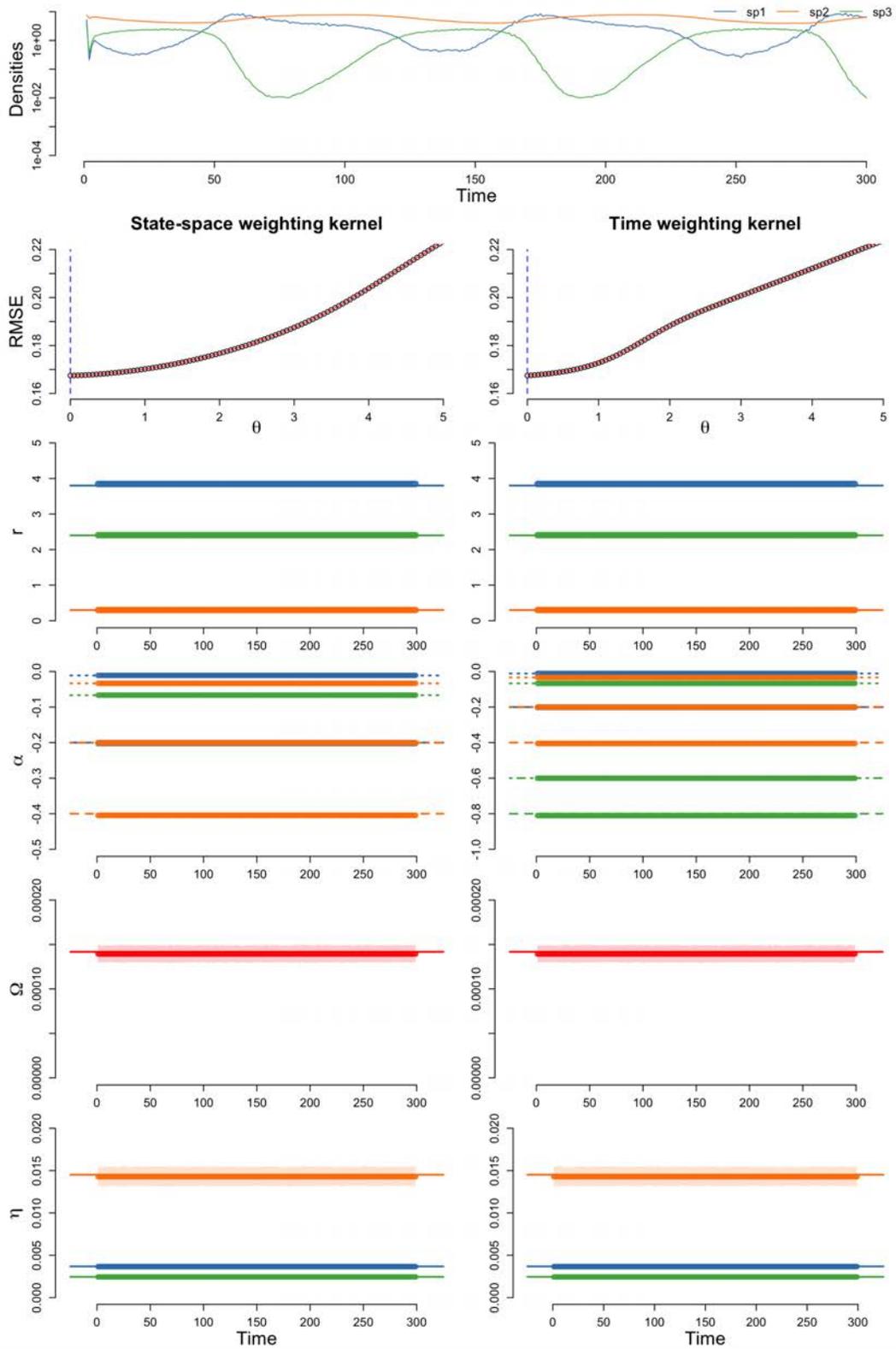


Fig. S5. Fixed environment and cyclic dynamics, with intermediate environmental noise  $\epsilon = 0.01$ .

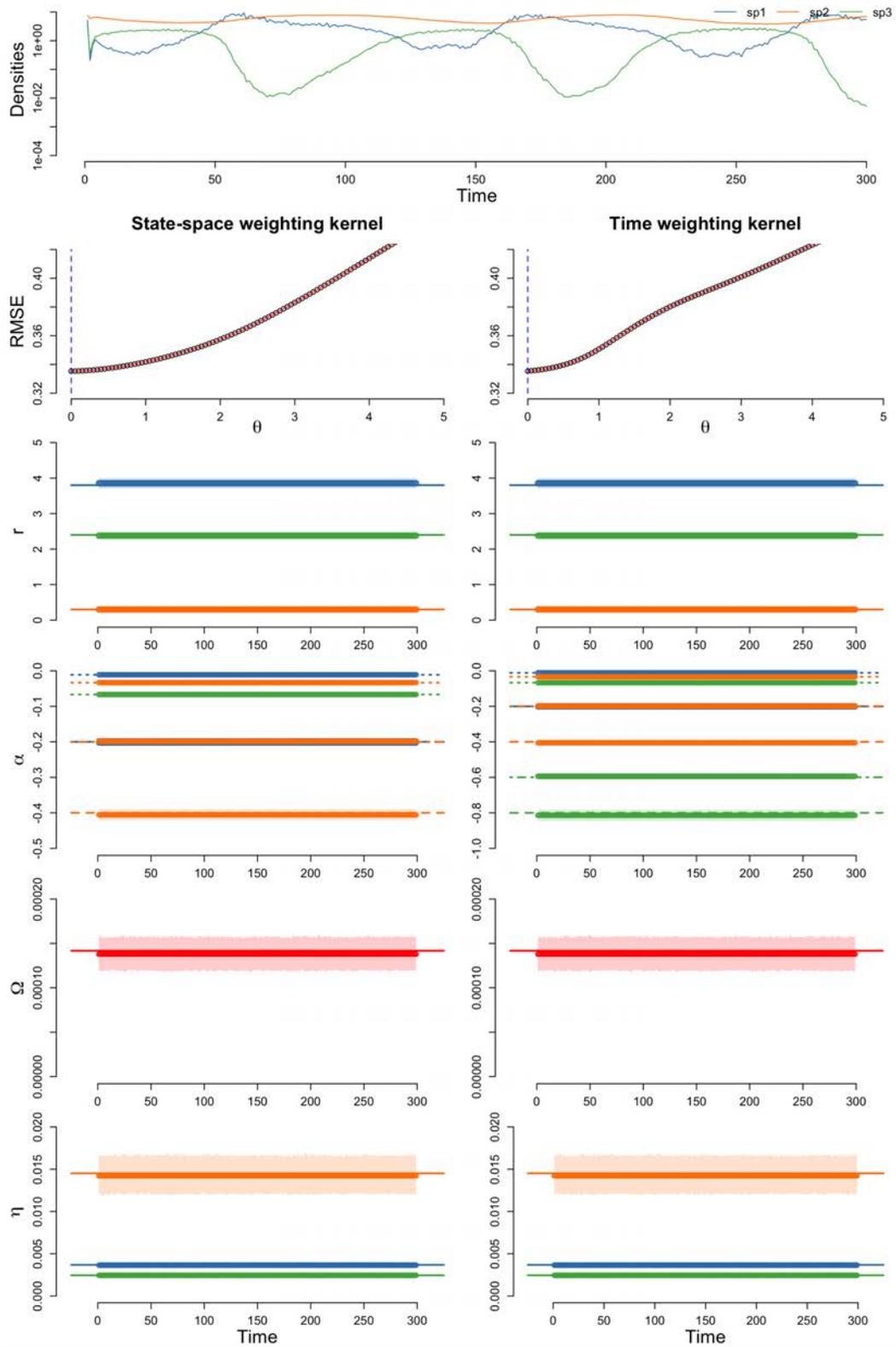


Fig. S6. Fixed environment and cyclic dynamics, with high environmental noise  $\epsilon = 0.02$ .

#### 4. Supporting Figures for Synthetic Data under Fixed Environment and Fixed Point Dynamics

For the fixed point dynamics, we used the following parameters:

$$\mathbf{r} = \begin{bmatrix} 0.7 \\ 0.81 \\ 1.19 \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} -1.05 & -0.161 & -0.07 \\ -0.21 & -0.7 & -0.21 \\ -0.28 & -0.455 & -0.7 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.005 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.02 \cdot r_i^2)$ .

On each figure:

- Top panel: population dynamics.
- Left column of below panels: results with the state-space weighing kernel.
- Right row of below panels: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

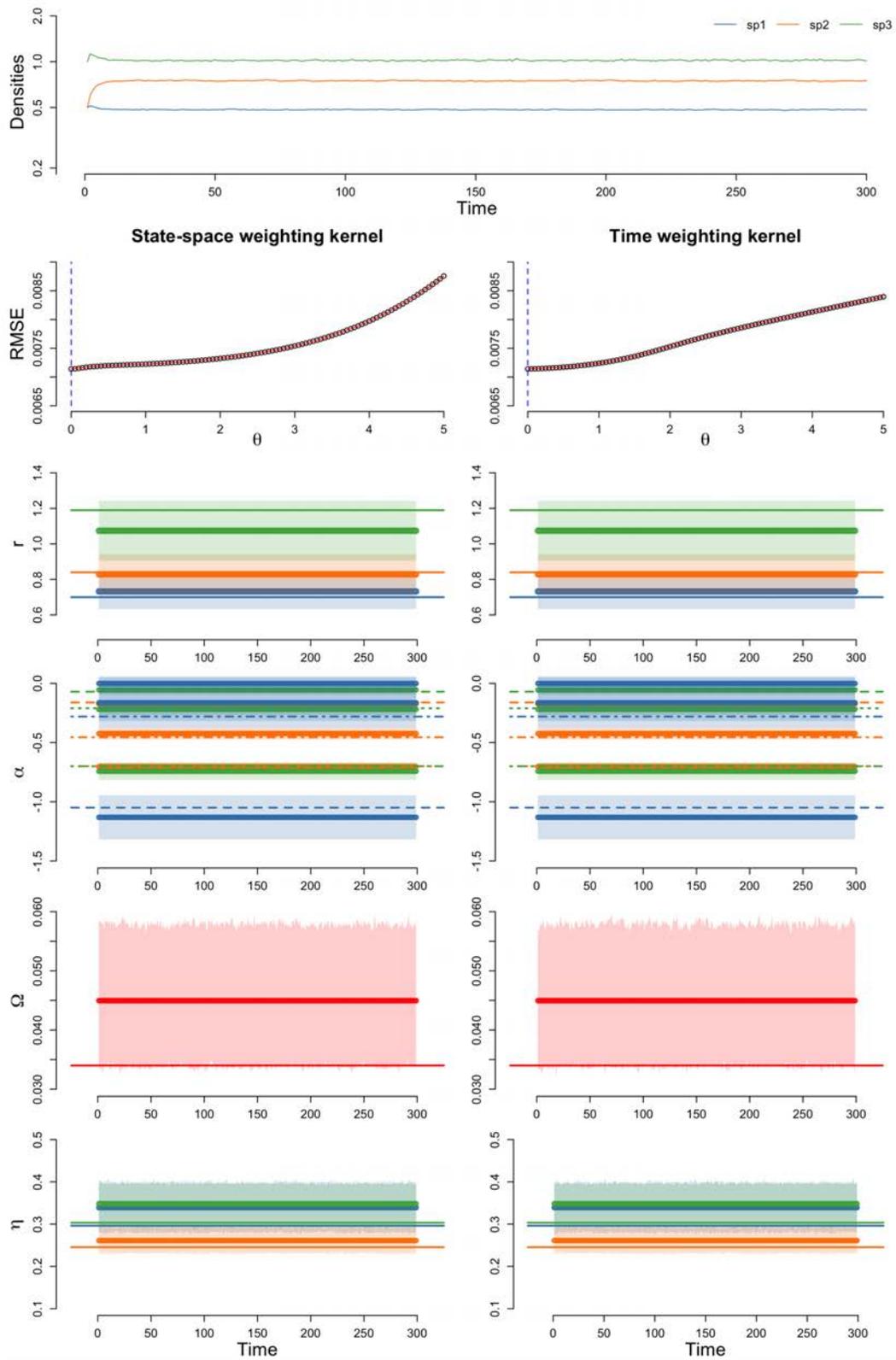


Fig. S7. Fixed environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.005$ .

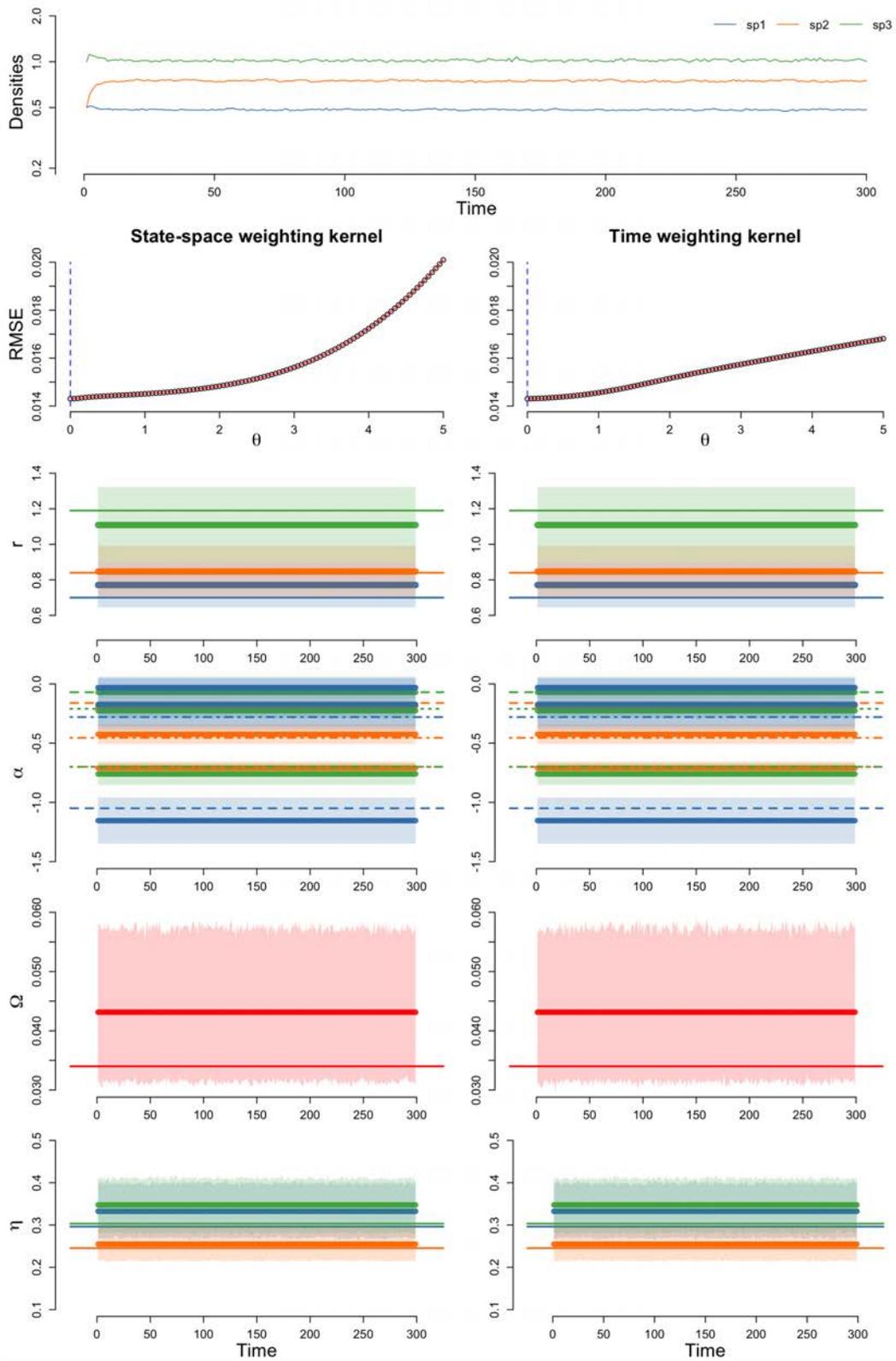


Fig. S8. Fixed environment and fixed point dynamics, with intermediate environmental noise  $\epsilon = 0.01$ .

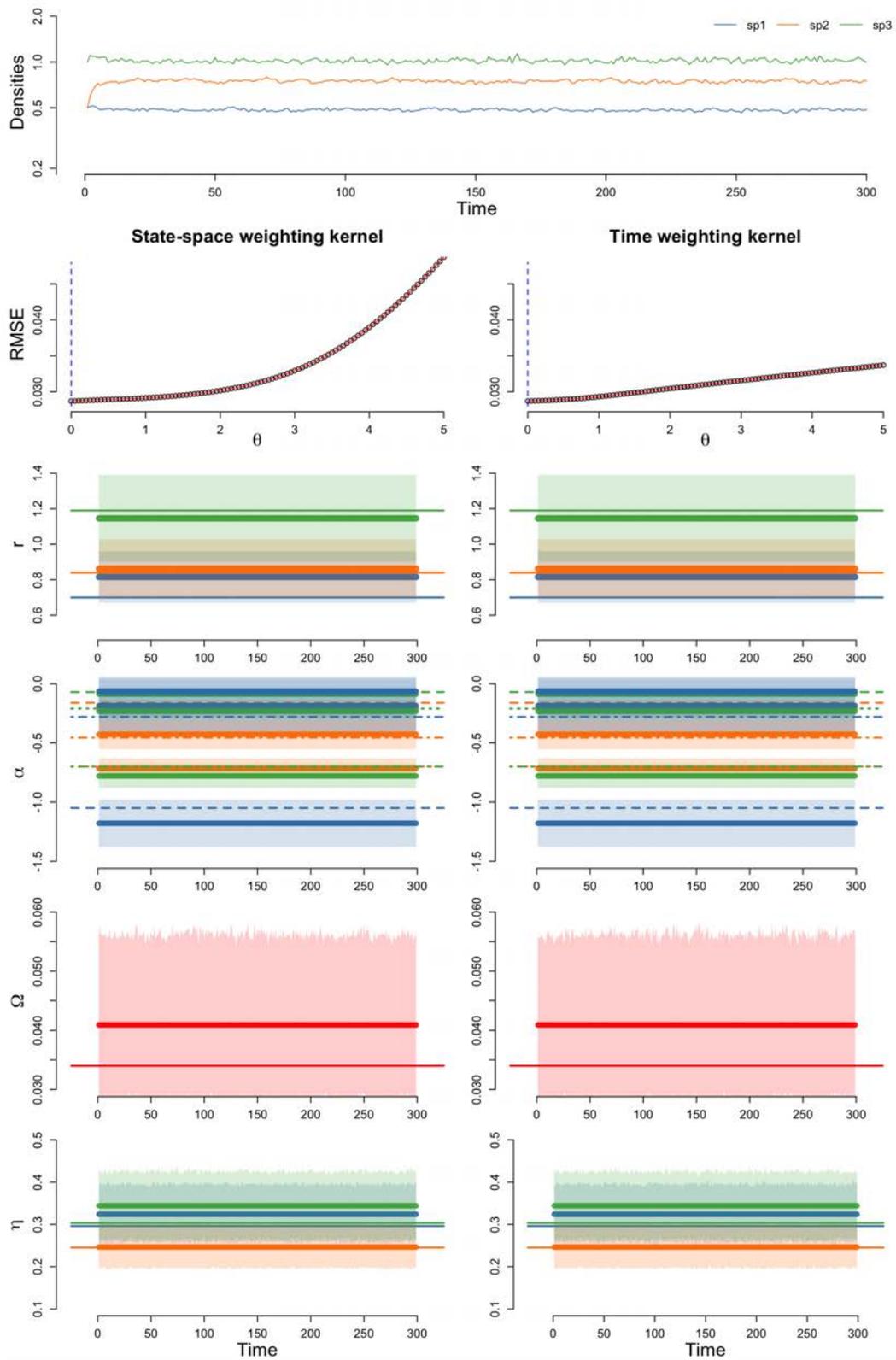


Fig. S9. Fixed environment and fixed point dynamics, with high environmental noise  $\epsilon = 0.02$ .

## 5. Supporting Figures for Synthetic Data under Changing Environment and Chaotic Dynamics

For the chaotic dynamics, we used the following parameters in environment 1 and environment 2, respectively:

$$\mathbf{r}_1 = \begin{bmatrix} 2.7 \\ 3.24 \\ 3.24 \end{bmatrix} \quad \mathbf{r}_2 = \begin{bmatrix} 2.7 \\ 3.24 \\ 1.89 \end{bmatrix}$$
$$\boldsymbol{\alpha}_1 = \begin{bmatrix} -4.05 & -0.621 & -0.27 \\ -0.81 & -2.7 & -0.81 \\ -1.08 & -1.755 & -2.70 \end{bmatrix} \quad \boldsymbol{\alpha}_2 = \begin{bmatrix} -4.05 & -0.621 & -0.27 \\ -0.81 & -2.7 & -0.81 \\ -1.08 & -0.945 & -2.70 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.005 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.007 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ .

For each level of environmental noise, two figures are provided.

The first figure shows the population dynamics as function of time and in the state-space.

The second figure shows:

- Left column: results with the state-space weighing kernel.
- Right column: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

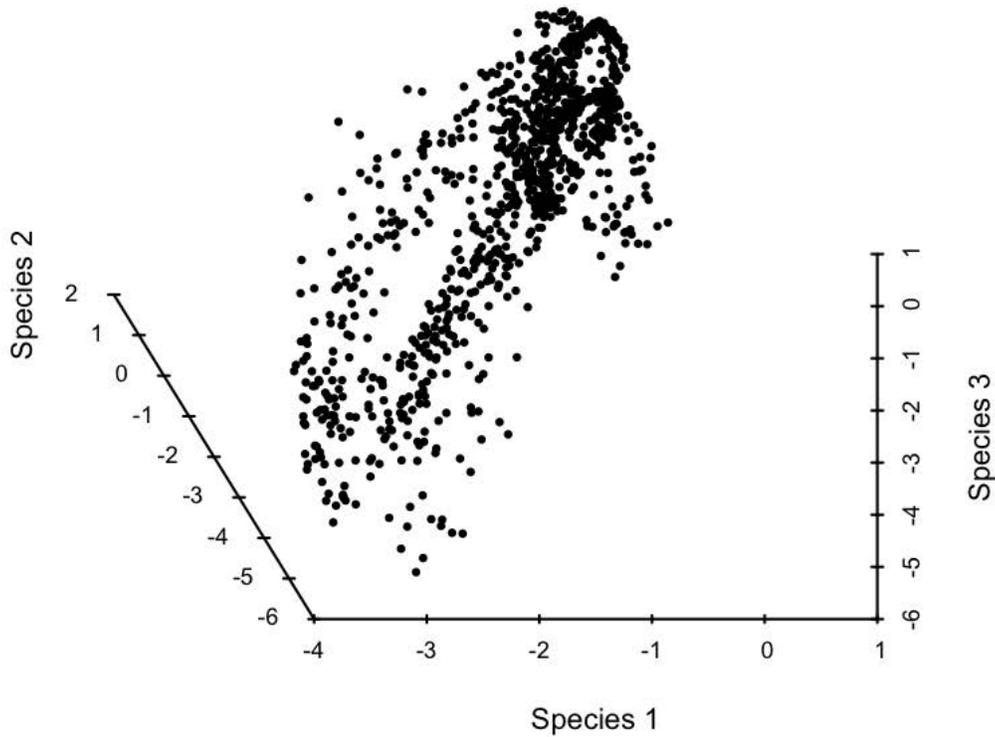
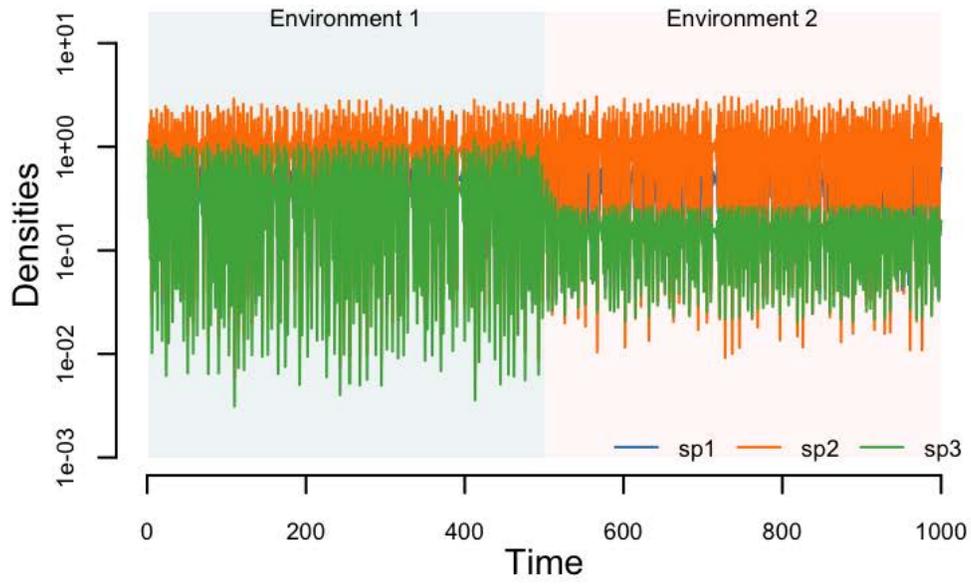


Fig. S10. Changing environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.005$ .

