

Supplementary information file for *Fast fitting of neural ordinary differential equations by Bayesian neural gradient matching to infer ecological interactions from time series data*

Willem Bonnaffé^{1,2} & Tim Coulson²

1. Big Data Institute, University of Oxford, Old Road Campus, Oxford OX3 7LF
2. Department of Biology, University of Oxford, Zoology Research and Administration Building, 11a Mansfield Road, Oxford OX1 3SZ

Contact: Willem Bonnaffé, 128 Southfield Park, Oxford, OX4 2BA, UK (w.bonnafe@gmail.com)

Statement of authorship: Willem Bonnaffé designed the method, performed the analysis, wrote the manuscript; Tim Coulson led investigations, provided input for the manuscript, commented on the manuscript.

Contents

A	Bayesian regularisation	1
B	Complementary results benchmark analysis	4
C	Complementary results case study 2 replicate A	9
D	Complementary results case study 2 replicate B	12
E	Complementary results case study 2 replicate C	15
F	Complementary results case study 3	18
G	Complementary results case study 4	21

A Bayesian regularisation

The fitting of the models is performed in a Bayesian framework, considering normal error structure for the residuals, and normal prior density distributions on the parameters,

$$p(\omega|\mathcal{D}) \propto p(\mathcal{D}|\omega)p(\omega), \quad (1)$$

where θ is the parameter vector of the model, and \mathcal{D} the evidence, namely the data that the model is fitted to. Assuming a normal likelihood for the residuals given the evidence we get

$$p(\mathcal{D}|\omega) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{e_i(\mathcal{D}, \omega)^2}{2\sigma^2} \right\}, \quad (2)$$

where $e_i(\mathcal{D}, \omega)$ are the residuals of the model given the parameters, and the evidence. In the case of the interpolation, the residuals correspond to the observation error $\varepsilon^{(o)}$ (Eq. 3). In the case of the NODE approximation, they correspond to the process error $\varepsilon^{(p)}$ (Eq. 7). N is the number of data points, either observations in the case of the interpolation, $N^{(o)}$, or interpolated points in the case of the NODE fitting, $N^{(p)}$.

The prior probability density functions for the parameters are given by

$$p(\omega) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi\delta_j^2}} \exp \left\{ -\frac{\omega_j^2}{2\delta_j^2} \right\}, \quad (3)$$

where M is the number of parameters in the models. The parameter δ_j controls the dispersion of the priors, and thereby the complexity/level of constraint of the model.

Bayesian regularisation consists in constraining the values of the parameters in the model to be close to a desired value. Usually, parameters are constrained by choosing normal priors centered about 0. In this case, the standard deviation of the normal priors, δ , governs the range of values that the parameters can take, and hence constrains more or less strongly the behaviour of the model (Cawley and Talbot 2007). Low values of dispersion may increase constraint on parameters too drastically, which would lead to underfitting, and result in a reduction of the variance of parameter estimates and bias mean estimates towards 0. In contrast, too high values of dispersion may lead to overfitting, by allowing for more complex shapes. To account for this, we optimise models on the second-level of inference. This means that we are finding the optimal value of δ , in addition to optimising the model parameters.

In practice, choosing the level of constraint is difficult, Cawley and Talbot hence developed a criterion to perform model selection on the second level of inference. They proposed to optimise the marginal posterior distribution by averaging out the dispersion of the priors. With an improper Jeffrey's hyperprior, $p(\xi) \propto \frac{1}{\xi}$, on the dispersion parameters, the dispersion can be integrated out, leaving us with a formula for the posterior that only depends on the parameters of the model,

$$\log P(\omega|\mathcal{D}) \propto -\frac{N}{2} \log \left(\sum_{i=1}^N e_i(\mathcal{D}, \omega)^2 \right) - \frac{M}{2} \log \left(\sum_{j=1}^M \omega_j^2 \right), \quad (4)$$

where $P(\omega|\mathcal{D})$ denotes the marginal posterior density, \mathcal{D} denotes the evidence, N and M denote the number of data points and parameters, respectively, e_i denote the residuals, and ω denote the parameters of the model. The construction is elegant because it is not sensitive to the choice of prior hyperparameters, and simple as it amounts to optimising the log of the sum of squares, rather than the sum of squares (in the case of normal ordinary least square).