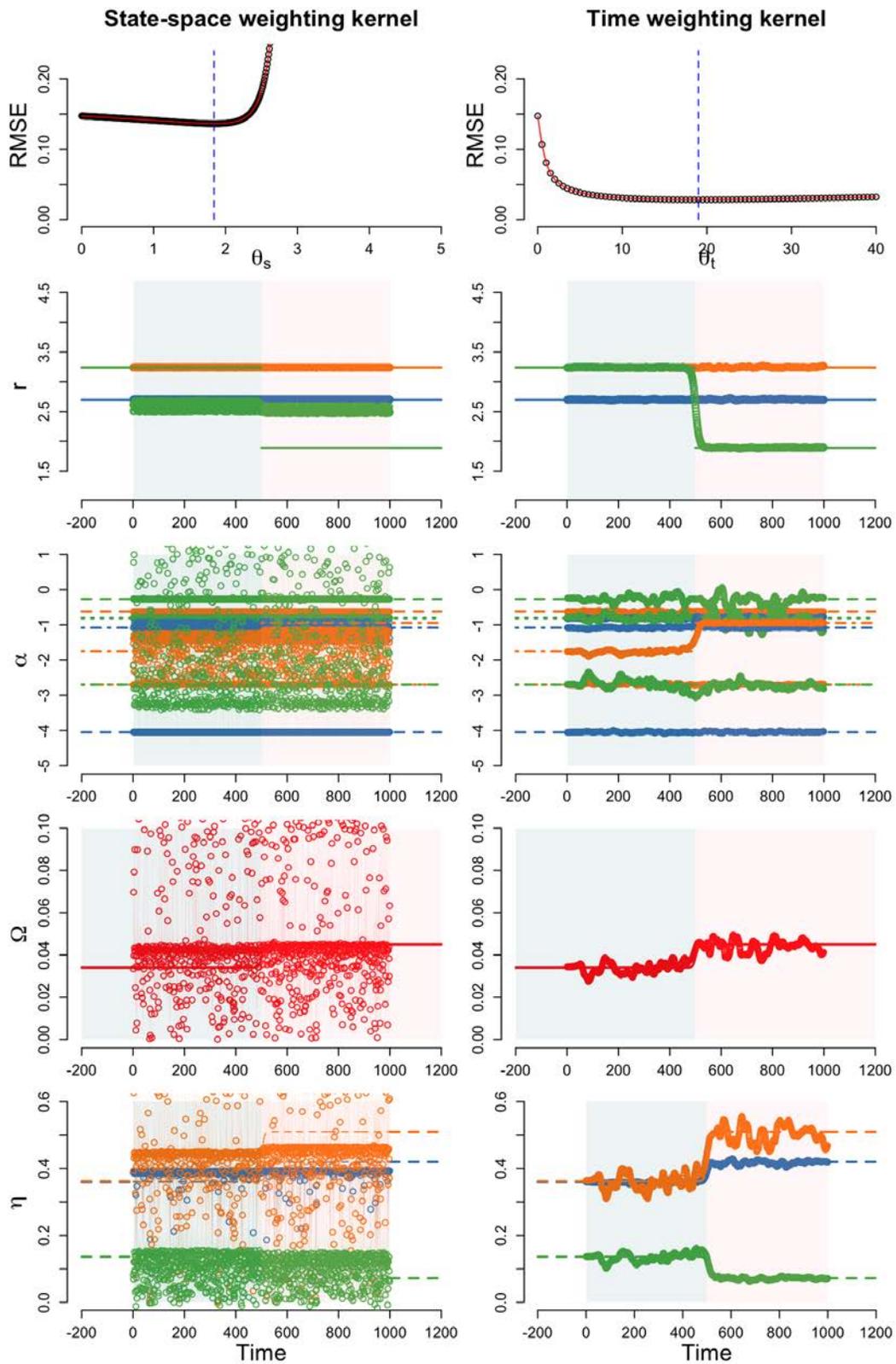


Fig. S11. Changing environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.005$ .

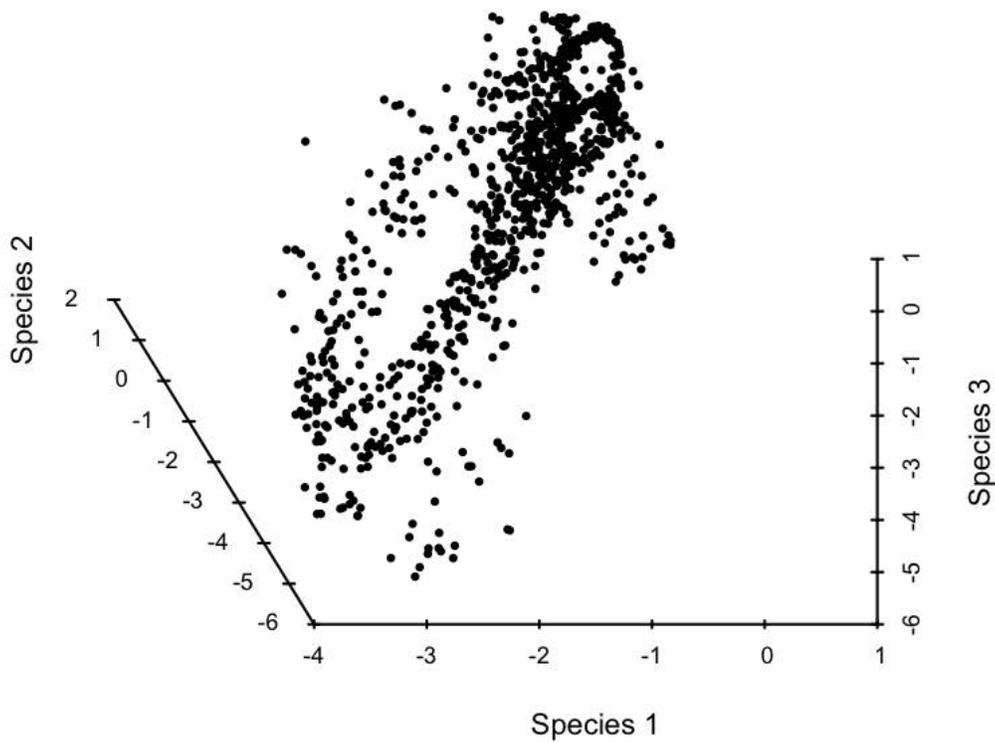
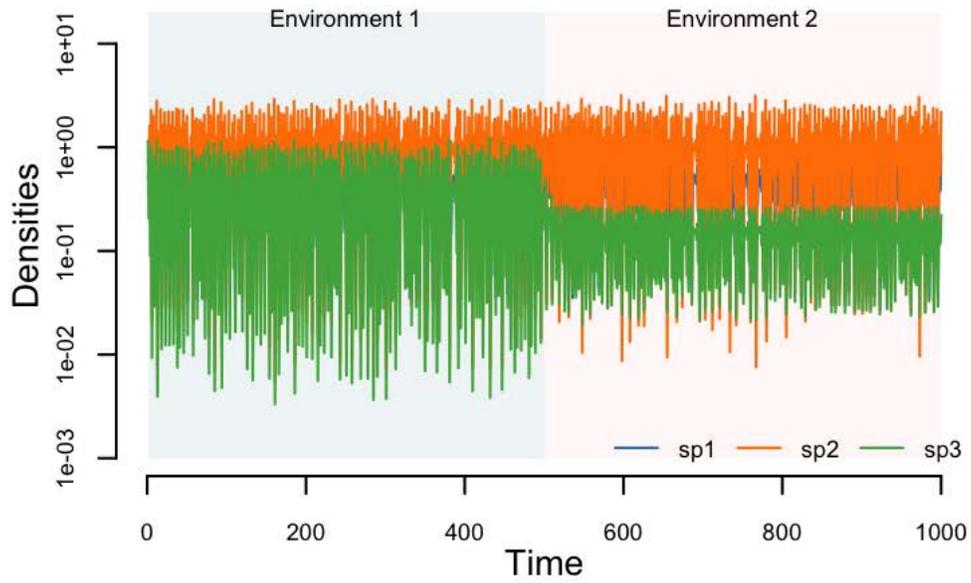


Fig. S12. Changing environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.007$ .

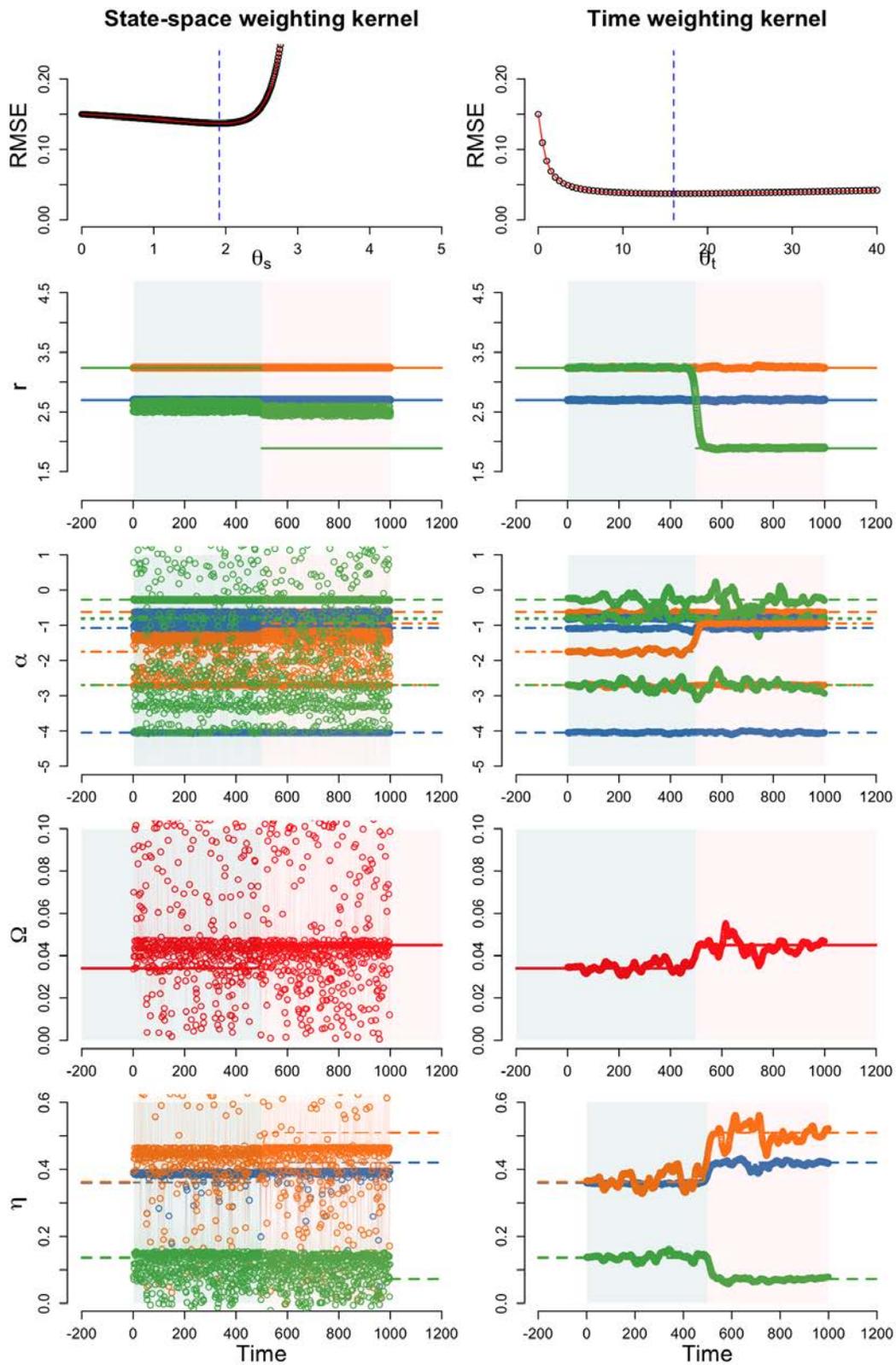


Fig. S13. Changing environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.007$ .

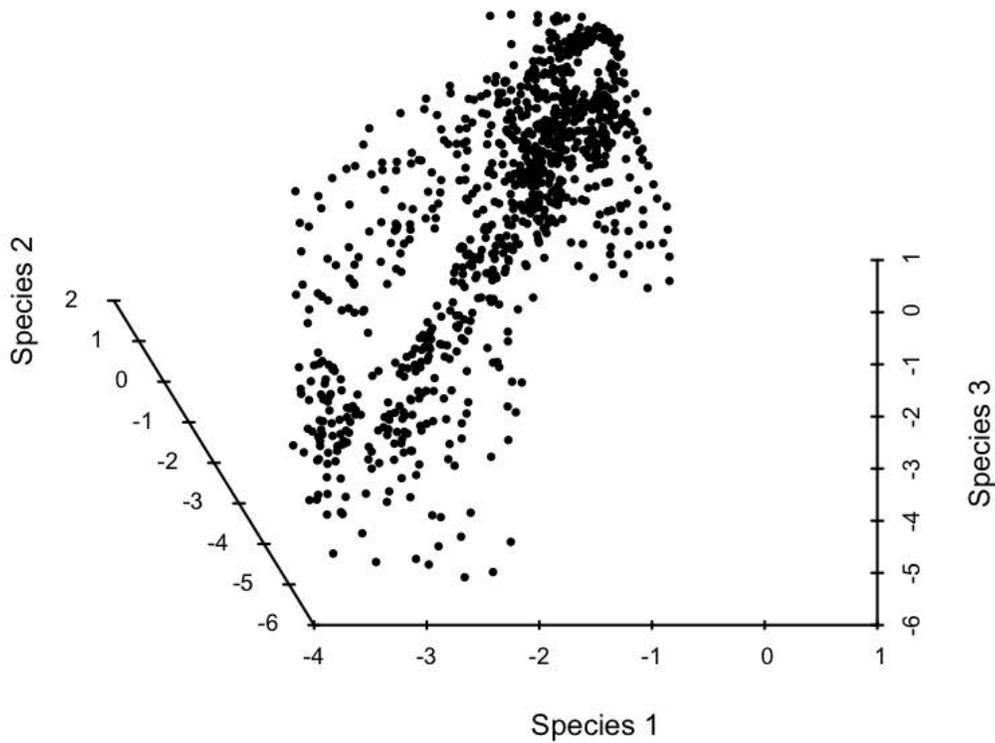
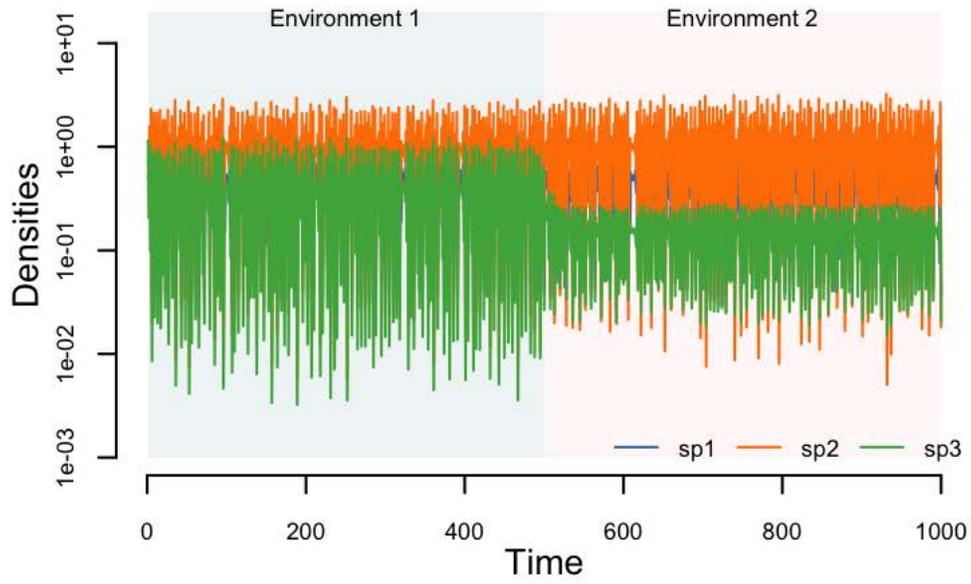


Fig. S14. Changing environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.01$ .

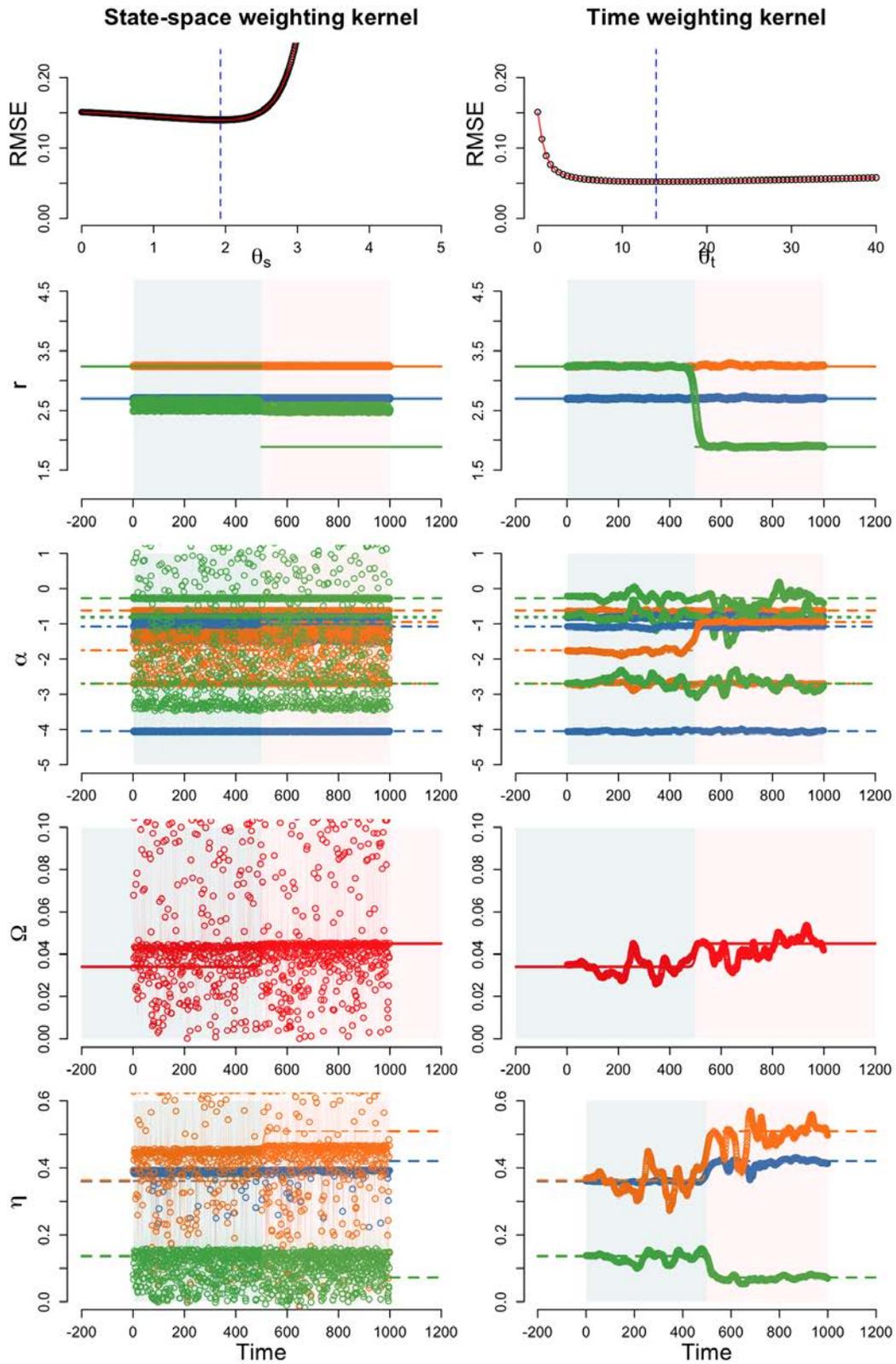


Fig. S15. Changing environment and chaotic dynamics, with small environmental noise  $\epsilon = 0.01$ .

## 6. Supporting Figures for Synthetic Data under Changing Environment and Cyclic Dynamics

For the cyclic dynamics, we used the following parameters in environment 1 and environment 2, respectively:

$$\mathbf{r}_1 = \begin{bmatrix} 3.8 \\ 0.3 \\ 2.4 \end{bmatrix} \quad \mathbf{r}_2 = \begin{bmatrix} 3.8 \\ 0.3 \\ 2.4 \end{bmatrix}$$

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} -0.2 & -0.4 & -0.8 \\ -0.0011 & -0.033 & -0.067 \\ -0.2 & -0.2 & -0.6 \end{bmatrix} \quad \boldsymbol{\alpha}_2 = \begin{bmatrix} -0.2 & -0.4 & -0.8 \\ -0.011 & -0.033 & -0.067 \\ -0.2 & -0.2 & -0.6 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.005 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.007 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ .

For each level of environmental noise, two figures are provided.

The first figure shows the population dynamics as function of time and in the state-space.

The second figure shows:

- Left column: results with the state-space weighing kernel.
- Right column: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

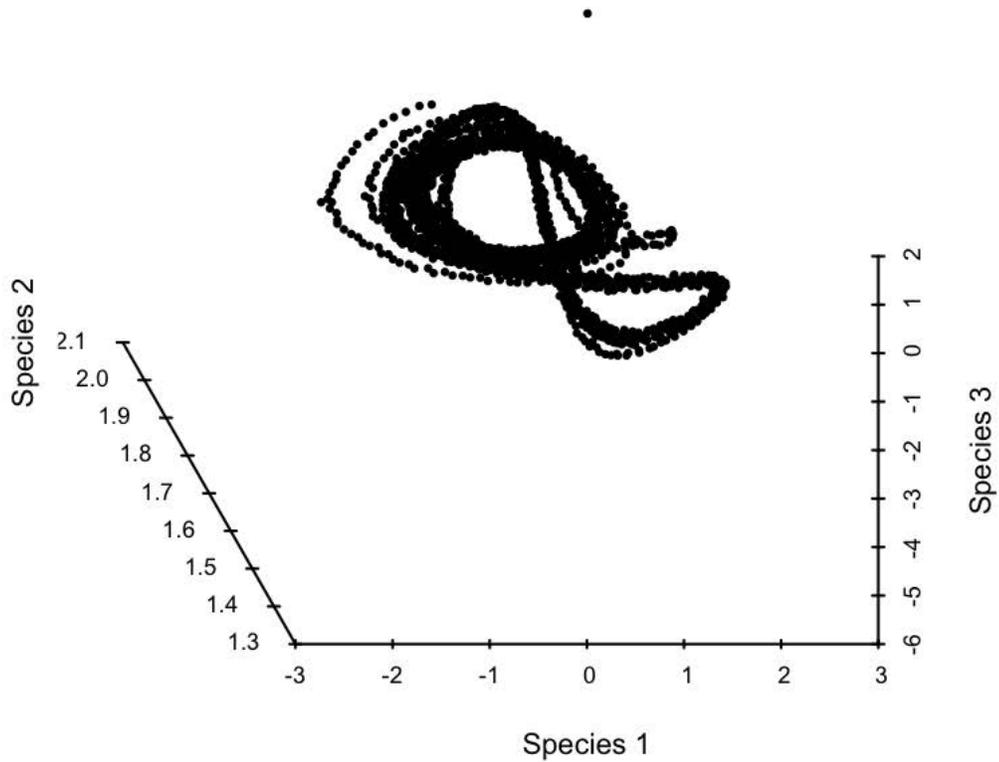
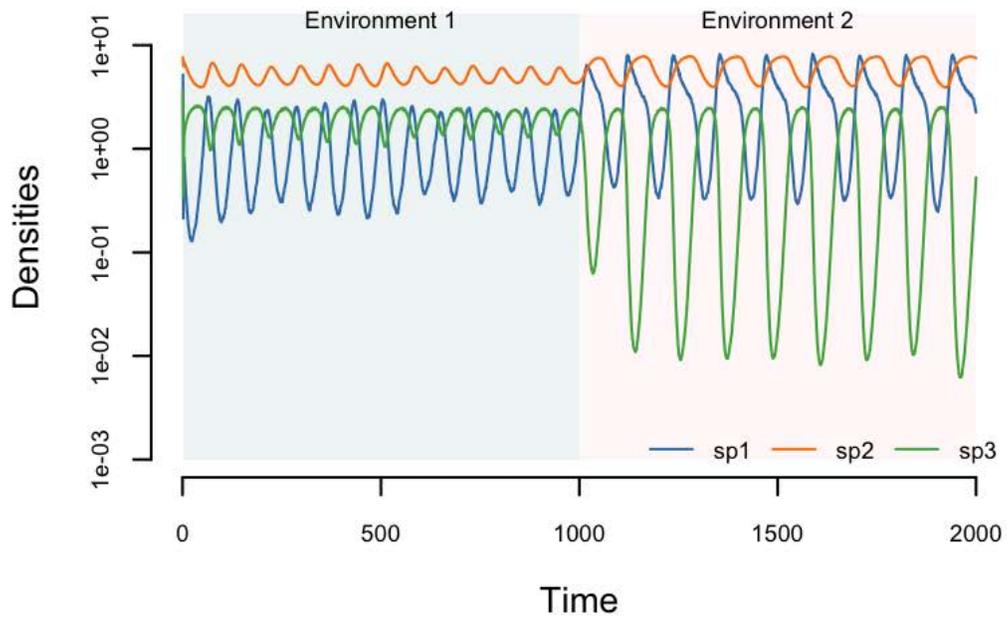


Fig. S16. Changing environment and cycle dynamics, with small environmental noise  $\epsilon = 0.005$ .