One of the drawbacks of this construction is that the marginal posterior density is not finite when the parameters are 0, which can lead to underfitting. In this paper we use a modified criterion, which corrects for that problem,

$$\log P(\omega|\mathcal{D}) \propto -\frac{N}{2} \log \left(1 + \sum_{i=1}^N e_i(\mathcal{D}, \omega)^2 \right) - \frac{M}{2} \log \left(1 + \sum_{j=1}^M \omega_j^2 \right), \quad (5)$$

where the marginal posterior density depends only on the residuals of the model when the parameters are equal to 0, and otherwise depends on both the parameters and the residuals. This construction can be obtained by adding an exponential correction to the improper Jeffrey's hyperprior that Cawley and Talbot used in their original study, namely $p(\xi) \propto \frac{1}{\xi} \exp\{-\xi/2\}$, where ξ is the regularisation parameter.

We provide below details of the calculation of this modified criterion, starting by averaging out the regularisation parameter. We let $\xi := 1/\delta^2$, assuming that all parameters are controlled by the same regularisation parameter, which gives the following equation for the prior distribution,

$$p(\omega|\xi) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi/\xi}} \exp \left\{ -\frac{\xi}{2} \omega_j^2 \right\}. \quad (6)$$

To follow the proof given by Cawley and Talbot we further use the following notation, $d := M$, $Z_\Omega := (2\pi/\xi)^{d/2}$, and $\Omega := \sum_j \omega_j^2$, which yields

$$p(\omega|\xi) = Z_\Omega(\xi)^{-1} \exp \left\{ -\frac{\xi}{2} \Omega(\omega) \right\}. \quad (7)$$

We can then integrate out the regularisation parameter by computing the marginal prior distribution,

$$P(\omega) = \int p(\omega|\xi) p(\xi) d\xi. \quad (8)$$

This expression can be solved analytically with the right choice of prior. Cawley and Talbot use $p(\xi) \propto 1/\xi$, but instead we choose $p(\xi) \propto 1/\xi \exp\{-\xi/2\}$. By assuming positive support for ξ and expanding out the marginal prior distribution we get,

$$\begin{aligned}
P(\omega) &= \int_0^\infty \left(\frac{2\pi}{\xi}\right)^{-d/2} \exp\left\{-\frac{\xi}{2}\Omega(\omega)\right\} \frac{1}{\xi} \exp\left\{-\frac{\xi}{2}\right\} d\xi \\
&= (2\pi)^{-d/2} \int_0^\infty \xi^{d/2-1} \exp\left\{-\frac{\xi}{2}(\Omega(\omega)+1)\right\} d\xi
\end{aligned} \tag{9}$$

We use the gamma integral, $\int_0^\infty x^{\nu-1} \exp\{-\mu x\} = \Gamma(\nu)/\mu^\nu$ (Cawley and Talbot 2007), to obtain,

$$P(\omega) = (2\pi)^{-d/2} \frac{\Gamma(d/2)}{(1/2)^{d/2} (\Omega(\omega)+1)^{d/2}} \tag{10}$$

Finally, by applying the logarithm and discarding constants, we obtain the final expression for the marginal prior distribution,

$$\begin{aligned}
\log P(\omega) &\propto -\frac{d}{2} \log(\Omega(\omega)+1) \\
&\propto -\frac{M}{2} \log\left(1 + \sum_j^M \omega_j^2\right)
\end{aligned} \tag{11}$$

The same procedure can be applied to derive the marginal likelihood, by noting $d := N$, $\xi := 1/\sigma^2$, $Z_\Omega := (2\pi/\xi)^{d/2}$, $\Omega(\omega) := \sum_i e_i^2$.