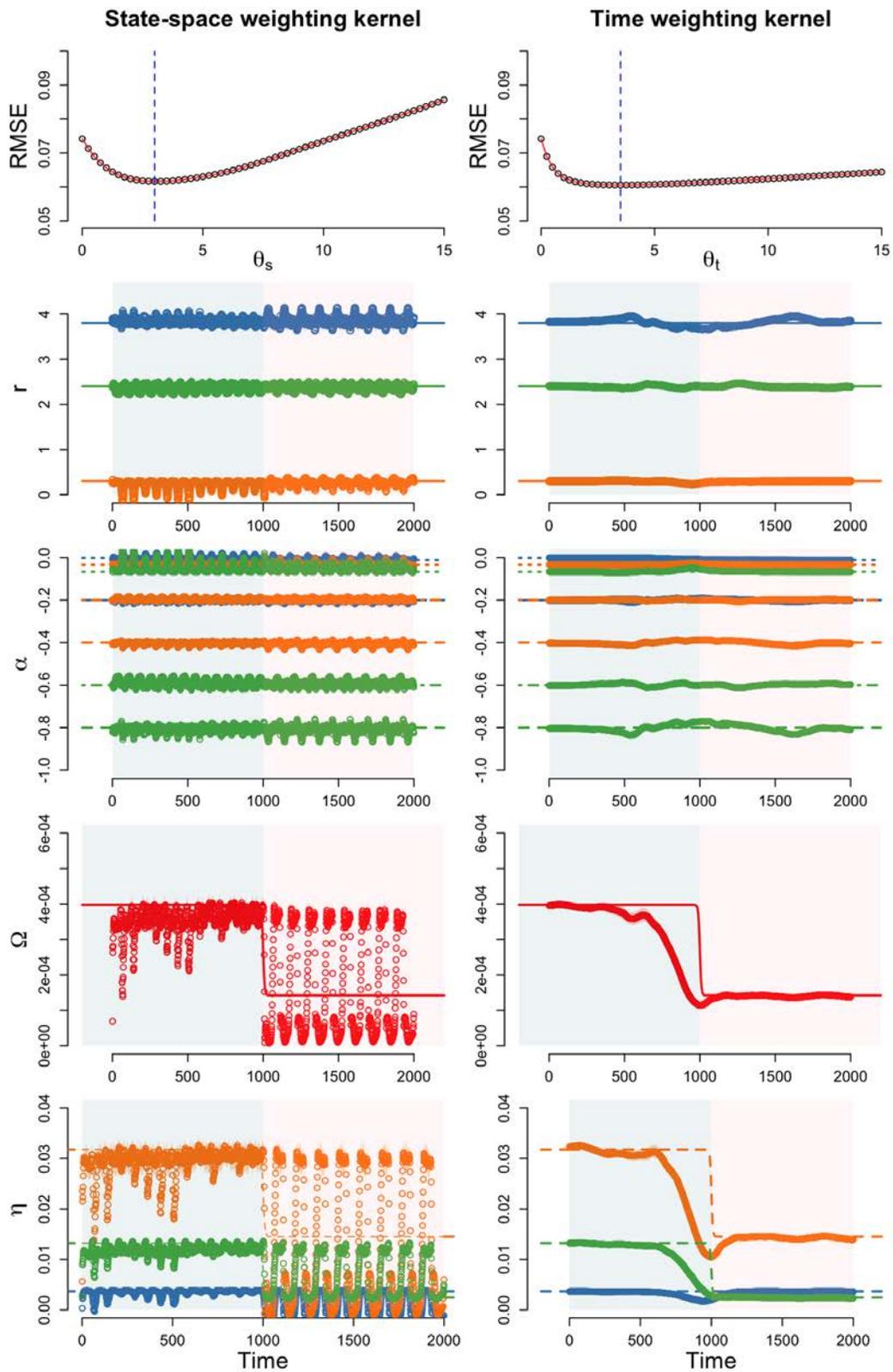


Fig. S17. Changing environment and cycle dynamics, with small environmental noise  $\epsilon = 0.005$ .

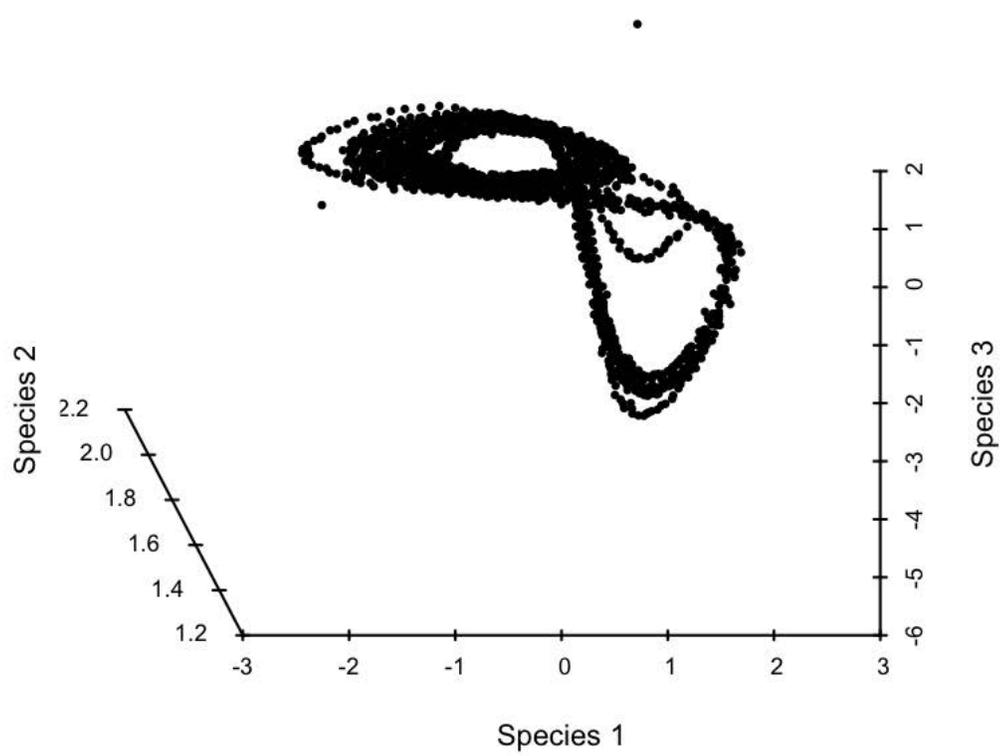
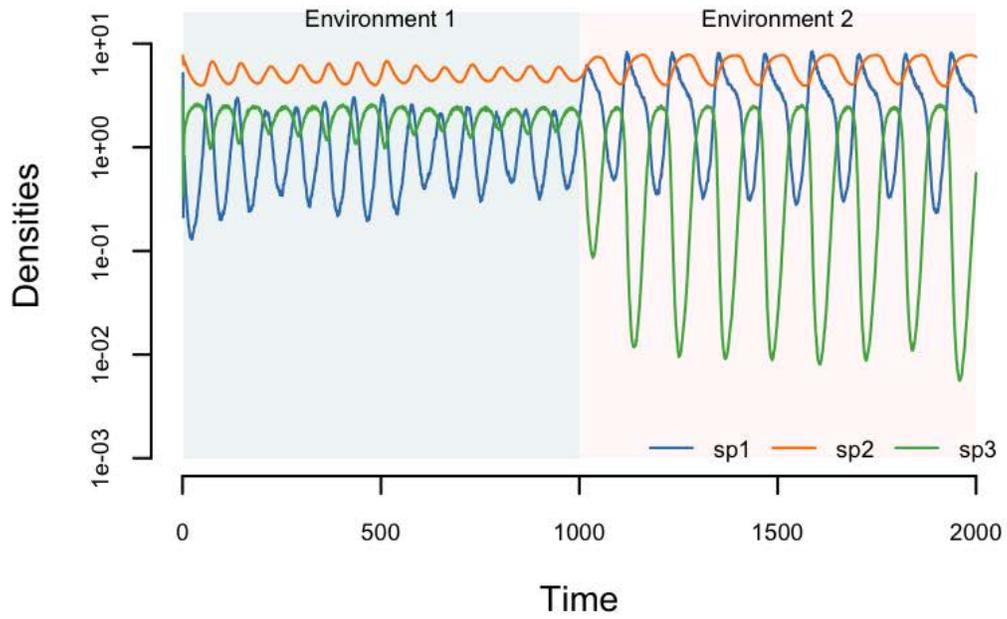


Fig. S18. Changing environment and cycle dynamics, with small environmental noise  $\epsilon = 0.007$ .

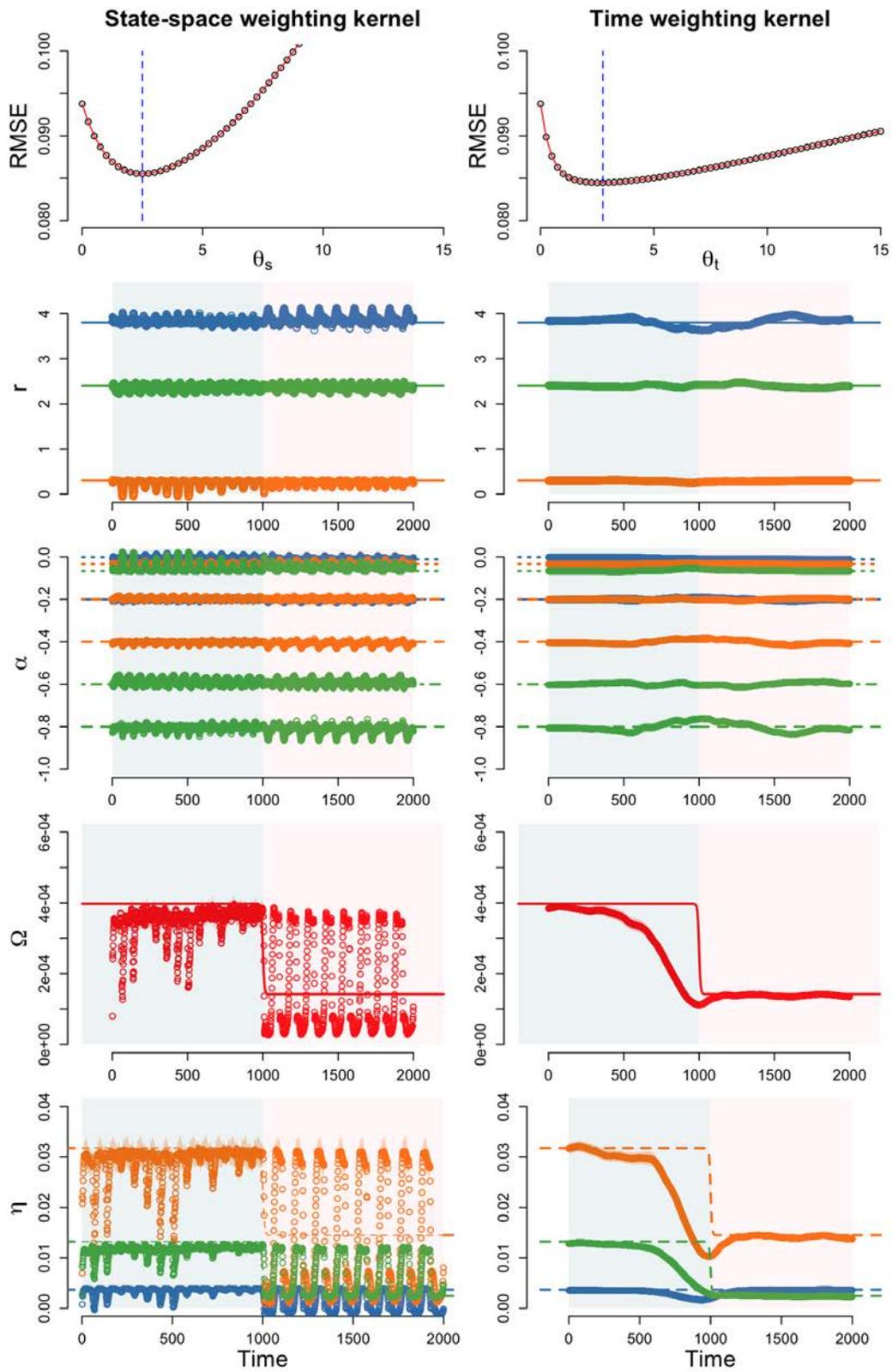


Fig. S19. Changing environment and cycle dynamics, with small environmental noise  $\epsilon = 0.007$ .

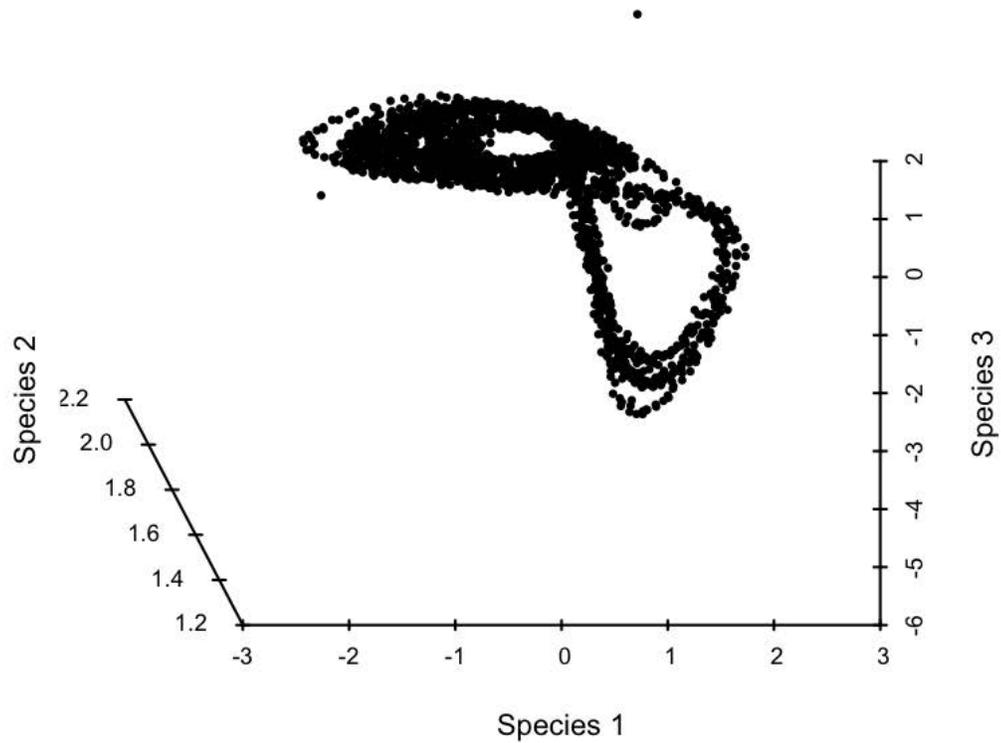
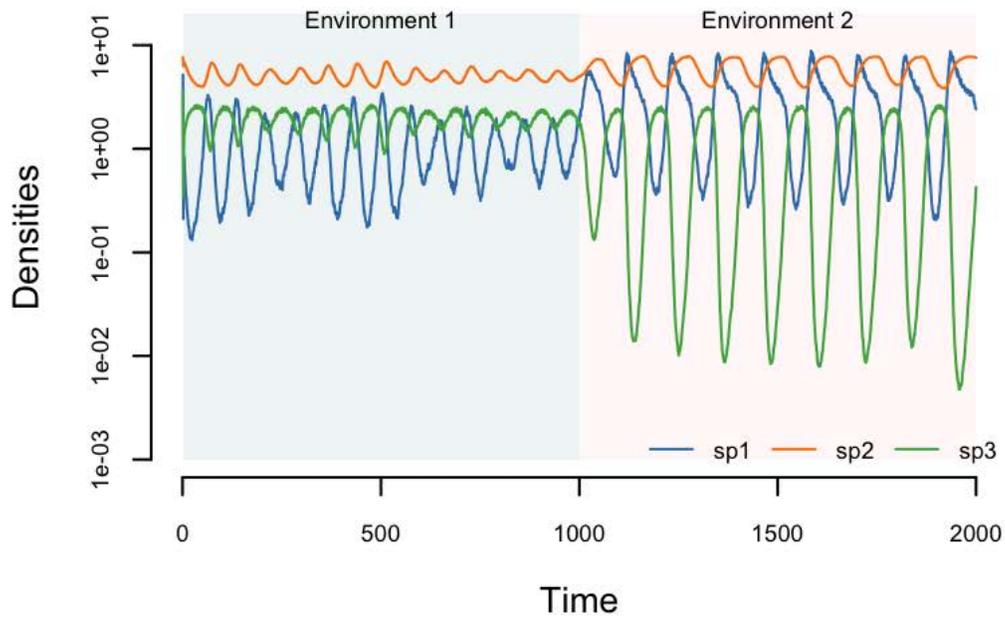


Fig. S20. Changing environment and cycle dynamics, with small environmental noise  $\epsilon = 0.01$ .

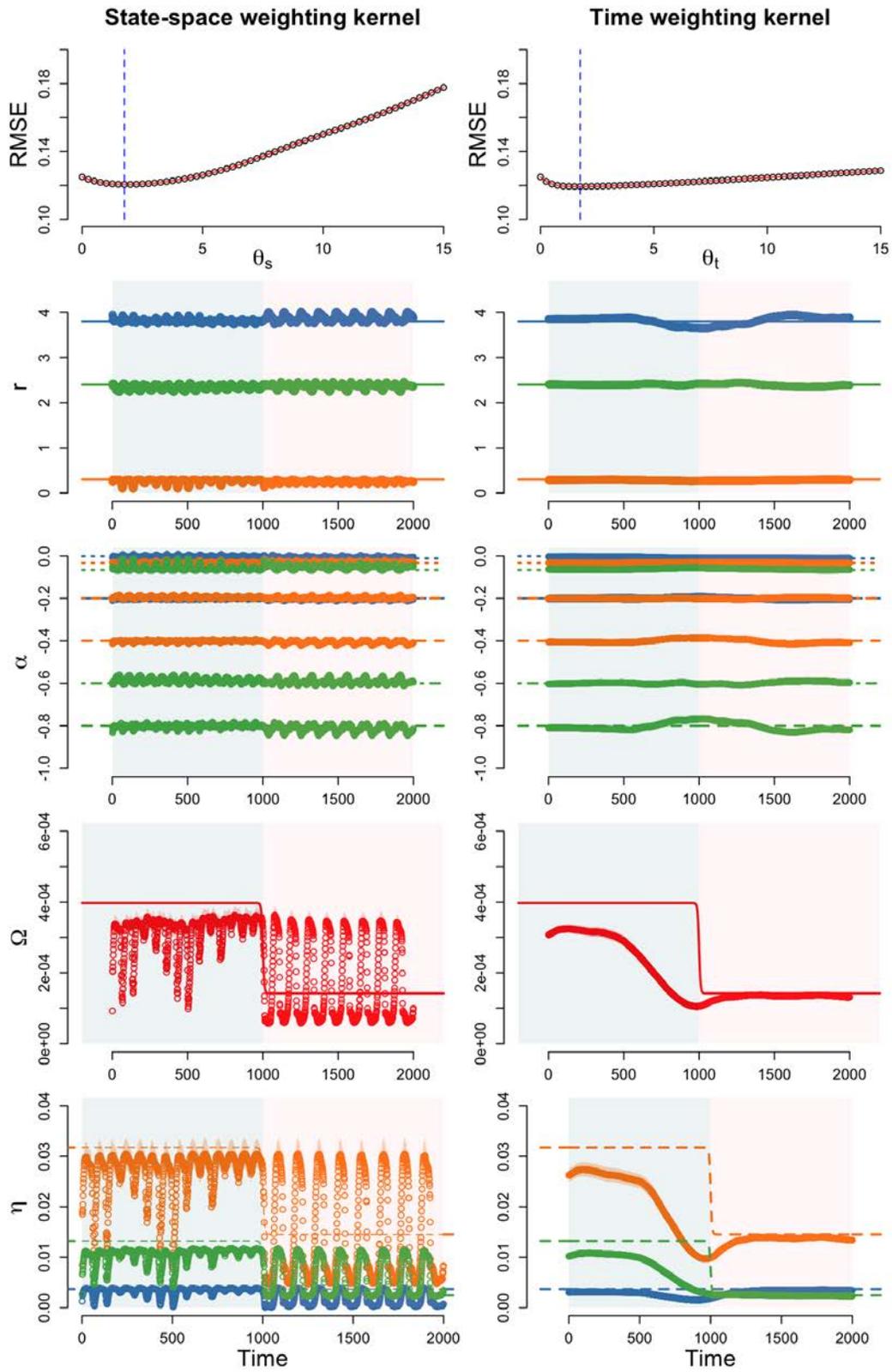


Fig. S21. Changing environment and cycle dynamics, with small environmental noise  $\epsilon = 0.01$ .

## 7. Supporting Figures for Synthetic Data under Changing Environment and Fixed Point Dynamics

For the fixed point dynamics, we used the following parameters in environment 1 and environment 2, respectively:

$$\mathbf{r}_1 = \begin{bmatrix} 0.7 \\ 0.84 \\ 1.19 \end{bmatrix} \quad \mathbf{r}_2 = \begin{bmatrix} 0.7 \\ 0.84 \\ 0.63 \end{bmatrix}$$
$$\boldsymbol{\alpha}_1 = \begin{bmatrix} -1.05 & -0.161 & -0.07 \\ -0.21 & -0.7 & -0.21 \\ -0.28 & -0.455 & -0.7 \end{bmatrix} \quad \boldsymbol{\alpha}_2 = \begin{bmatrix} -1.05 & -0.161 & -0.07 \\ -0.21 & -0.7 & -0.21 \\ -0.28 & -0.245 & -0.7 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.004 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.007 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ .

For each level of environmental noise, two figures are provided.

The first figure shows the population dynamics as function of time and in the state-space.

The second figure shows:

- Left column: results with the state-space weighing kernel.
- Right column: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

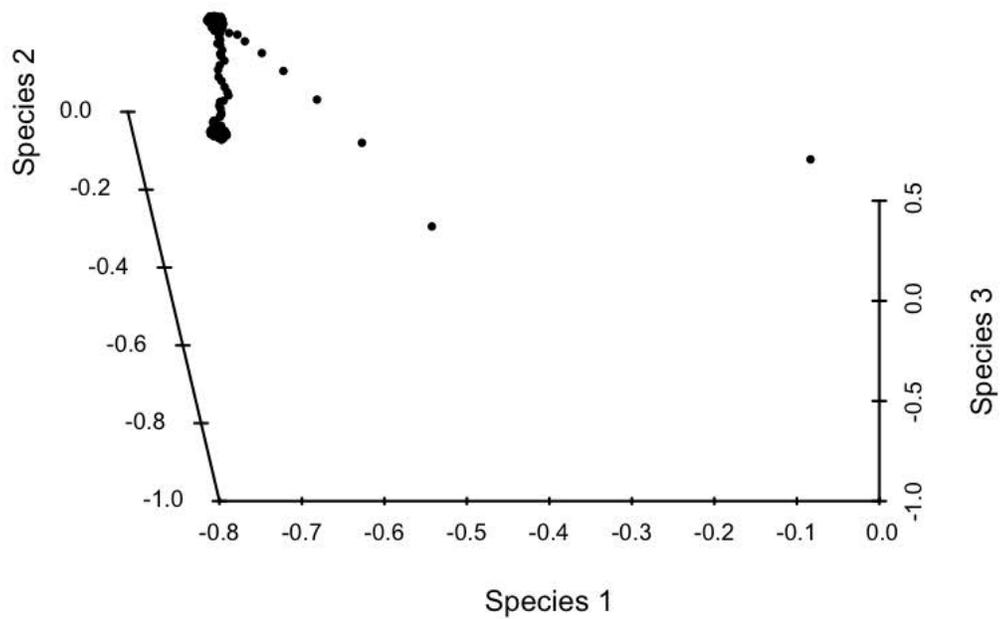
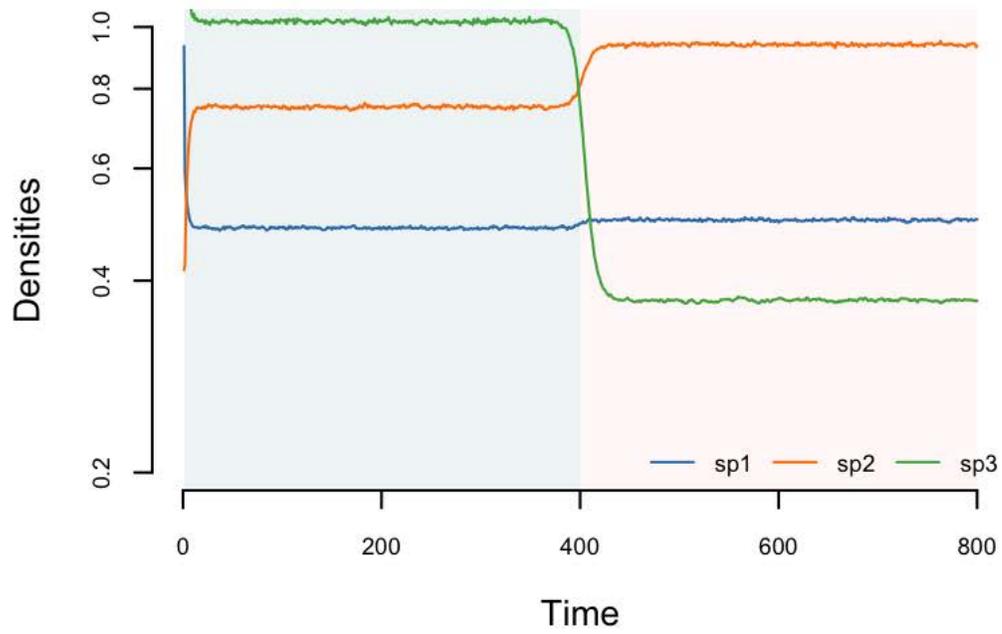


Fig. S22. Changing environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.004$ .

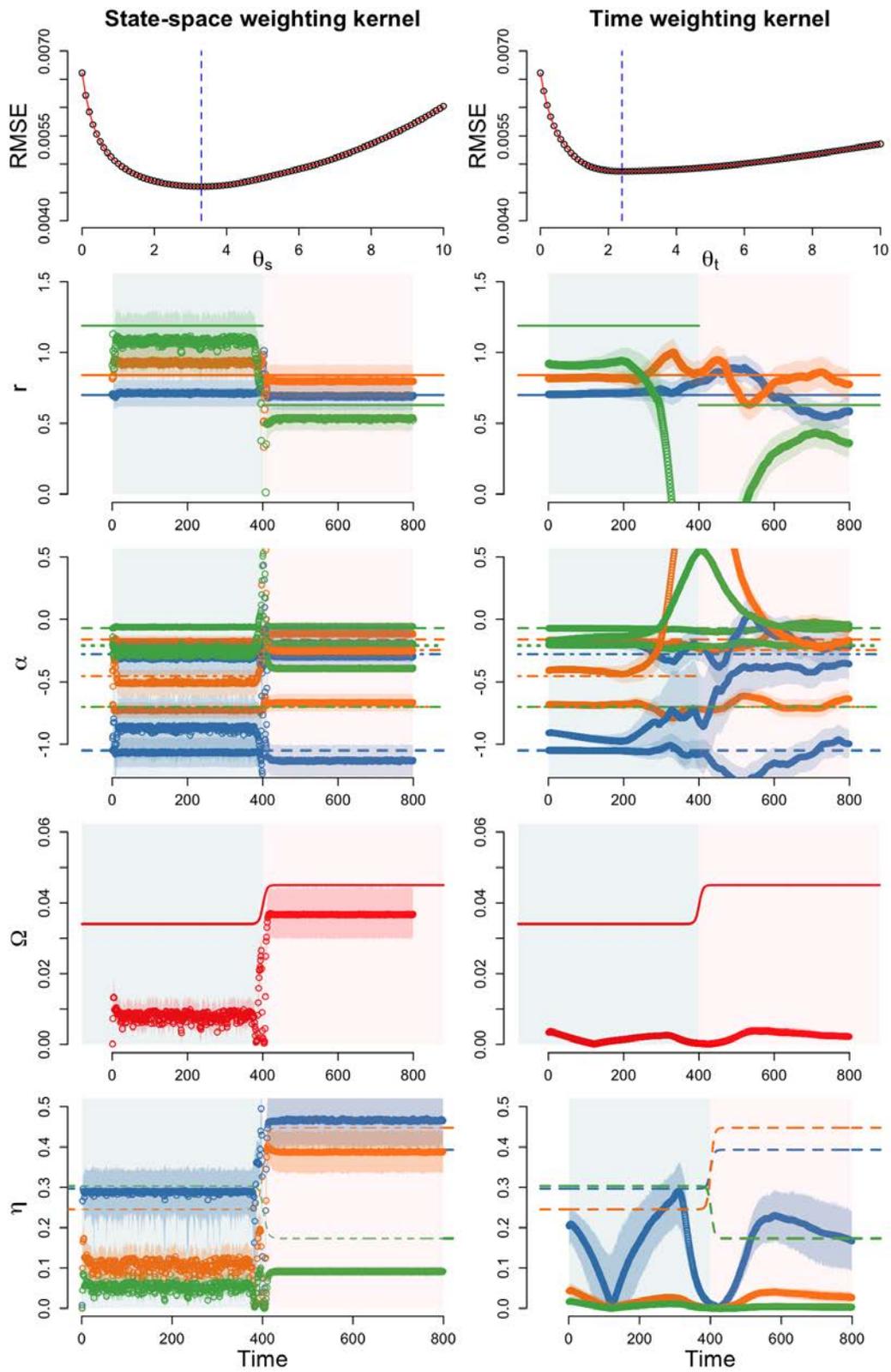


Fig. S23. Changing environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.004$ .

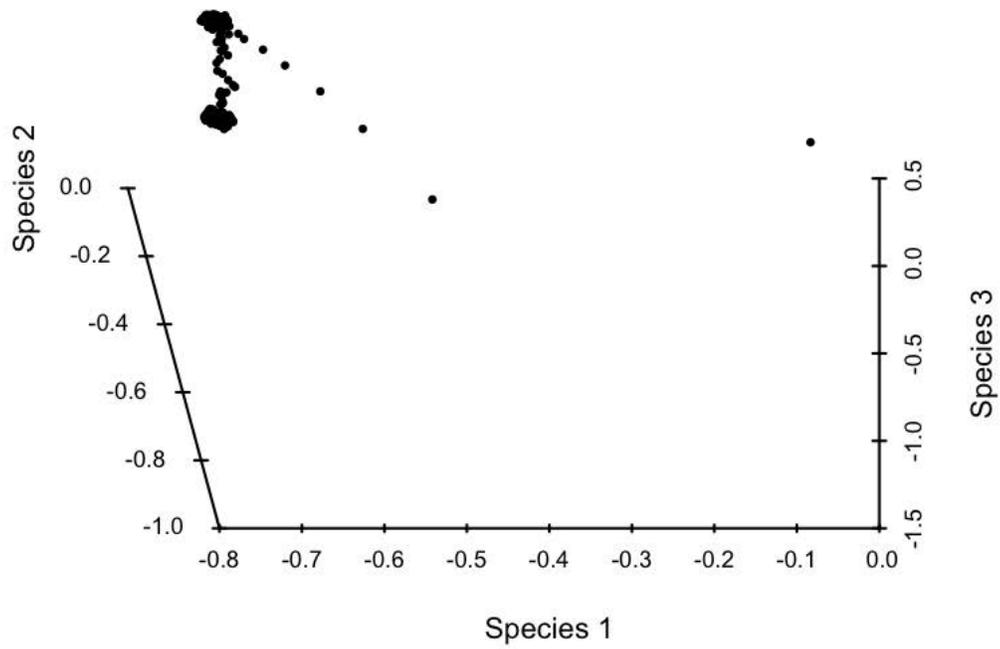
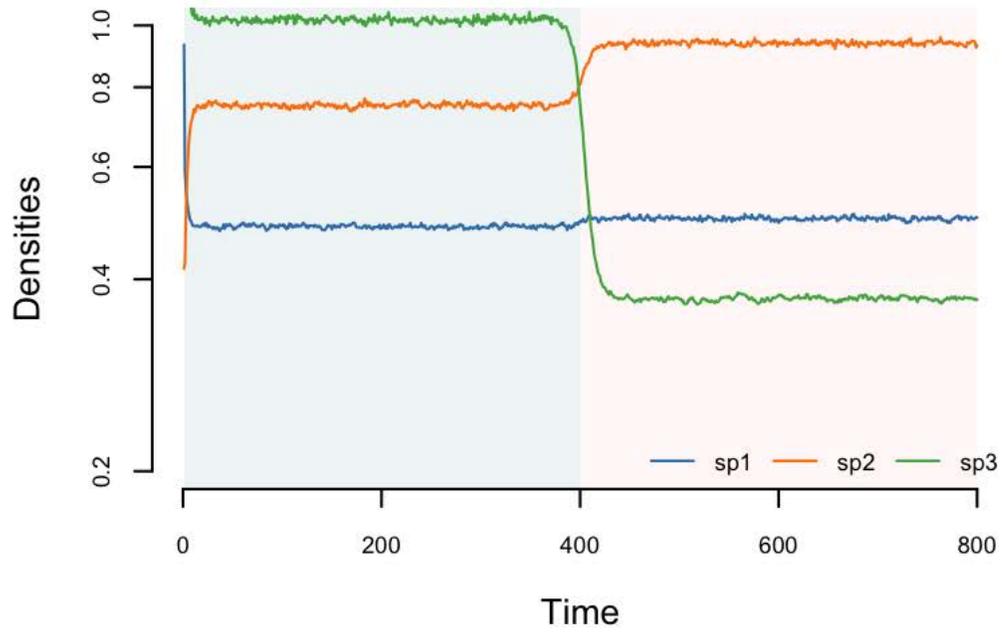


Fig. S24. Changing environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.007$ .

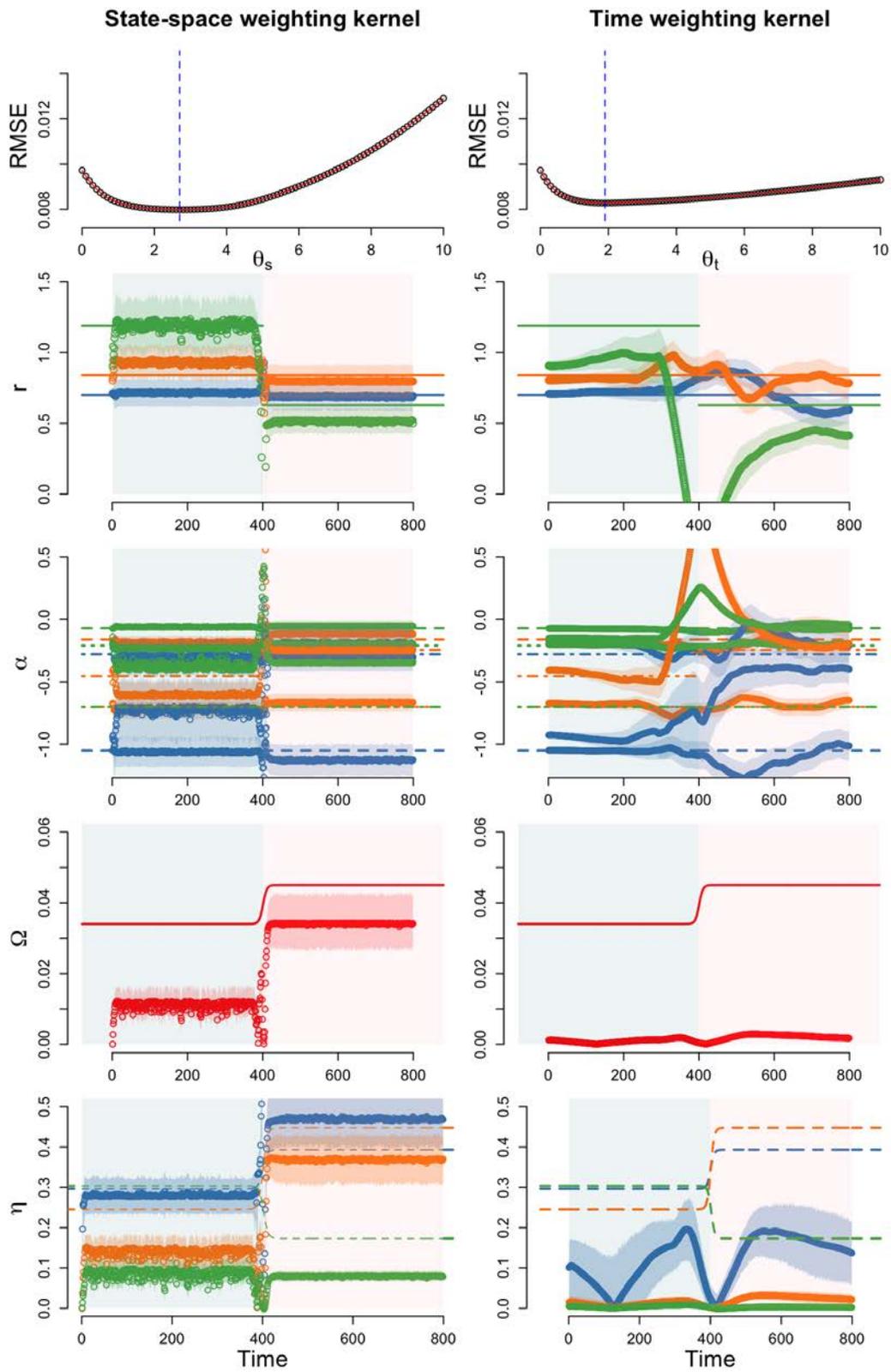


Fig. S25. Changing environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.007$ .

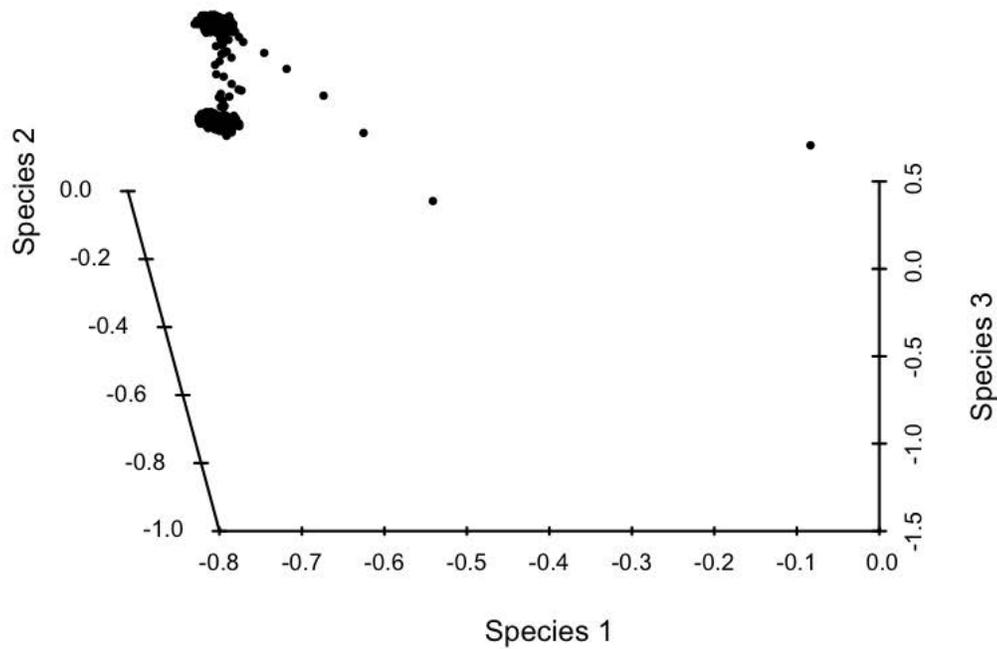
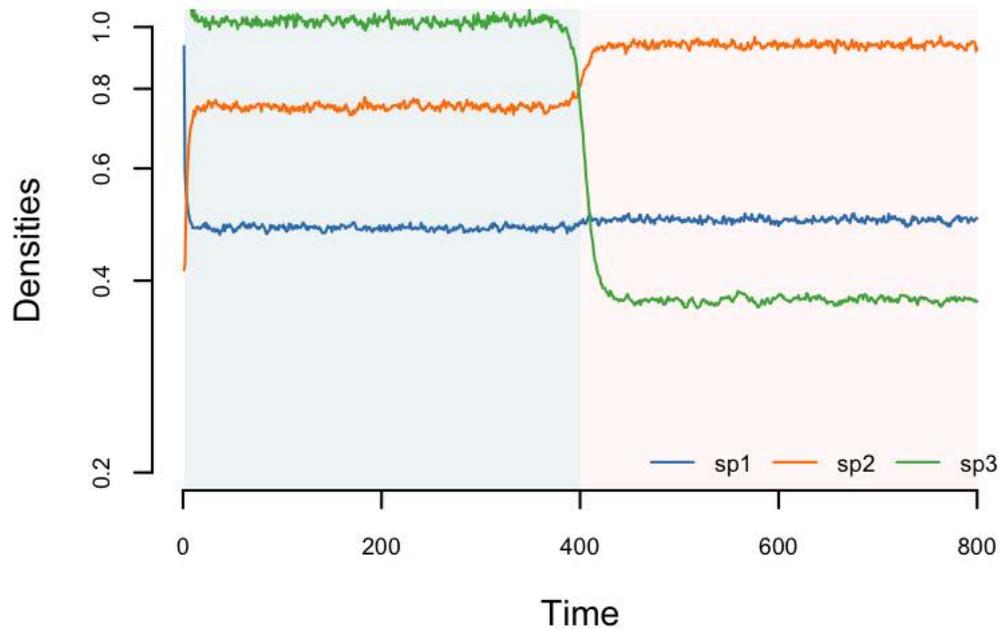


Fig. S26. Changing environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.01$ .

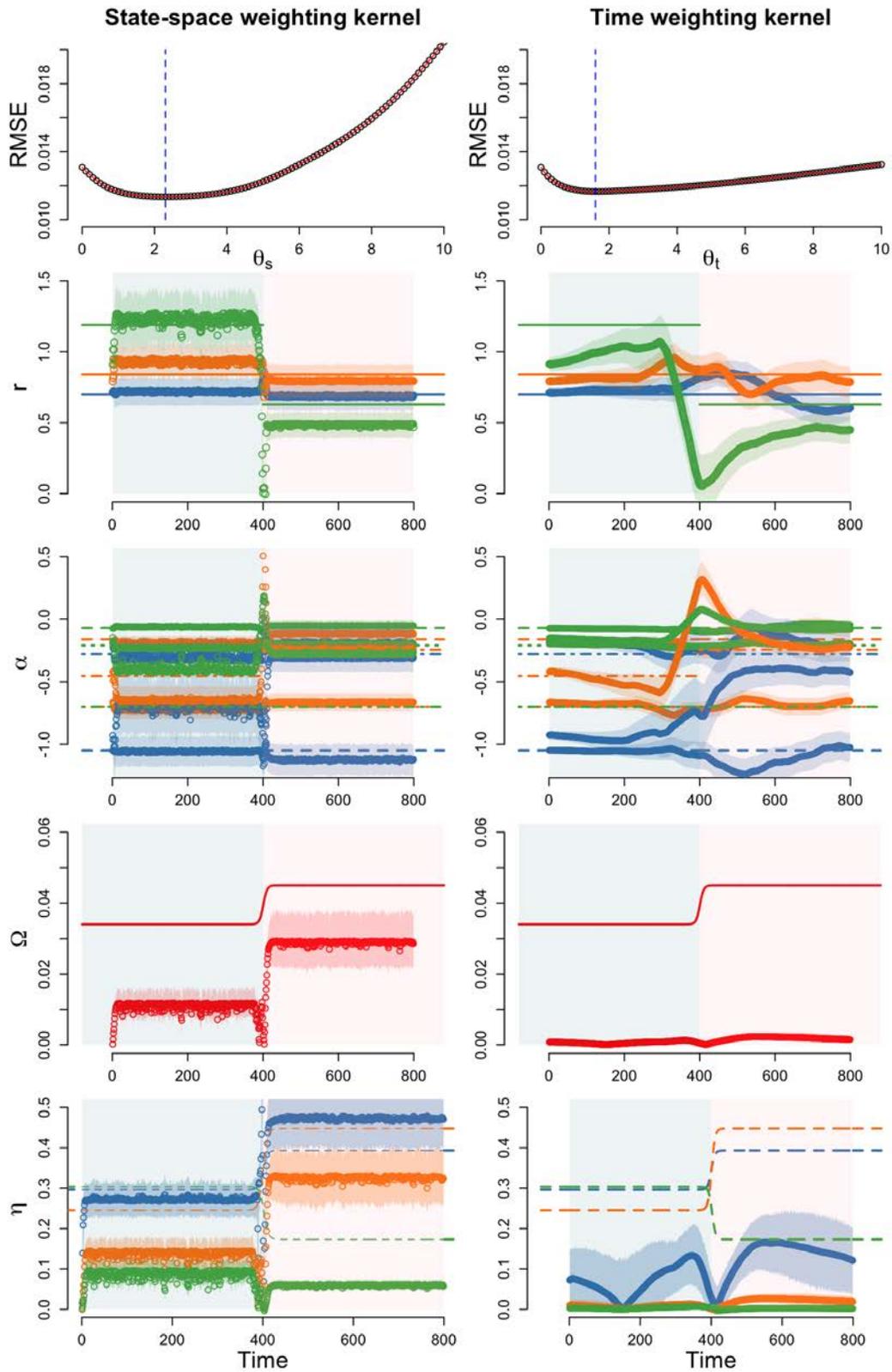


Fig. S27. Changing environment and fixed point dynamics, with small environmental noise  $\epsilon = 0.01$ .

## 8. Supporting Figures for Synthetic Data with Extinction

For the dynamics with an extinction, we used the following parameters in environment 1 and environment 2, respectively:

$$\mathbf{r}_1 = \begin{bmatrix} 2.7 \\ 3.24 \\ 4.05 \end{bmatrix} \quad \mathbf{r}_2 = \begin{bmatrix} 2.7 \\ 3.24 \\ 1.512 \end{bmatrix}$$
$$\boldsymbol{\alpha}_1 = \begin{bmatrix} -4.05 & -0.621 & -0.27 \\ -0.81 & -2.7 & -0.81 \\ -1.08 & -1.755 & -2.70 \end{bmatrix} \quad \boldsymbol{\alpha}_2 = \begin{bmatrix} -4.05 & -0.621 & -0.27 \\ -0.81 & -2.7 & -0.81 \\ -1.08 & -0.945 & -2.70 \end{bmatrix}.$$

We performed the simulation under three levels of environmental noise: (i) small  $\epsilon_i \sim \mathcal{N}(0, 0.005 \cdot r_i^2)$ , (ii) intermediate  $\epsilon_i \sim \mathcal{N}(0, 0.007 \cdot r_i^2)$ , and (iii) large  $\epsilon_i \sim \mathcal{N}(0, 0.01 \cdot r_i^2)$ .

For each level of environmental noise, two figures are provided.

The first figure shows the population dynamics as function of time and in the state-space.

The second figure shows:

- Left column: results with the state-space weighing kernel.
- Right column: results with the time weighing kernel.
- Second row: cross-validation results. The vertical dashed blue line indicates the value of  $\theta$  with the smallest RMSE.
- Third row: intrinsic growth rate.
- Fourth row: *per capita* interaction strength.
- Fifth row: niche difference.
- Last row: resistance angles.
- Thick lines, dashed lines, dotted lines, and dashed dotted lines indicate true values of the parameters. Points indicate inferred parameters. Colored areas indicate the 95 % confidential interval.

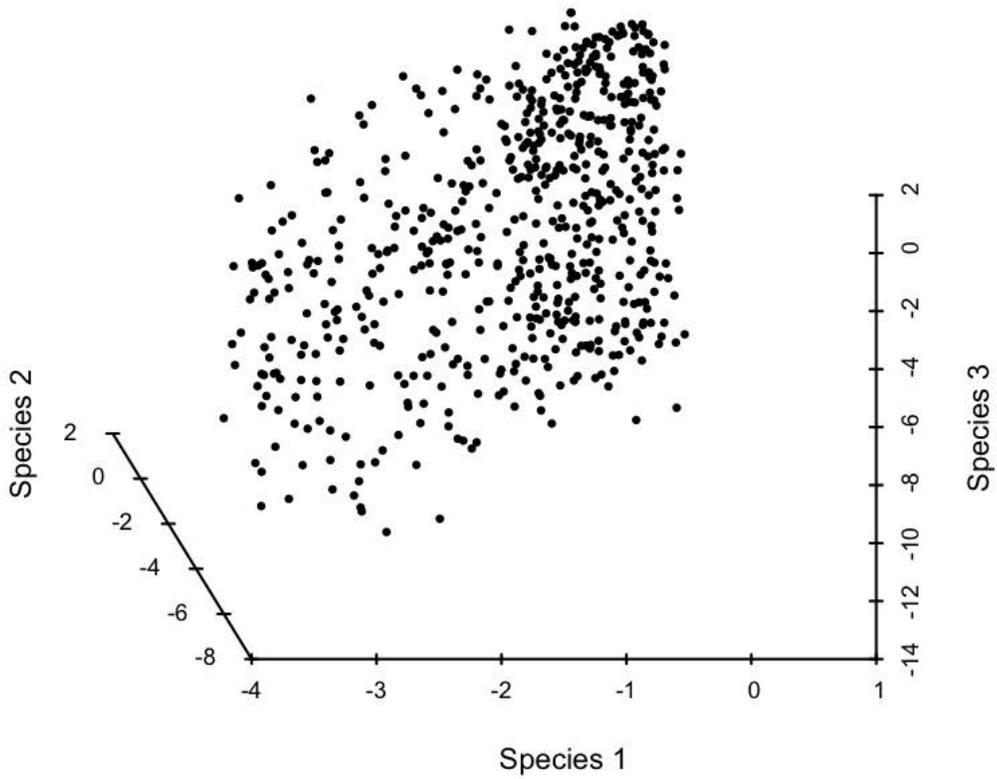
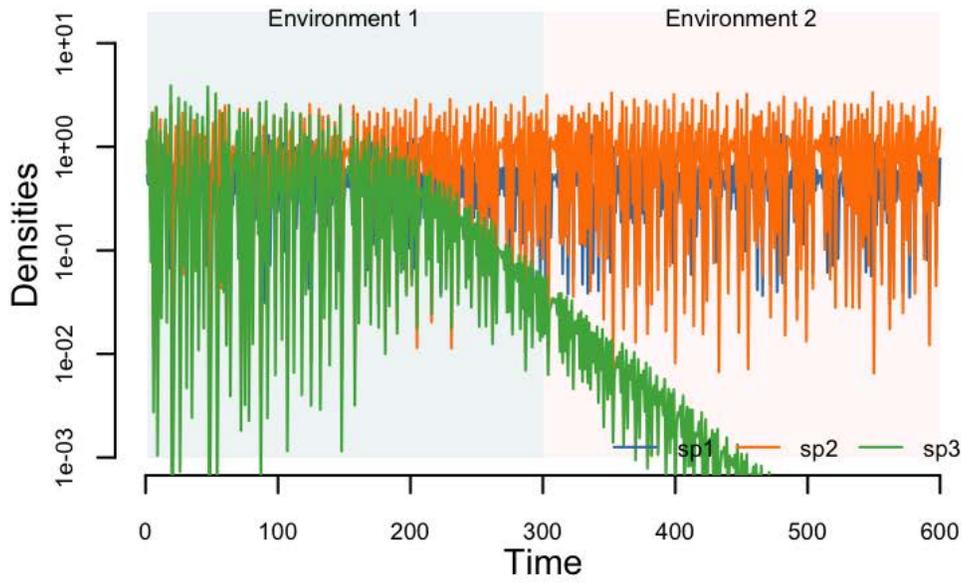


Fig. S28. Dynamics with extinction and small environmental noise  $\epsilon = 0.005$ .

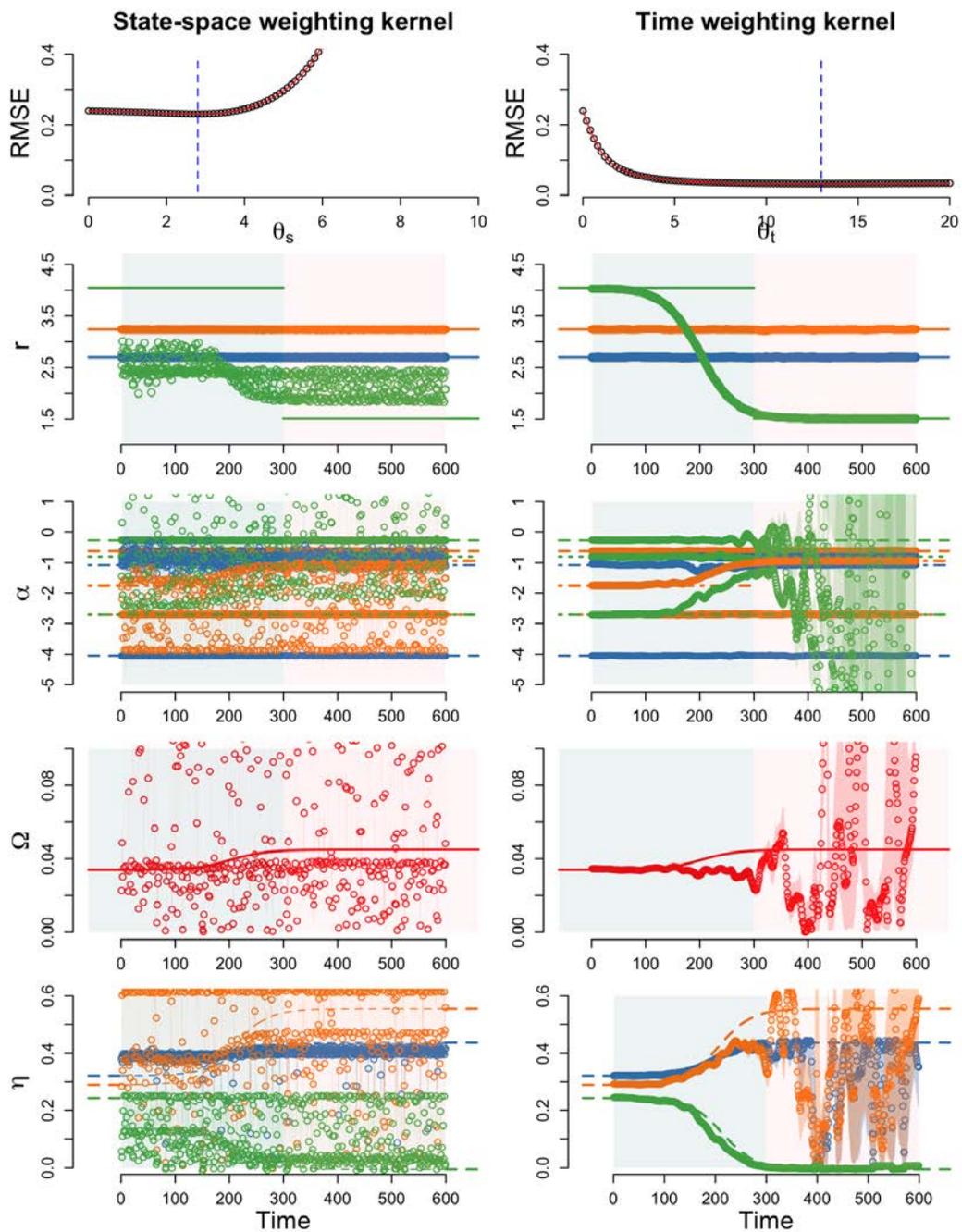


Fig. S29. Dynamics with extinction and small environmental noise  $\epsilon = 0.004$ .

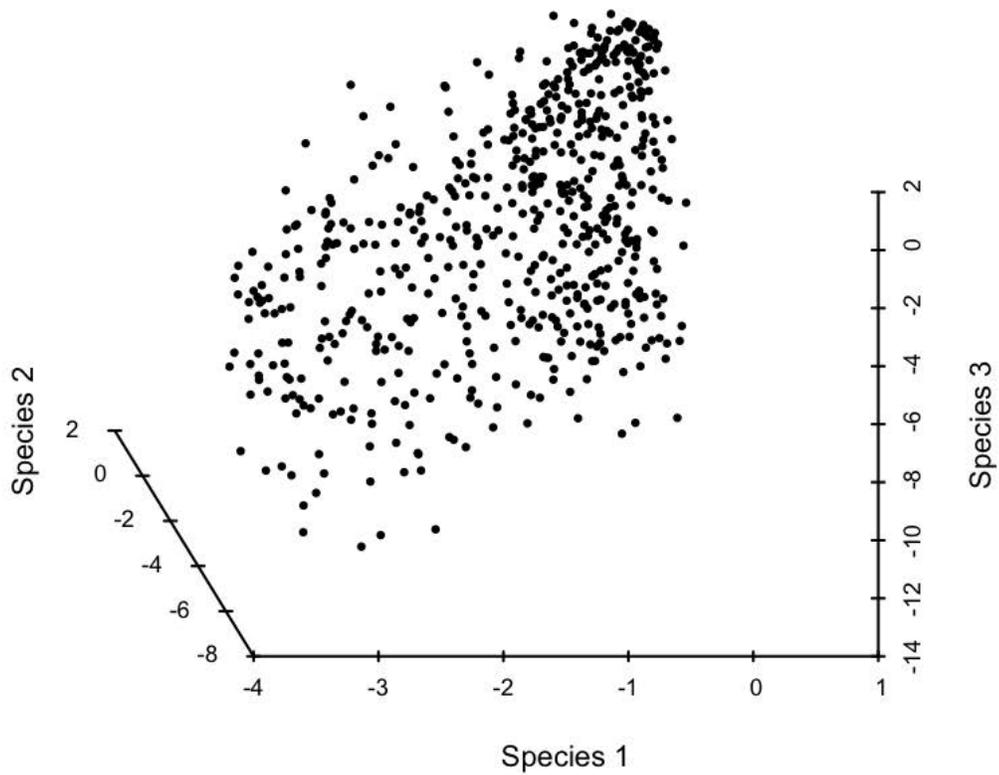
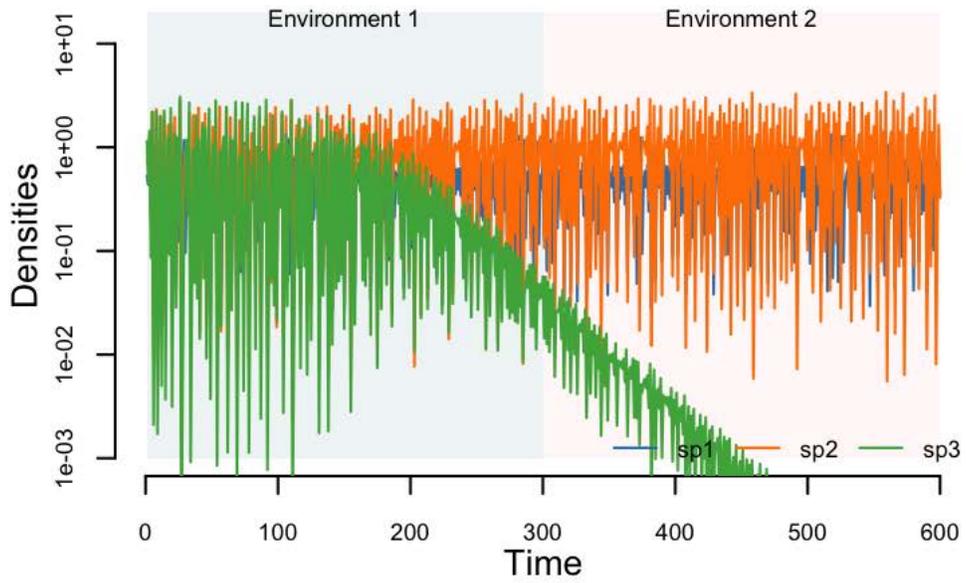


Fig. S30. Dynamics with extinction and small environmental noise  $\epsilon = 0.007$ .

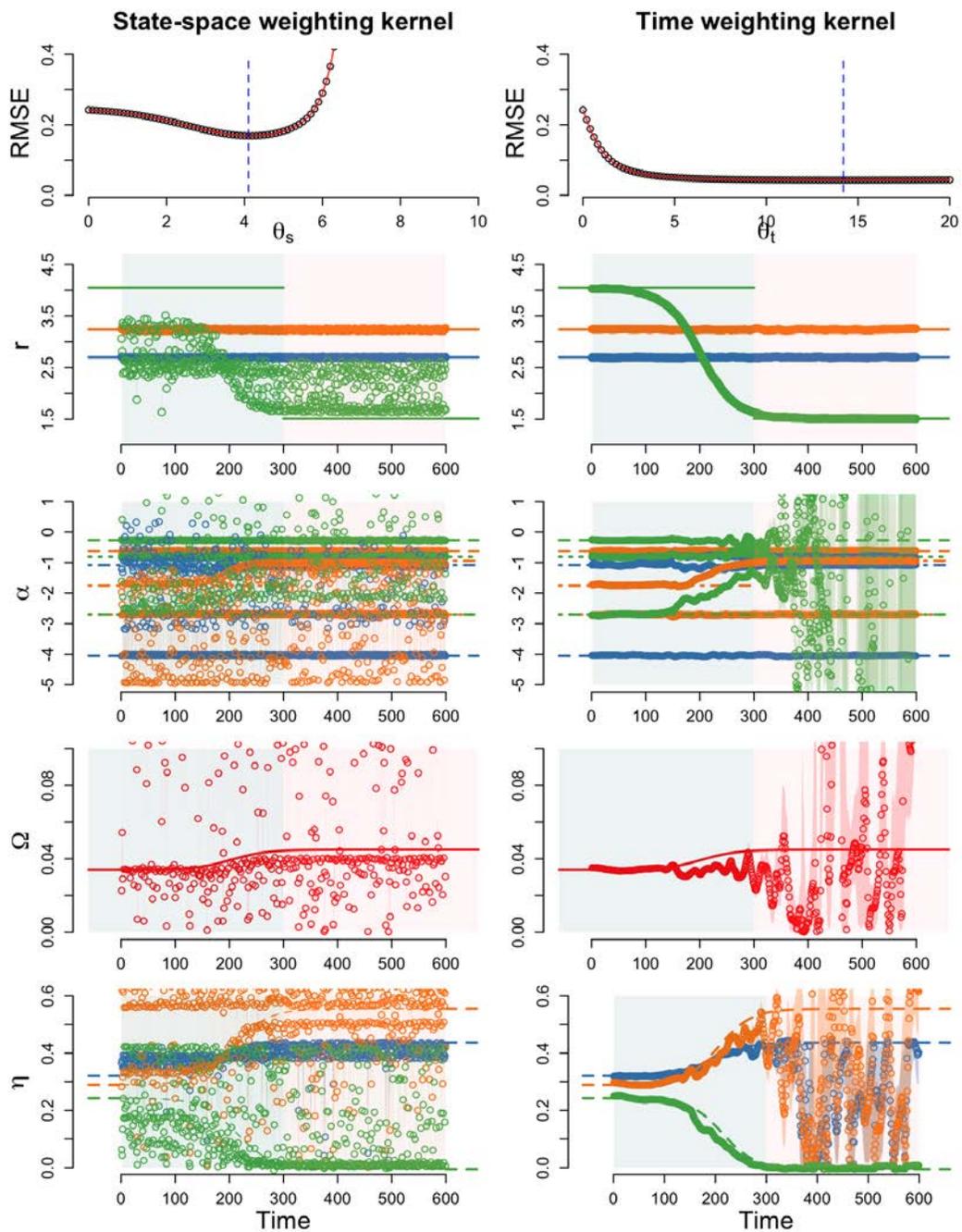


Fig. S31. Dynamics with extinction and small environmental noise  $\epsilon = 0.007$ .

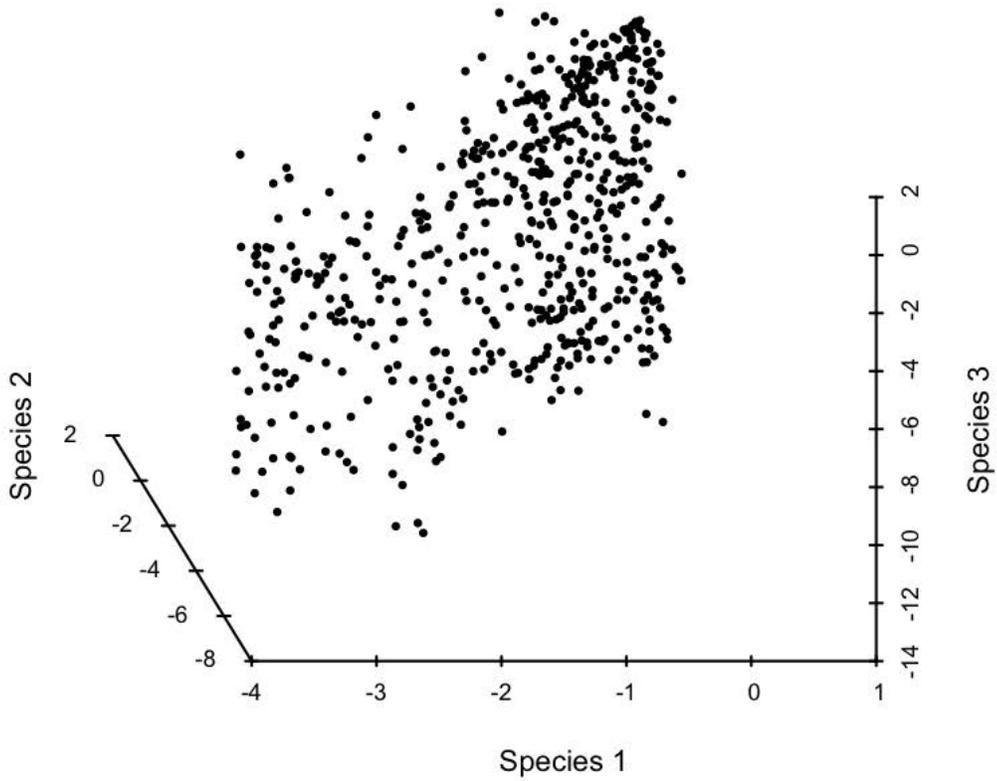
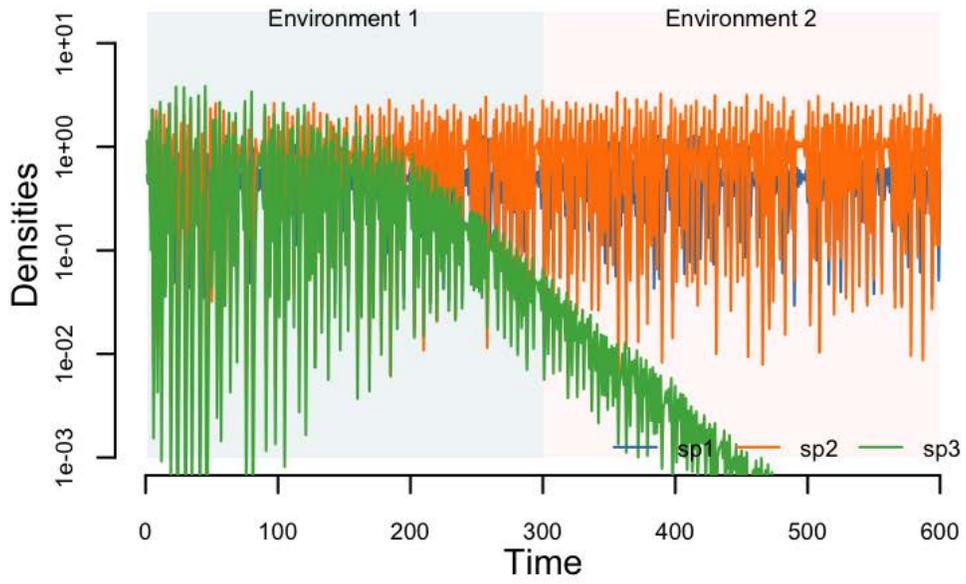


Fig. S32. Dynamics with extinction and small environmental noise  $\epsilon = 0.01$ .

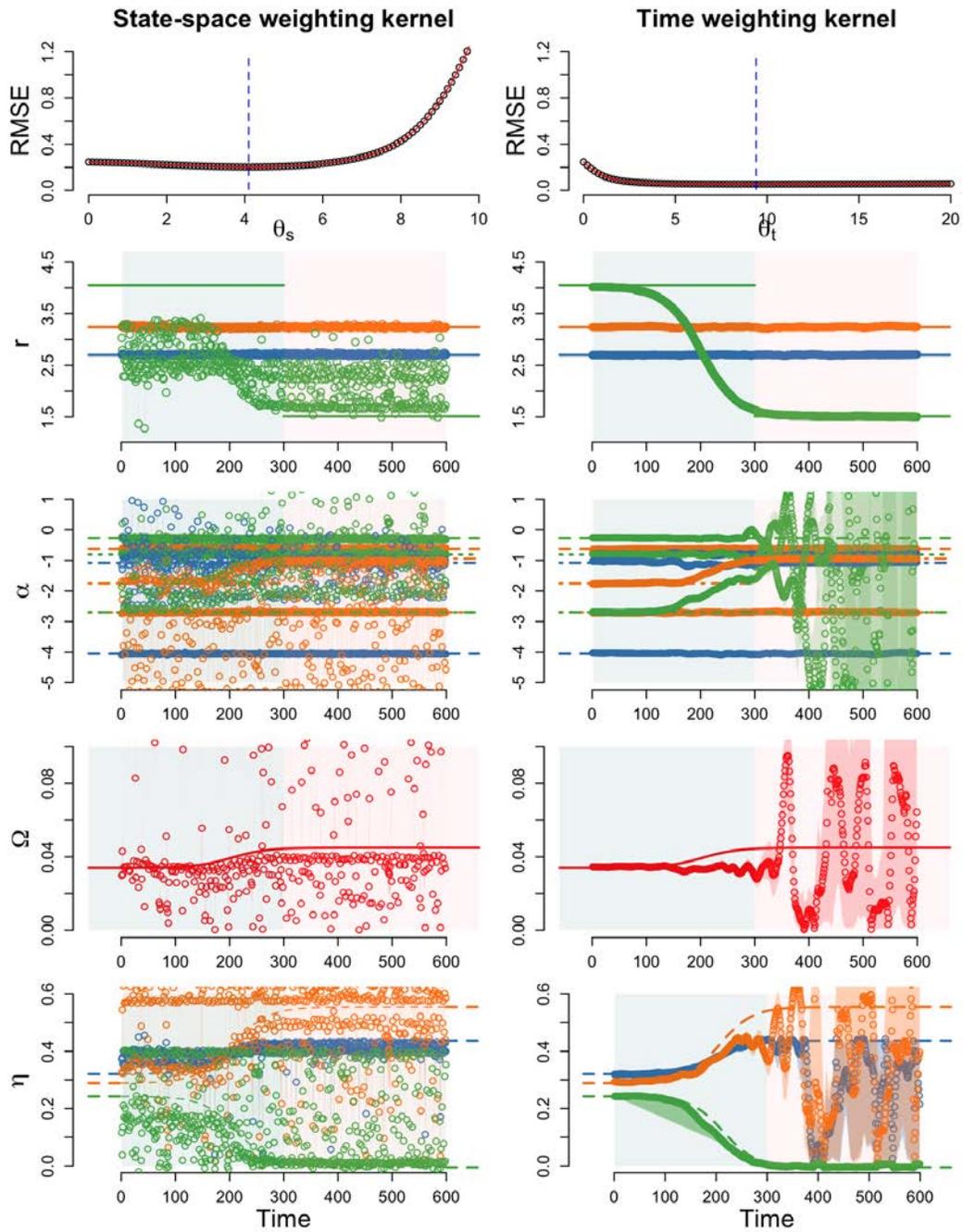
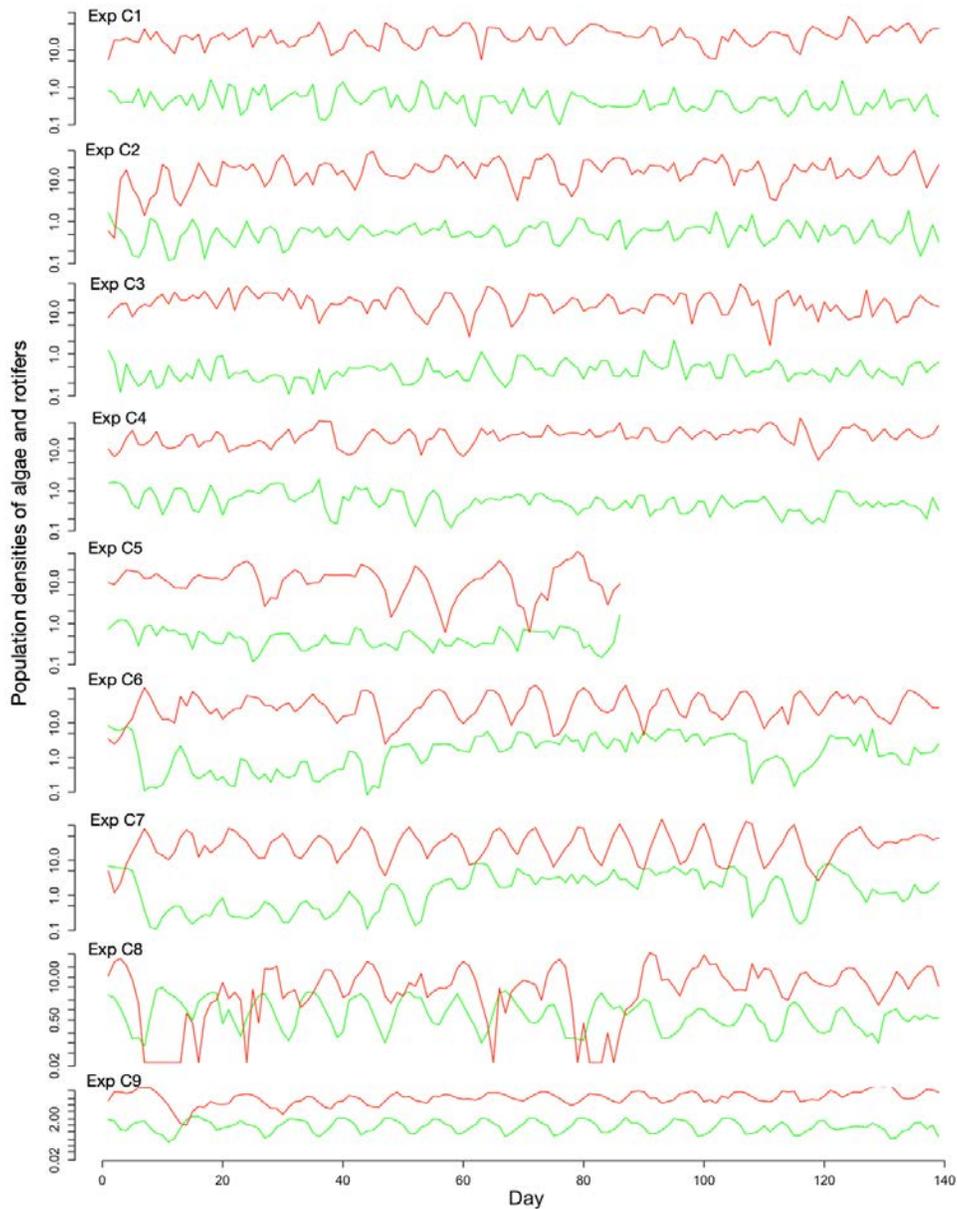


Fig. S33. Dynamics with extinction and small environmental noise  $\epsilon = 0.01$ .

## 9. Supporting Figures for Experimental Data

### A. Results using data from Blasius et al. 2020.



**Fig. S34.** Population dynamics for all nine experiments (rotifers in red and algae in green). Exp C1–C4: constant environmental conditions with influx  $80\mu\text{mol}/\text{l}$ , outflux  $0.55/\text{day}$ . Exp C5: constant environmental conditions with influx  $80\mu\text{mol}/\text{l}$ , outflux  $0.66/\text{day}$ . Exp C6 – C7: alternated nutrition influx between  $160\mu\text{mol}/\text{l}$  and  $0\mu\text{mol}/\text{l}$  every 8 days, and outflux rate of  $0.55/\text{day}$ . Exp C8–C9: constant environmental conditions with influx  $160\mu\text{mol}/\text{l}$  and outflux  $0.55/\text{day}$ . In the first seven experiments C1–C7, *M. minutum* was used whereas in exp C8–C9, *C. vulgaris* was used.

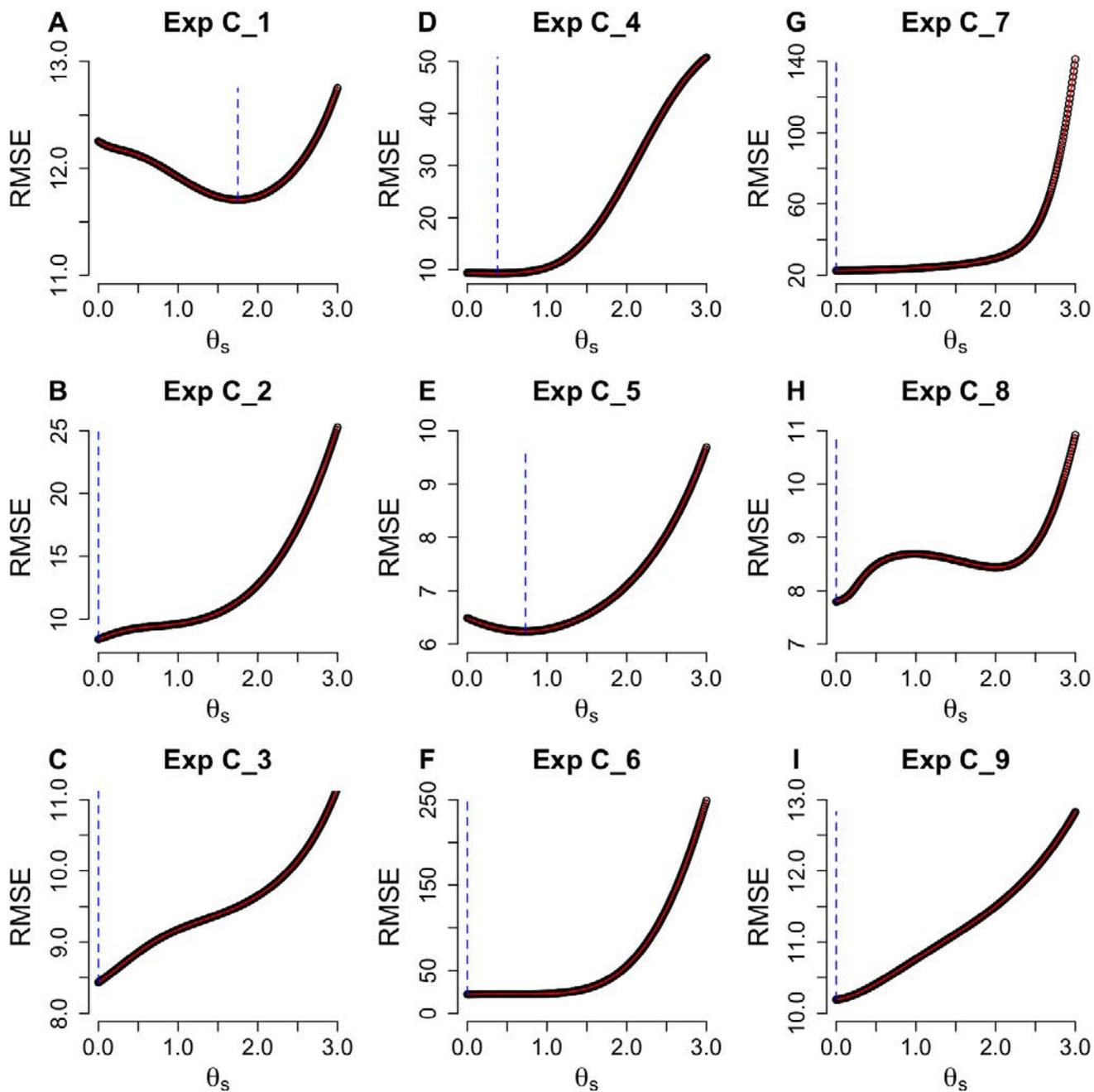


Fig. S35. Cross-validation results using state-space weighting kernel. The vertical dashed blue lines indicate the smallest values of the RMSE.

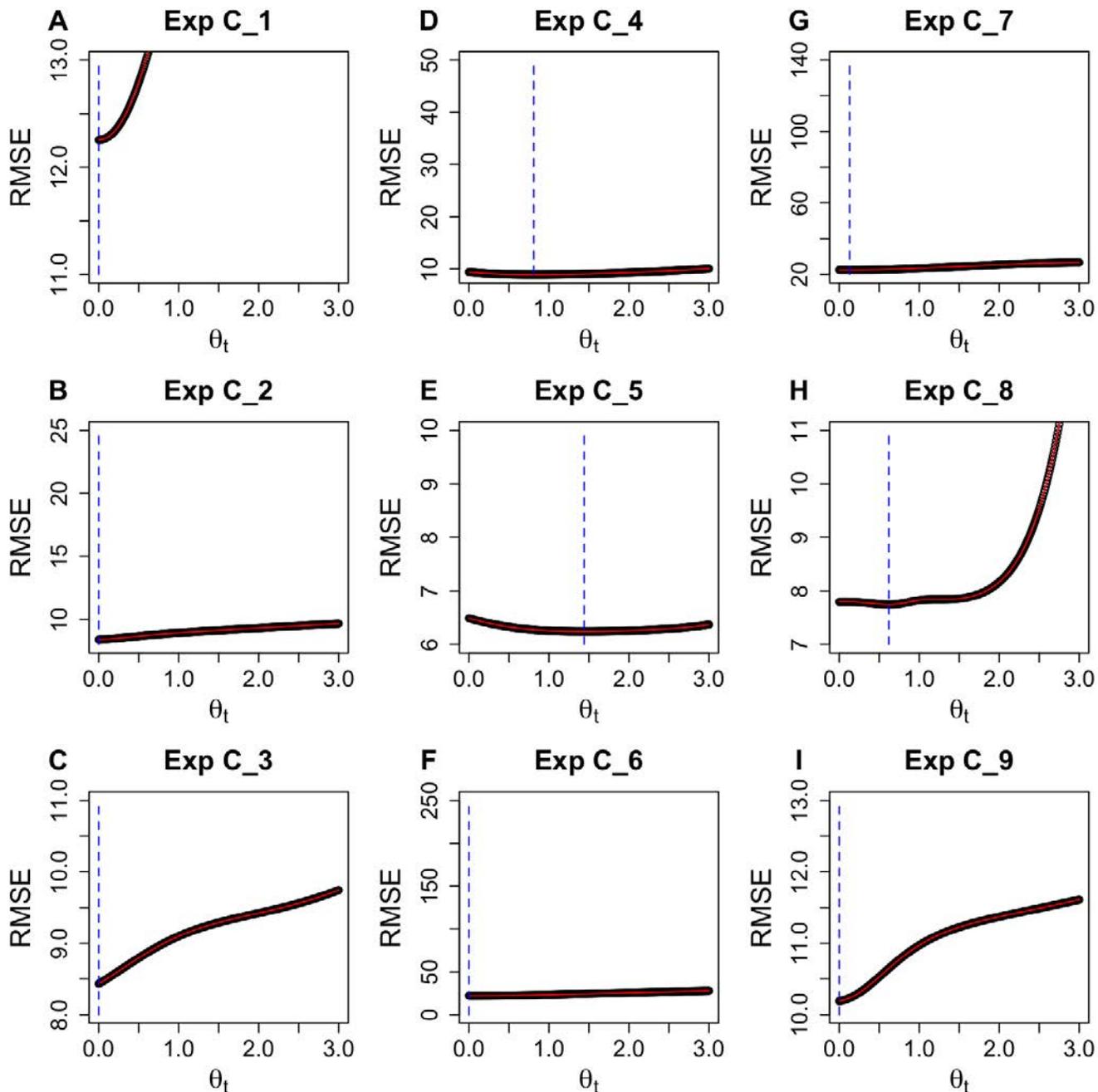


Fig. S36. Cross-validation results using time weighting kernel. The vertical dashed blue lines indicate the smallest values of the RMSE

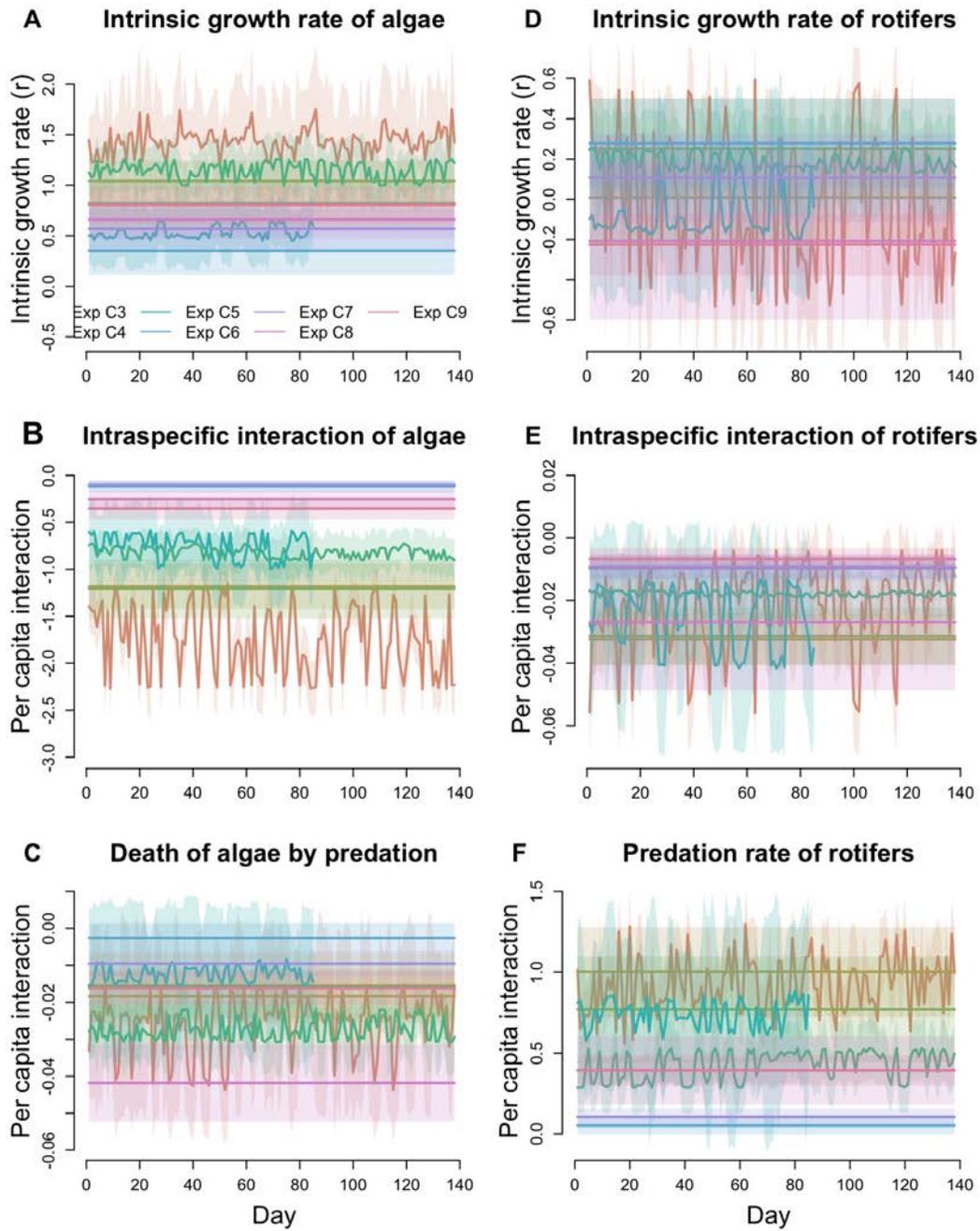


Fig. S37. Inference of the intrinsic growth rate  $r$  and *per capita* interactions of rotifers and algae, using state-space weighting kernel.

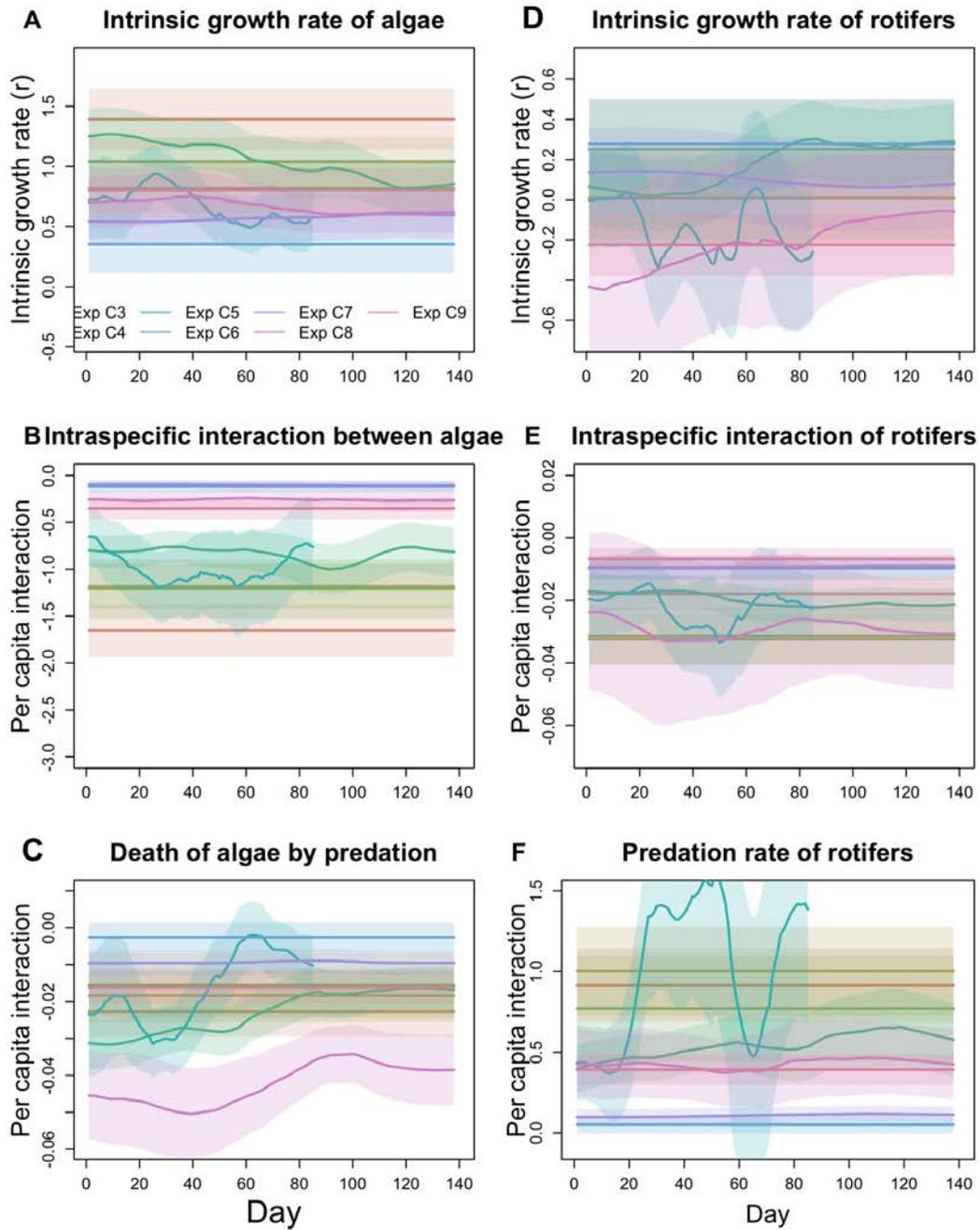
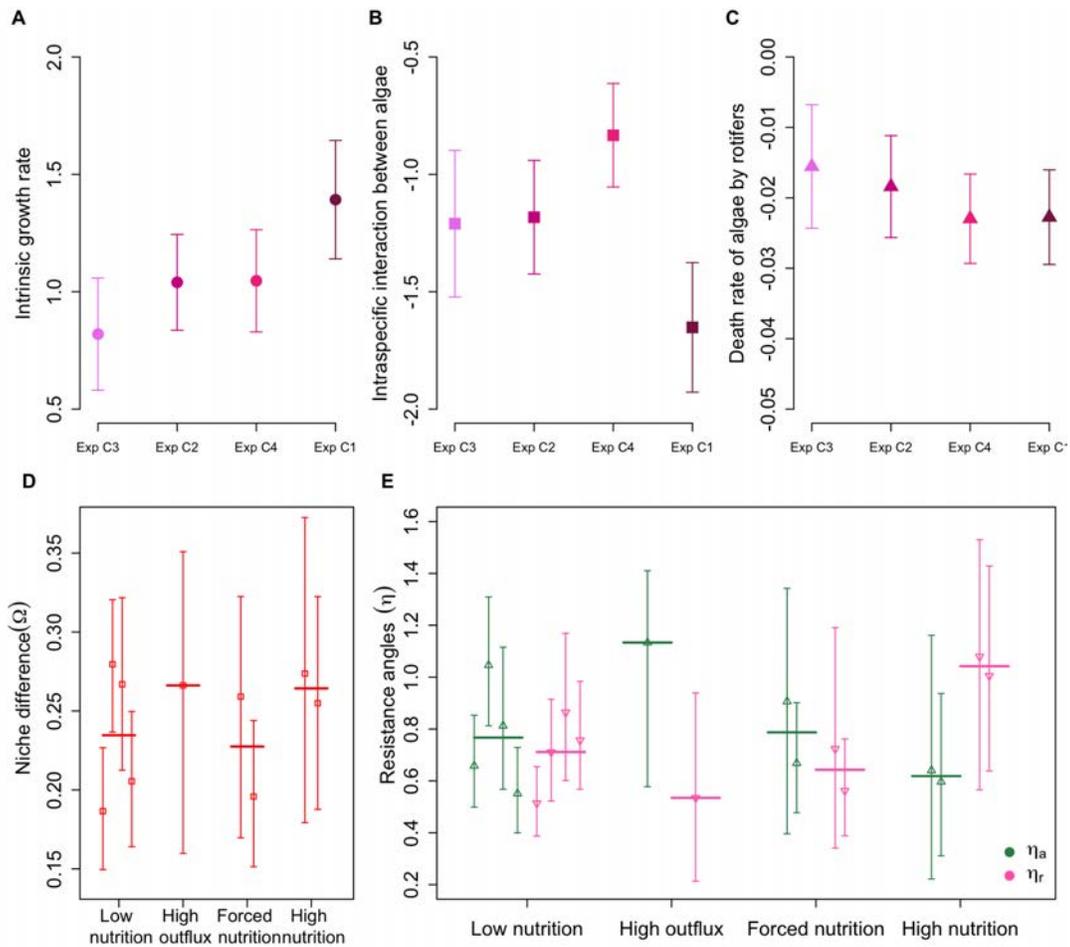


Fig. S38. Inference of the intrinsic growth rate  $r$  and *per capita* interactions of rotifers and algae, using time weighting kernel.



**Fig. S39.** Same as figure 3 in the main text, but using the time weighting kernel for the inference. Trade-off between the inferred intrinsic growth rate of algae and grazer defense (i.e. *per capita* interaction of rotifers on algae)

B. Results using data from Yoshida et al. 2003.

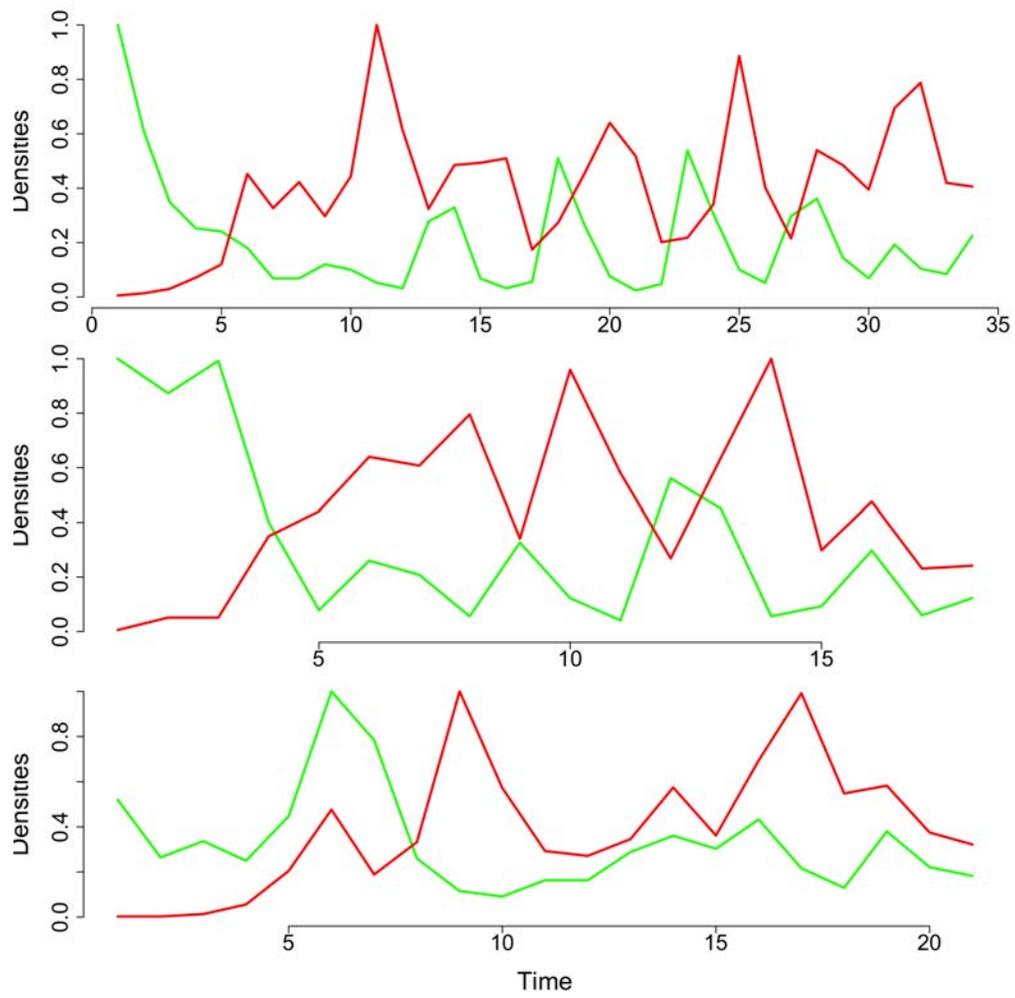


Fig. S40. Population dynamics for three experiments (rotifers in red and algae in green).

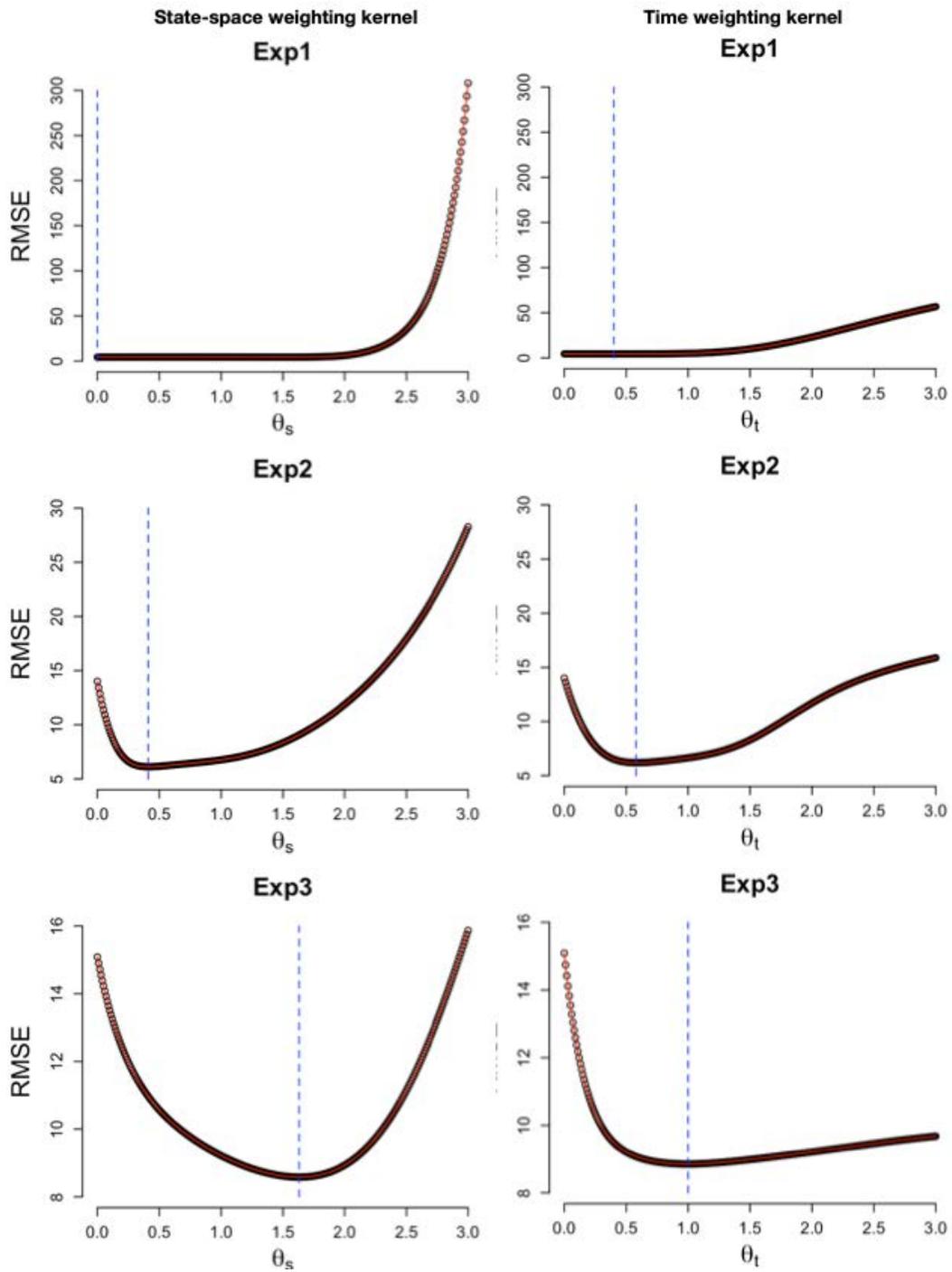
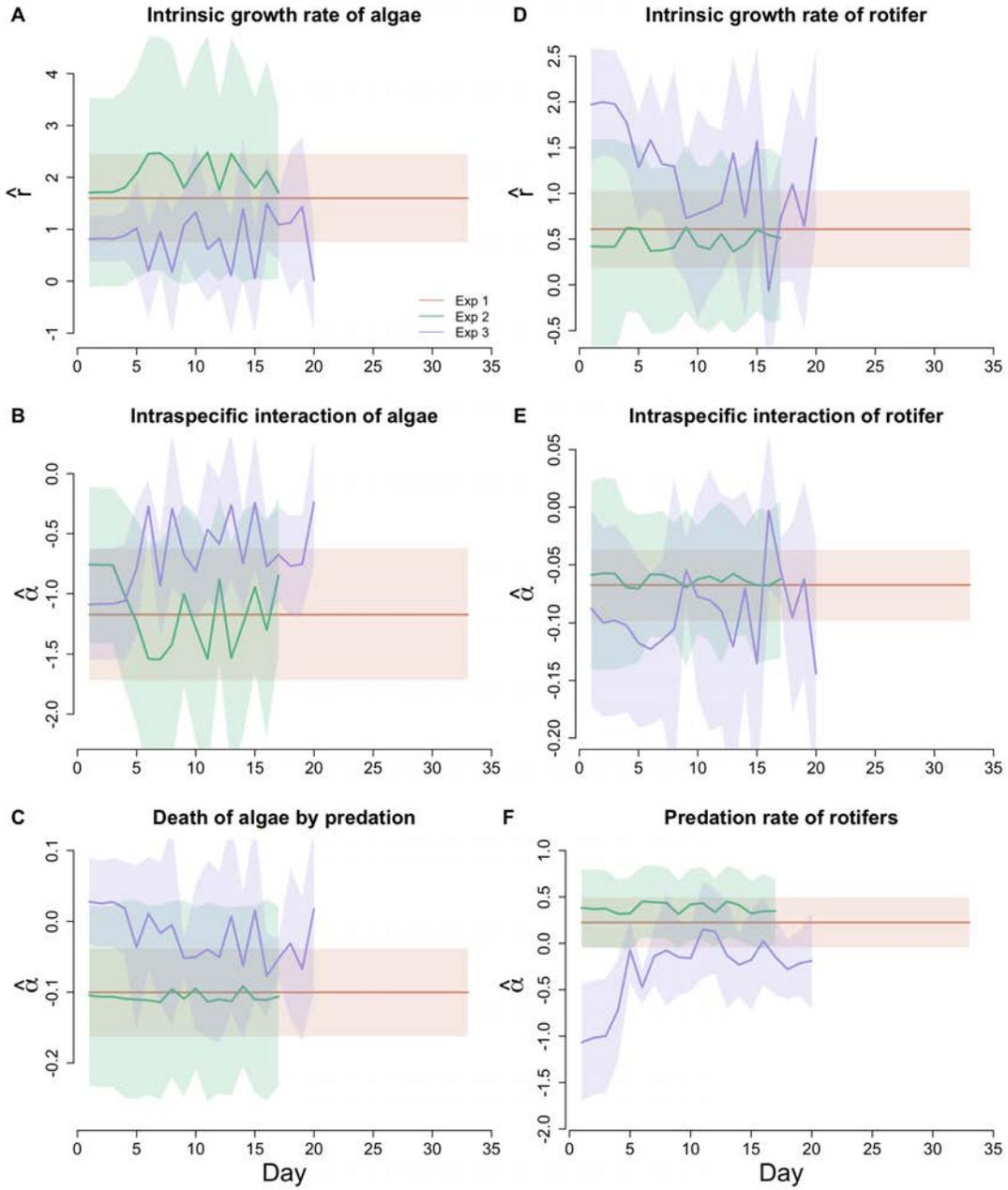


Fig. S41. Cross-validation results. The vertical dashed blue lines indicate the minimum values of the RMSE.



**Fig. S42.** Inference of the intrinsic growth rate  $r$  and *per capita* interactions of rotifers and algae, using state-space weighting kernel.

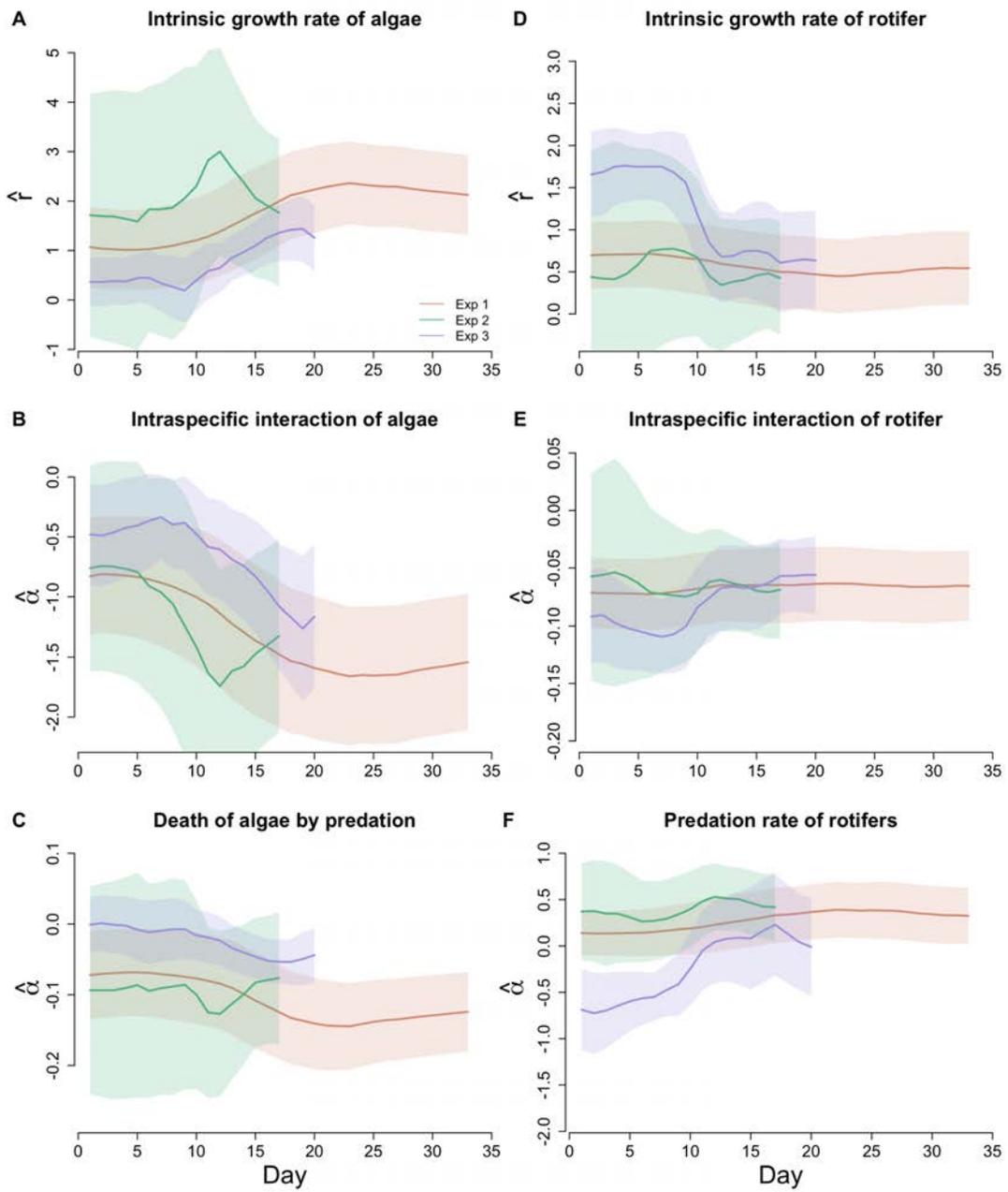


Fig. S43. Inference of the intrinsic growth rate  $r$  and *per capita* interactions of rotifers and algae, using time weighting kernel.

## 10. Supporting Figures for Observational Data

### A. Population dynamics.

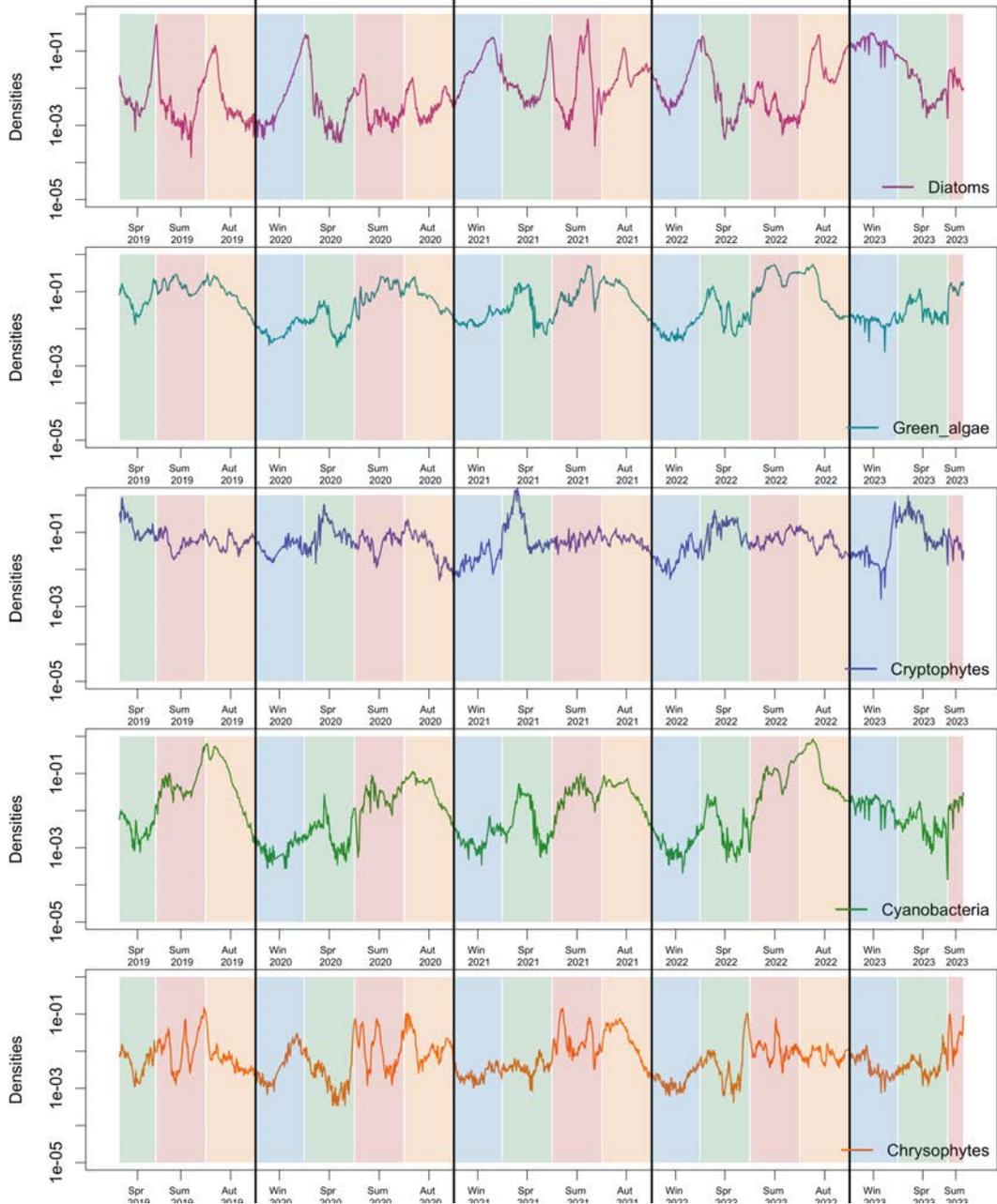


Fig. S44. Population densities of five autotrophic guilds: Diatoms, Green algae, Cryptophytes, Cyanobacteria, and Chrysophytes.

## B. Cross-validation results.

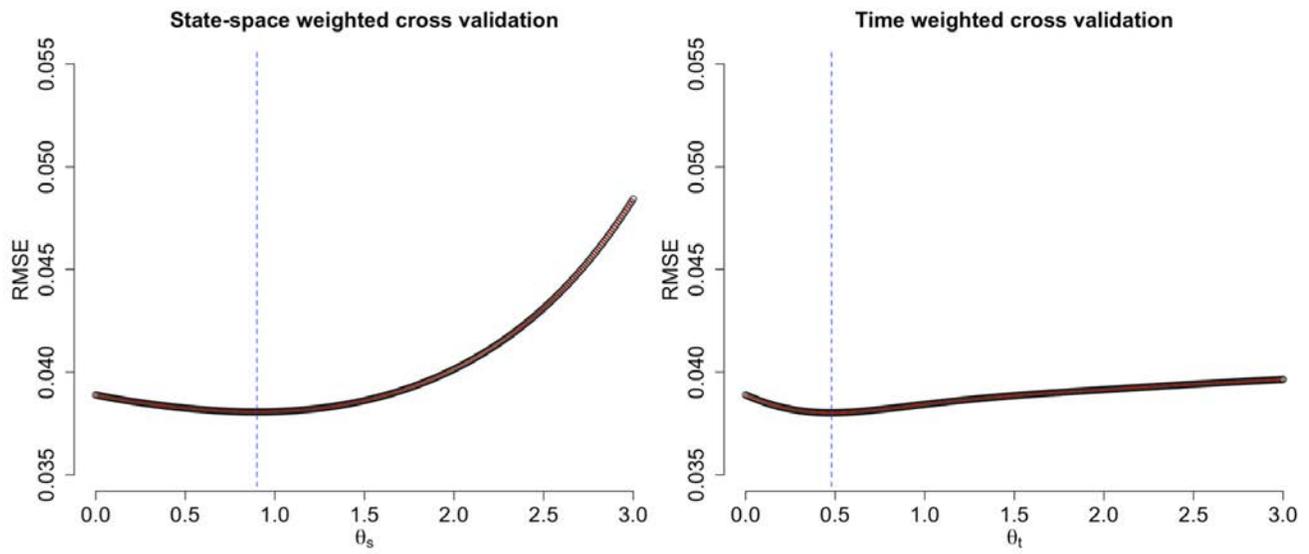


Fig. S45. Cross-validation for state-space weighting kernel ( $\theta_s$ ), and time weighting kernel ( $\theta_t$ ). The vertical dashed blue line indicate the minimum values of the RMSE.

C. Inference results – Ecological parameters.

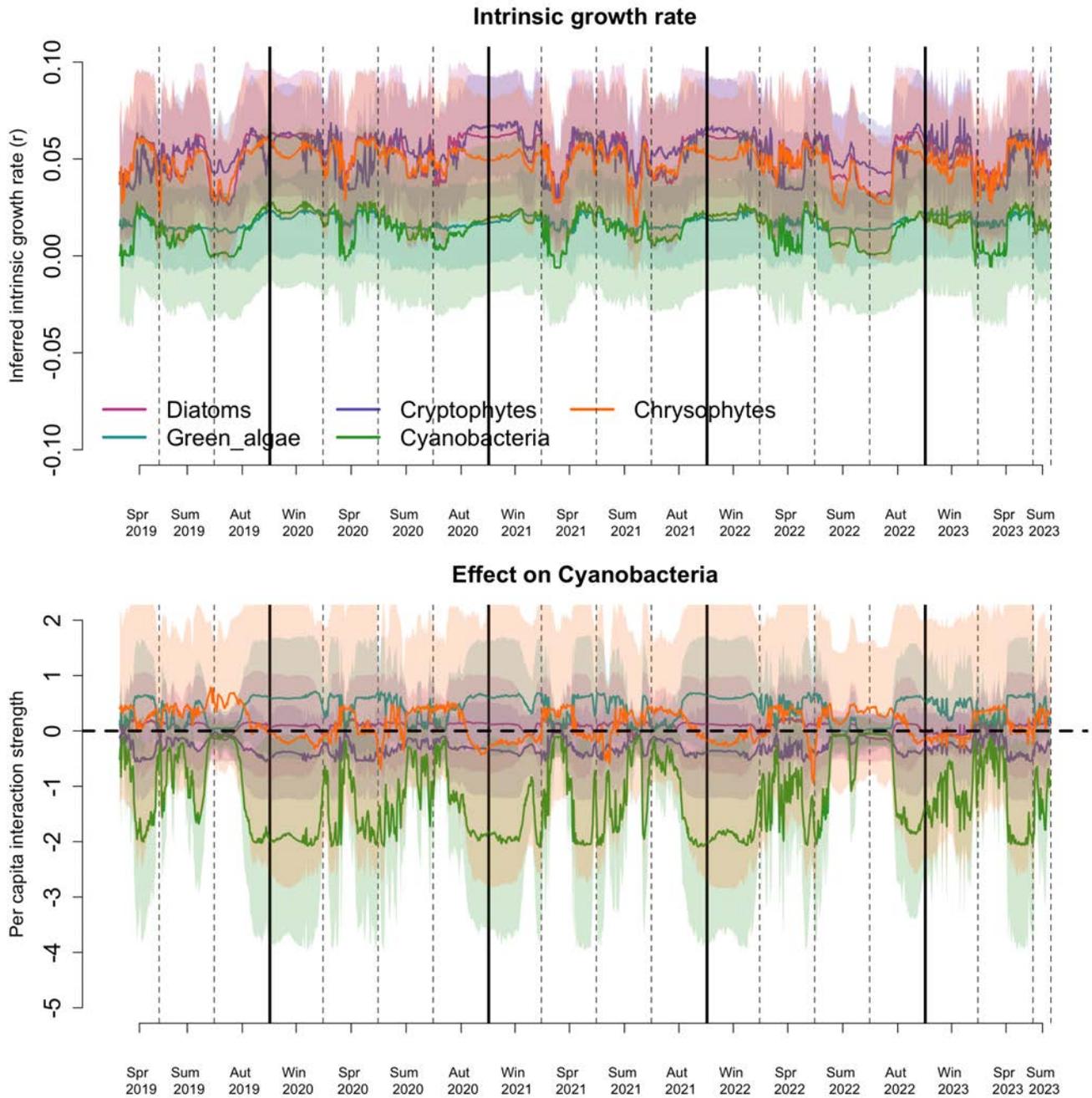
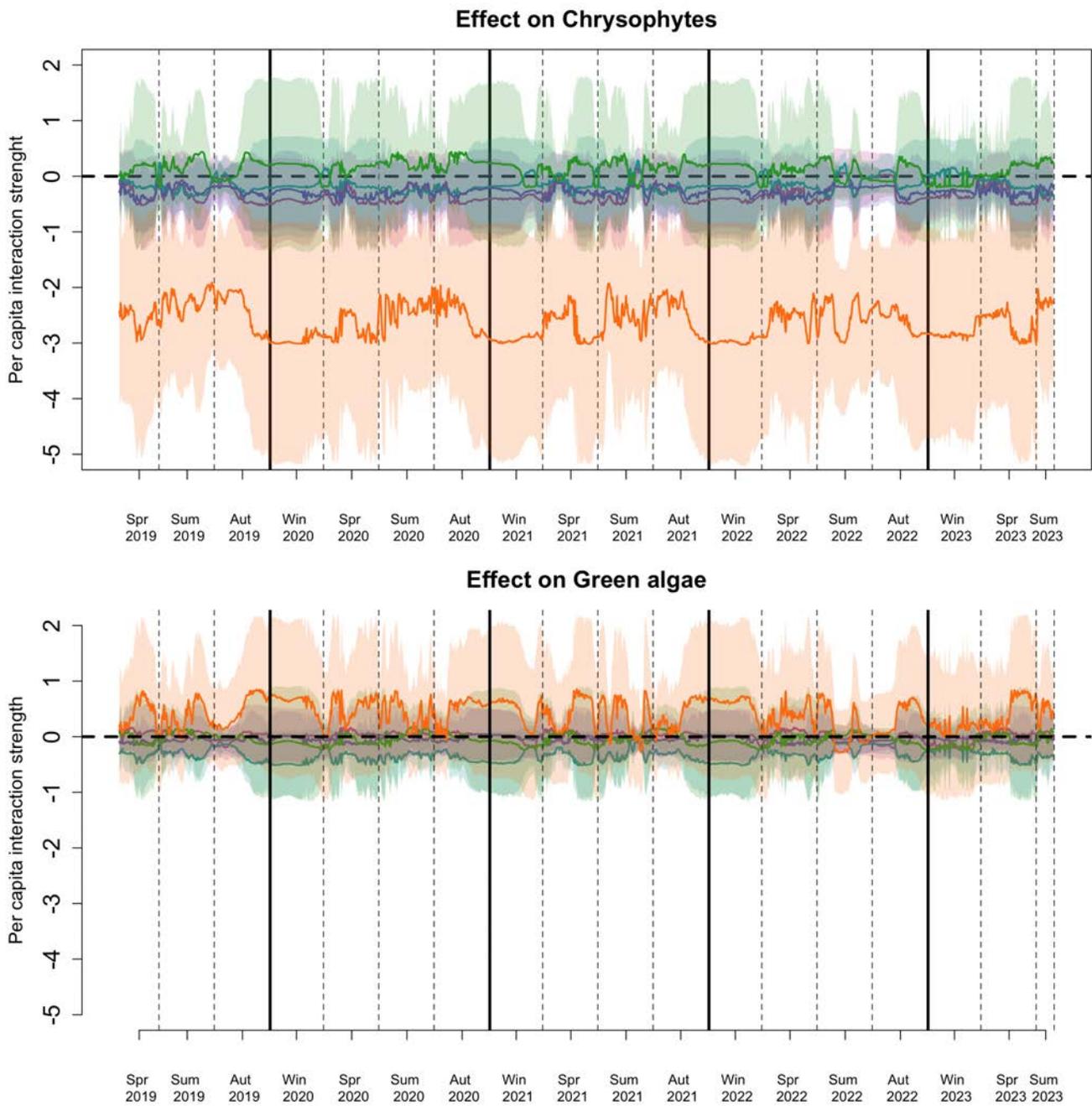
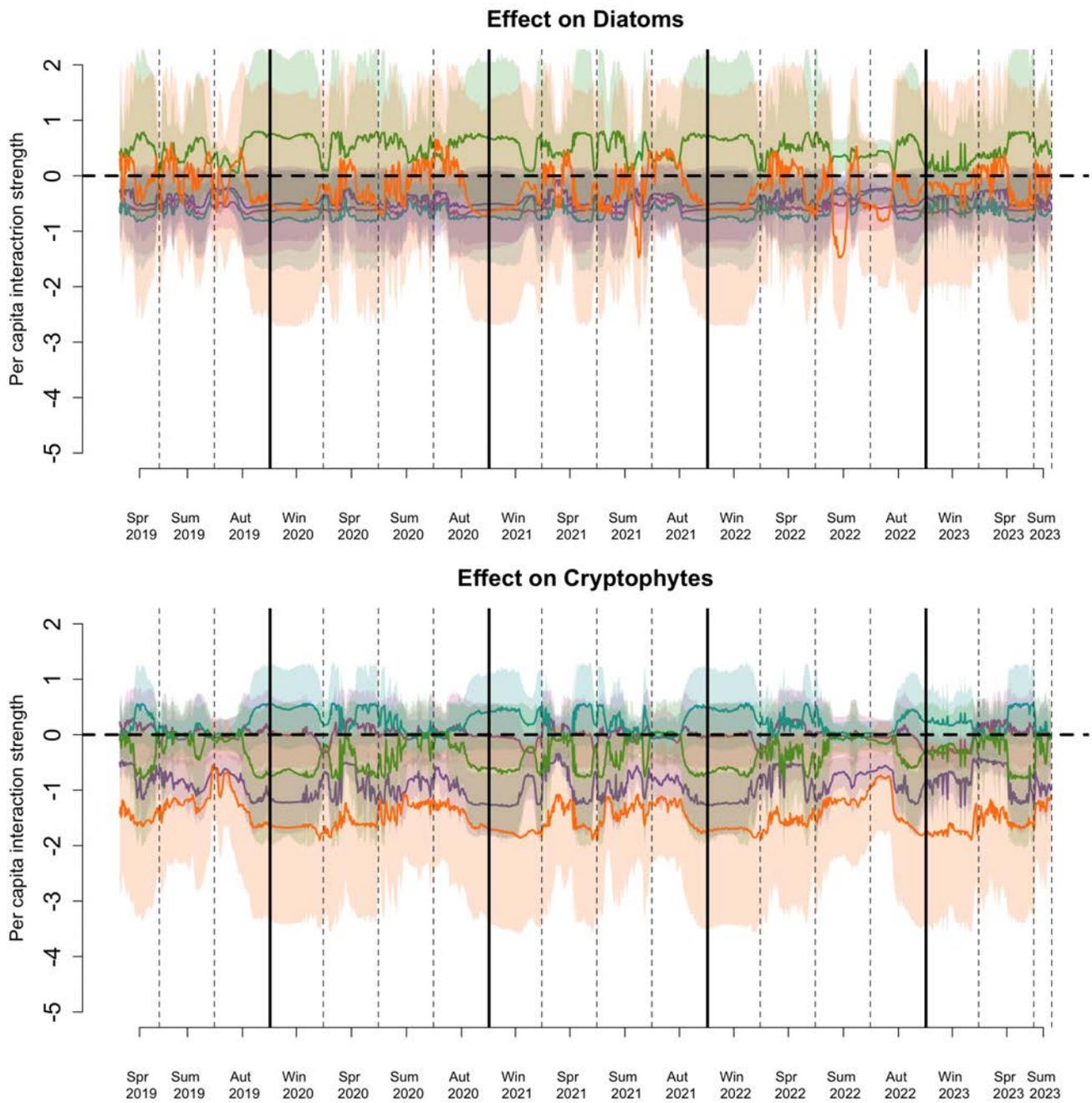


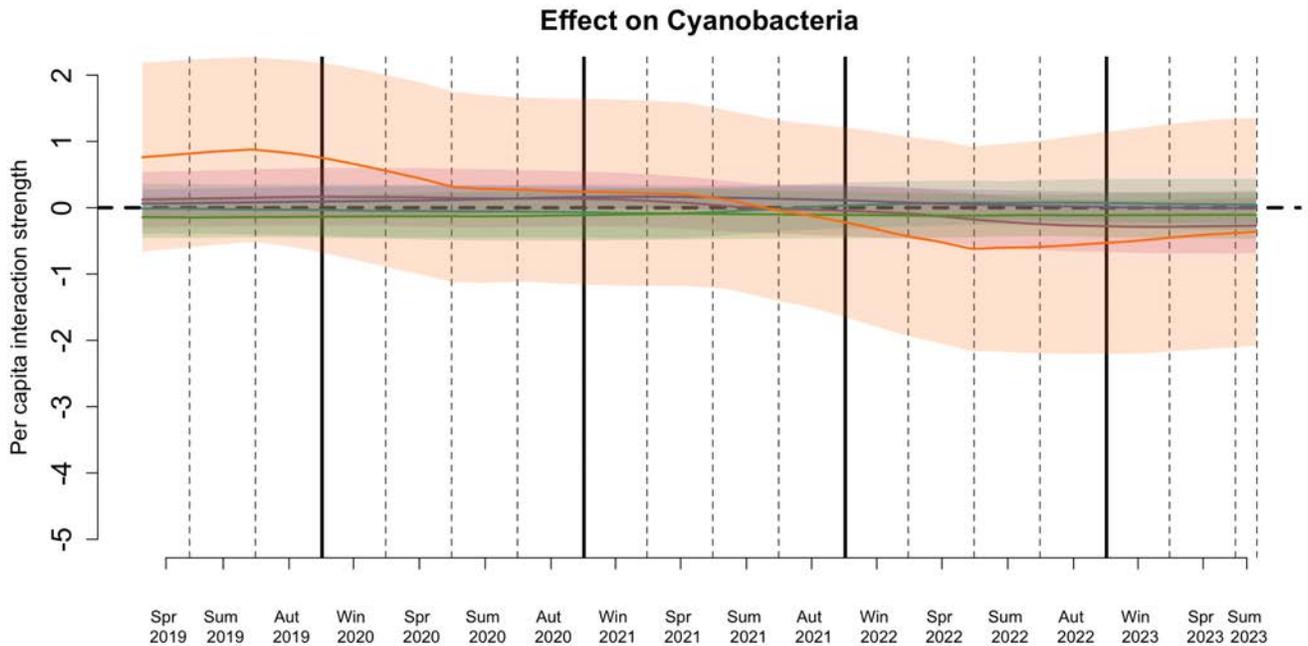
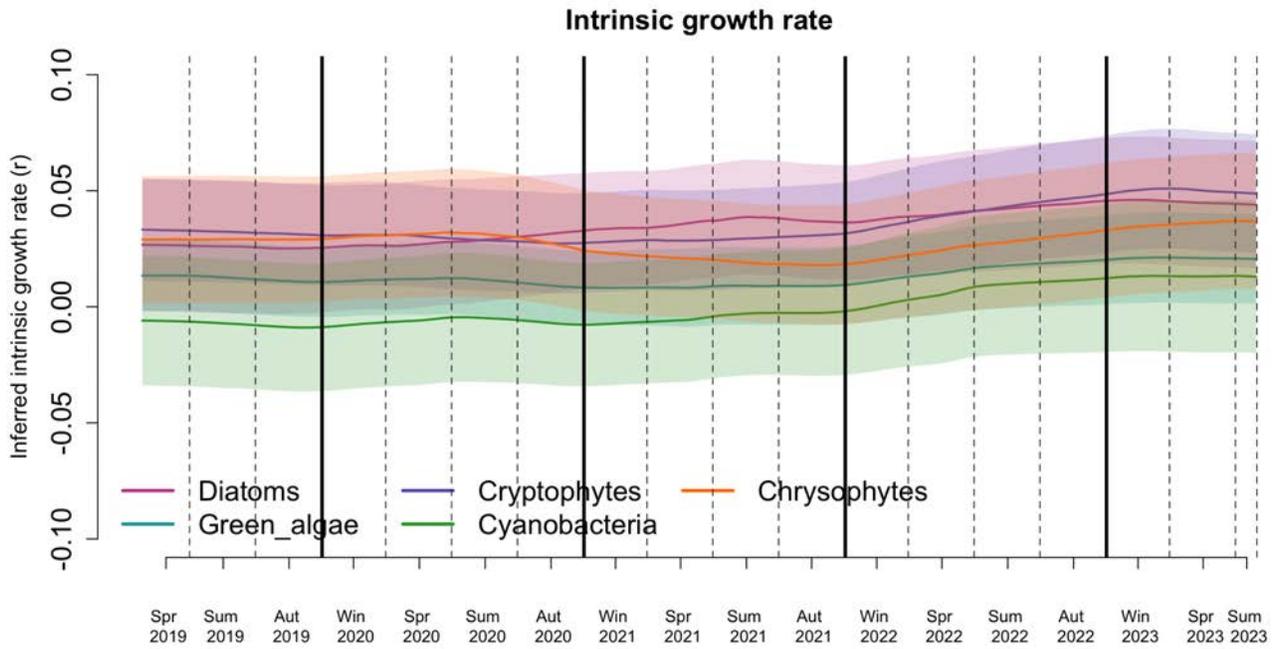
Fig. S46. Inference of the intrinsic growth rate and the *per capita* interactions of other guilds on Cyanobacteria, using the state-space weighting kernel.



**Fig. S47.** Inference of the *per capita* interactions of other guilds on Chrysophytes and Green algae, using the state-space weighting kernel.



**Fig. S48.** Inference of the *per capita* interactions of other guilds on Diatoms and Cryptophytes, using the state-space weighting kernel.



**Fig. S49.** Inference of the intrinsic growth rate and the *per capita* interactions of other guilds on Cyanobacteria, using the time weighting kernel.

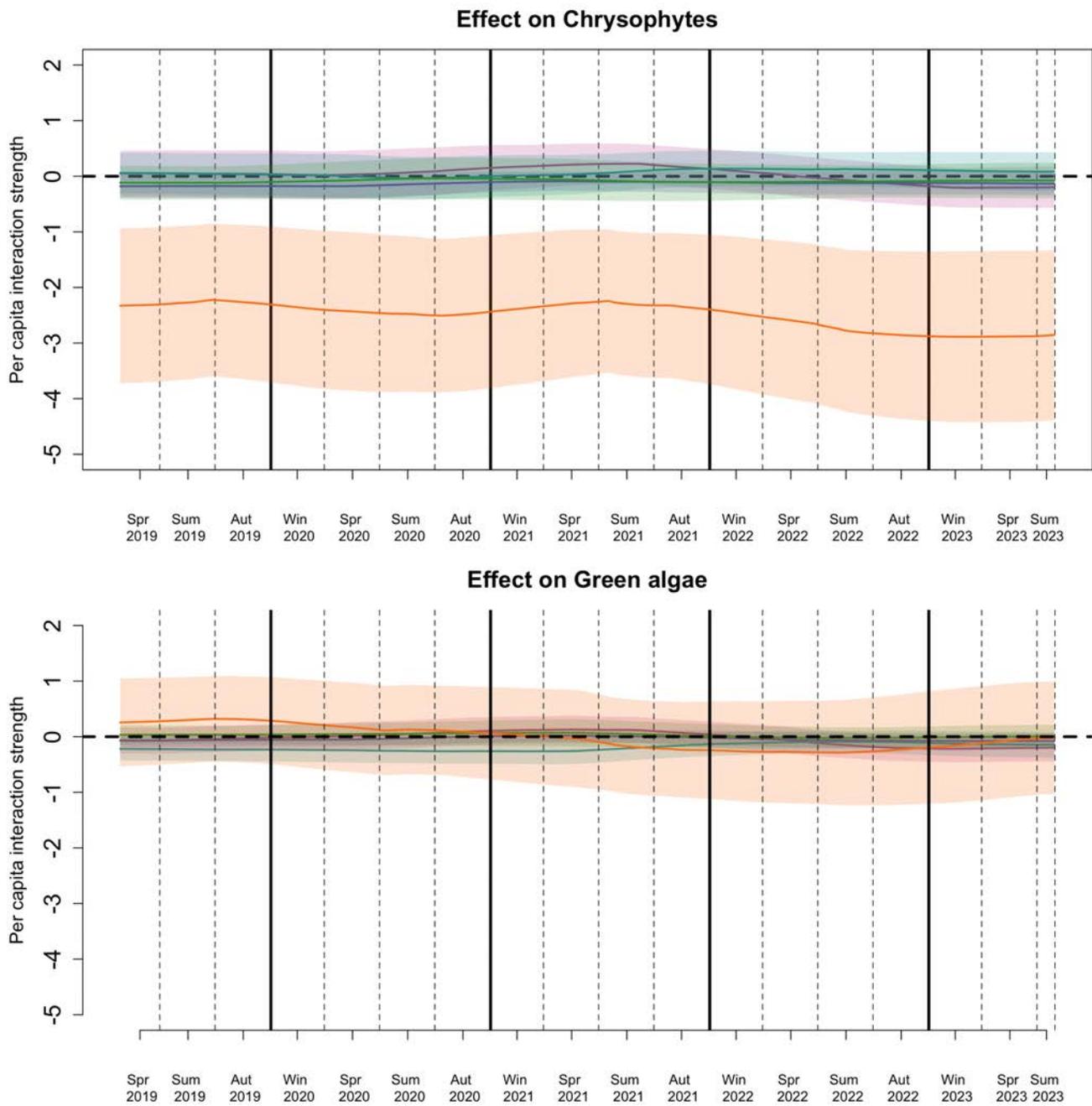


Fig. S50. Inference of the *per capita* interactions of other guilds on Chrysophytes and Green algae, using the time weighting kernel.

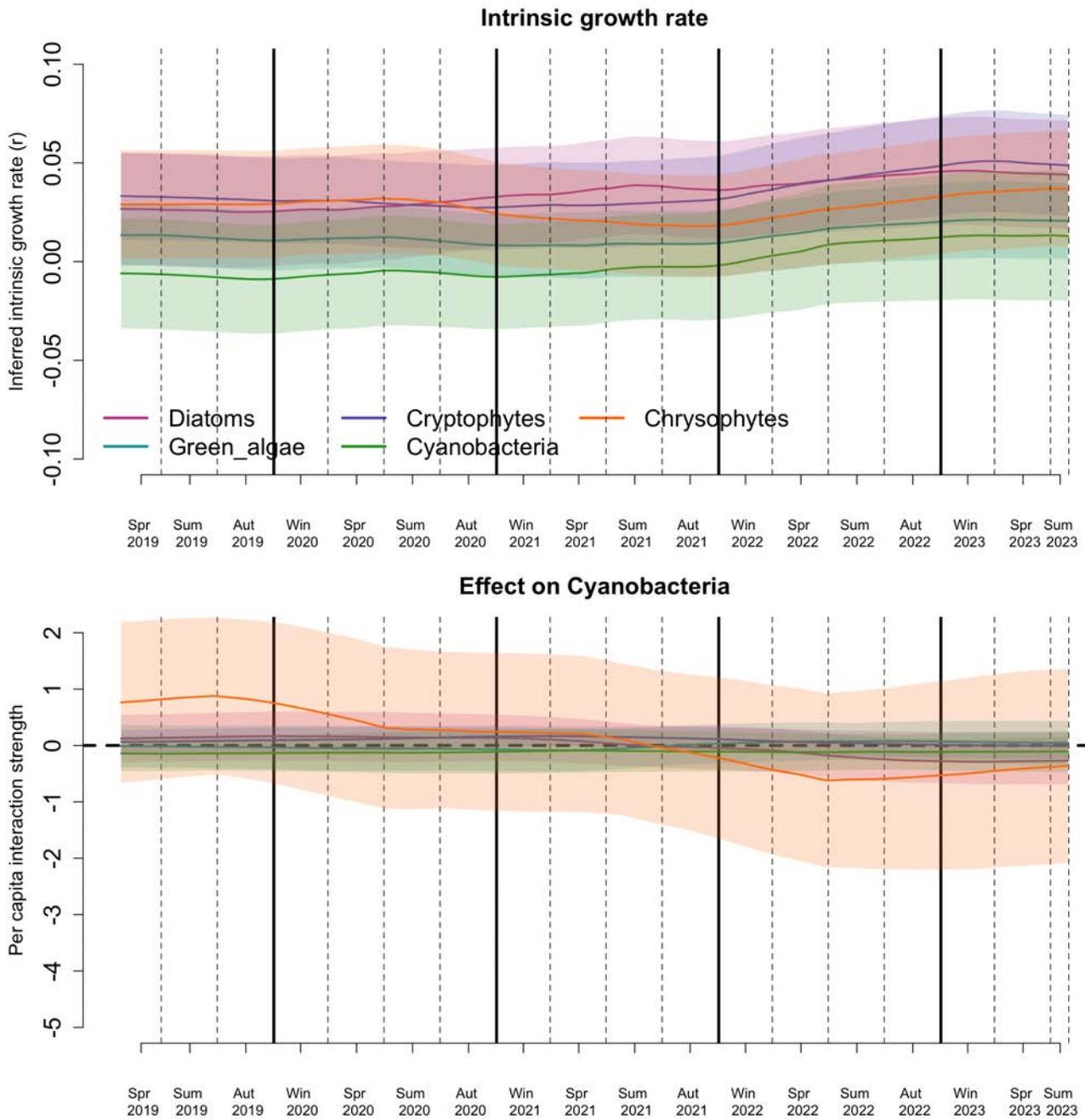


Fig. S51. Inference of the *per capita* interactions of other guilds on Diatoms and Cryptophytes, using the time weighting kernel.

D. Inference results – Coexistence metrics.

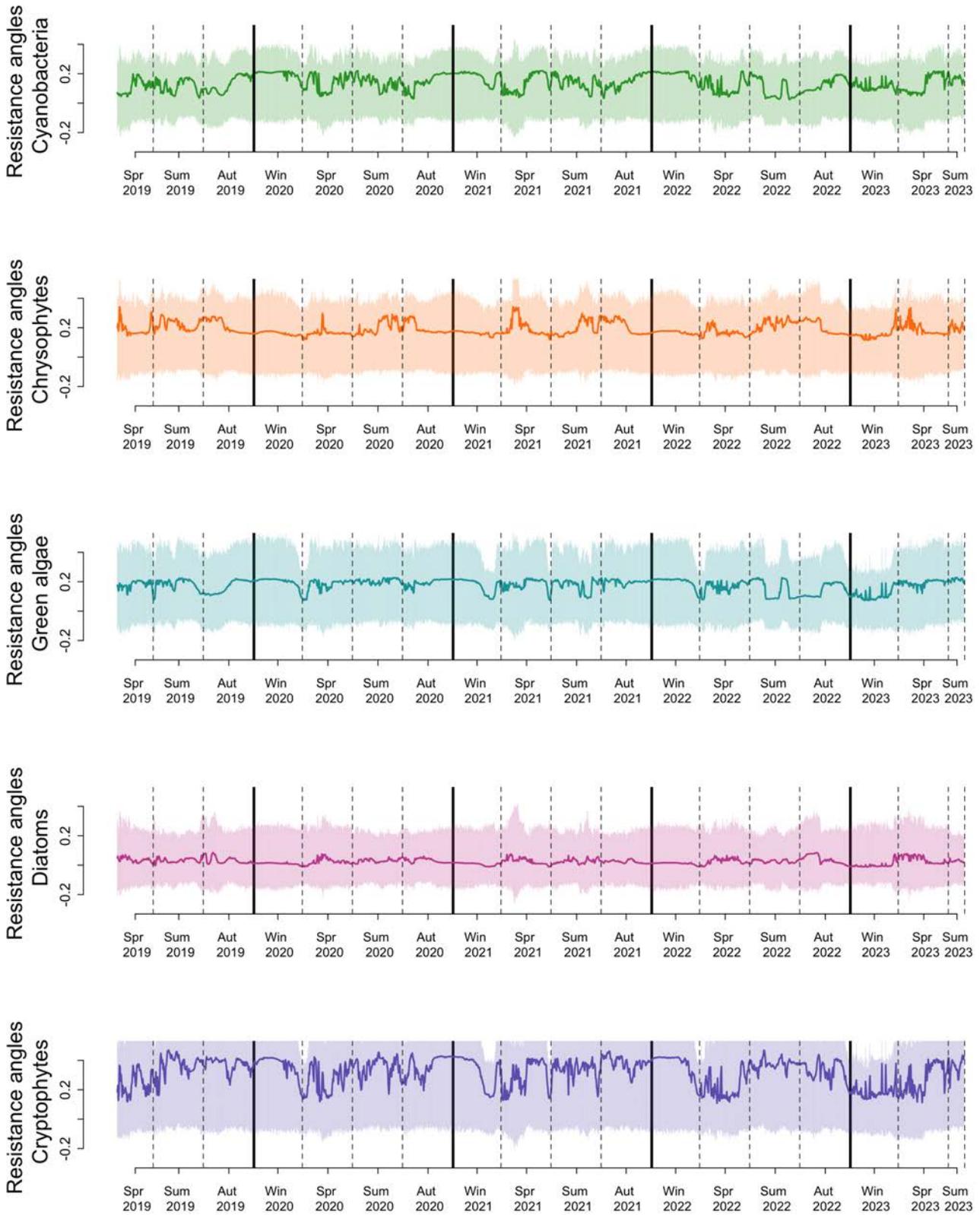


Fig. S52. Inference of the resistance angles of all guilds, using the state-space weighting kernel.

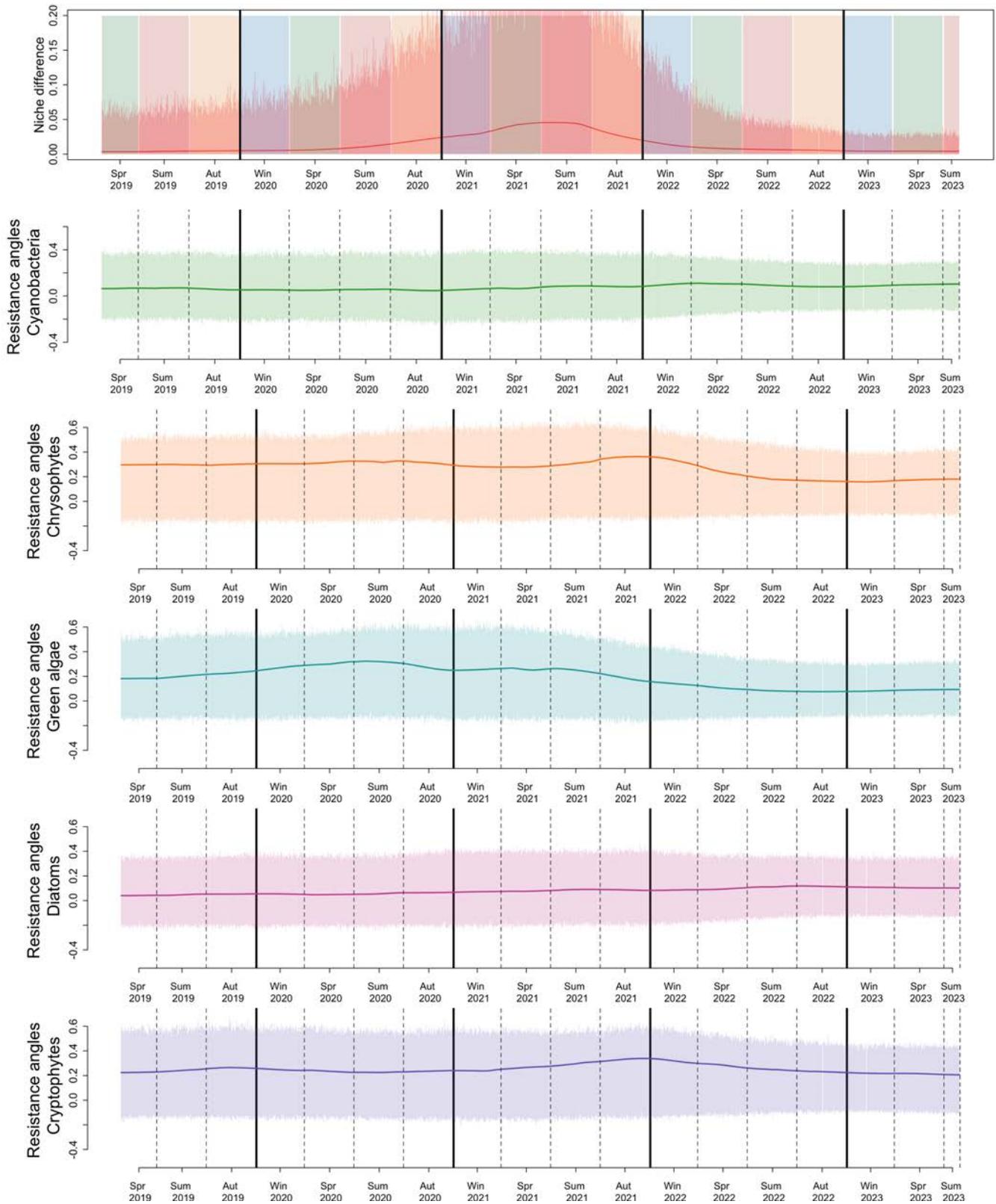


Fig. S53. Inference of the niche difference and resistance angles of all guilds, using the time weighting kernel.

## References

1. R Arditi, YV Tyutyunov, LI Titova, RP Rohr, LF Bersier, The Dimensions and Units of the Population Interaction Coefficients. *Front. Ecol. Evol.* **9** (2021).
2. T Hastie, R Tibshirani, J Friedman, *The Elements of Statistical Learning*. (Springer New York Inc.), (2001).
3. G Sugihara, Nonlinear Forecasting for the Classification of Natural Time Series. *Philos. Transactions: Phys. Sci. Eng.* **348**, 477–495 (1994).
4. ER Deyle, RM May, SB Munch, G Sugihara, Tracking and forecasting ecosystem interactions in real time. *Proc. Royal Soc. B: Biol. Sci.* **283**, 20152258 (2016).
5. N Raillard, M Prevosto, P Ailliot, Modeling process asymmetries with Laplace moving average. *Comput. Stat. & Data Analysis* **81**, 24–37 (2015).
6. L Fahrmeir, T Kneib, S Lang, BD Marx, *Regression: Models, Methods and Applications*. (Springer Berlin Heidelberg, Berlin, Heidelberg), (2021).
7. K Mardia, J Kent, J Bibby, *Multivariate analysis*. (Acad. Press), (1979).
8. MJ Ribando, Measuring solid angles beyond dimension three. *Discret. & Comput. Geom.* **36**, 479–487 (2006).
9. S Saavedra, et al., A structural approach for understanding multispecies coexistence. *Ecol. Monogr.* **87**, 470–486 (2017).
10. A Genz, F Bretz, *Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics, Vol. 195*. (Springer-Verlage, Heidelberg), (2009).
11. A Genz, et al., *mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-5*. (R Foundation for Statistical Computing), (2016).
12. F Eriksson, On the measure of solid angles. *Math. Mag.* **63**, 184–187 (1990).
13. VJ Lepori, N Loeuille, RP Rohr, Robustness versus productivity during evolutionary community assembly: short-term synergies and long-term trade-offs. *Proc. Royal Soc. B: Biol. Sci.* **291** (2024).
14. AC Davison, HD V., *Bootstrap Methods and their Application*. (Cambridge University Press), (1997).
15. P Turchin, *Complex population dynamics: a theoretical/empirical synthesis*, Monographs in population biology. (Princeton University Press, Princeton, N.J) No. 35, (2003).
16. B Blasius, L Rudolf, G Weithoff, U Gaedke, GF Fussmann, Long-term cyclic persistence in an experimental predator–prey system. *Nature* **577**, 226–230 (2020).
17. T Yoshida, LE Jones, SP Ellner, GF Fussmann, NG Hairston, Rapid evolution drives ecological dynamics in a predator–prey system. *Nature* **424**, 303–306 (2003).
18. E Merz, et al., Underwater dual-magnification imaging for automated lake plankton monitoring. *Water Res.* **203**, 117524 (2021).
19. SP Kyathanahally, et al., Deep learning classification of lake zooplankton. *Front. Microbiol.* **12** (2021).