B Complementary results benchmark analysis

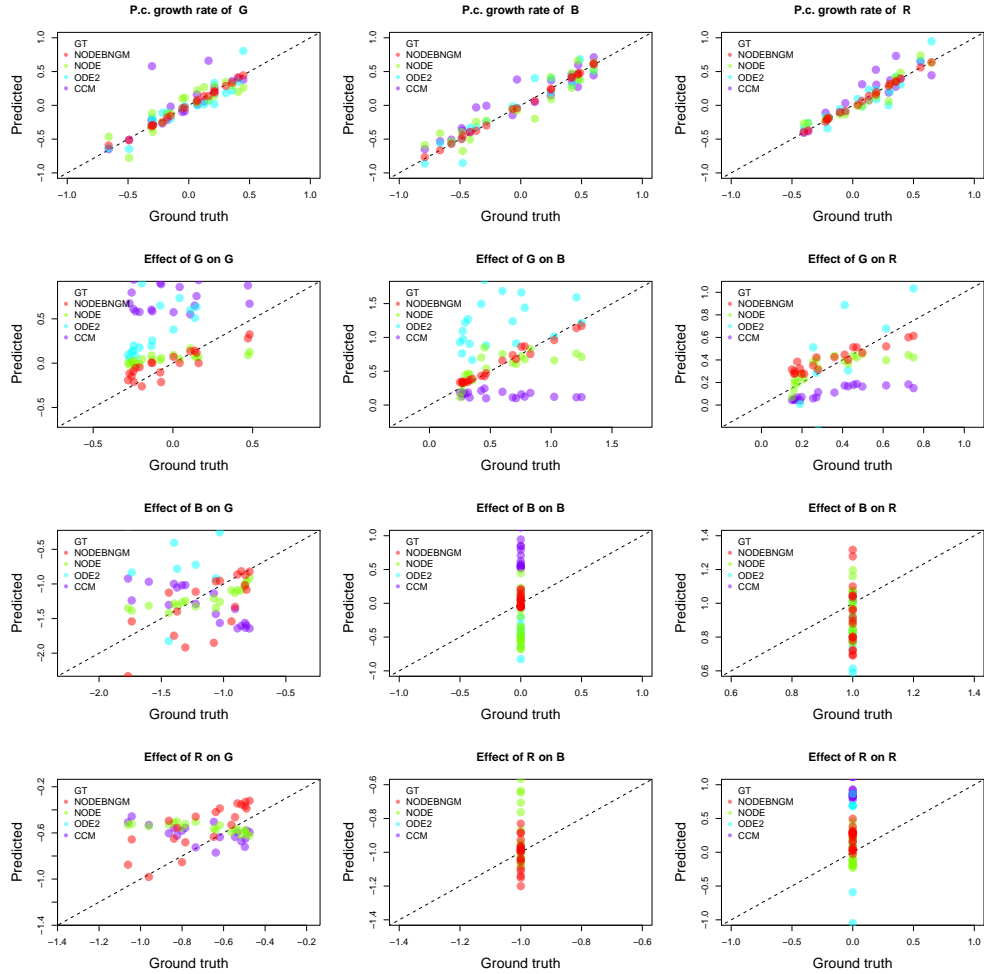


Figure S1: Train set accuracy of predicted per-capita growth rates and effects estimated by NODEBNGM, standard NODEs, ODE2, and CCM. The NODEBNGM method (nonparametric) involves fitting a NODE system by Bayesian neural gradient matching (BNGM). The NODE method (nonparametric) involves fitting a NODE system with an ODE solver. The ODE2 method (parametric) involves fitting an ODE system with polynomial functions of species densities with an ODE solver. The CCM method (nonparametric) involves computing locally linear approximations of the state space. For each method, we trained 30 models on the two first thirds of the artificial time series where ground truth is known (Fig. 2). For each plot, the x-axis corresponds to the ground truth, known from the equations that generated the artificial time series, and the y-axis corresponds to the prediction of the best model. Effects are computed as the sensitivity (i.e. derivative) of the per-capita growth rate with respect to each species density G, B, and R, namely the prey, intermediate and top predator.

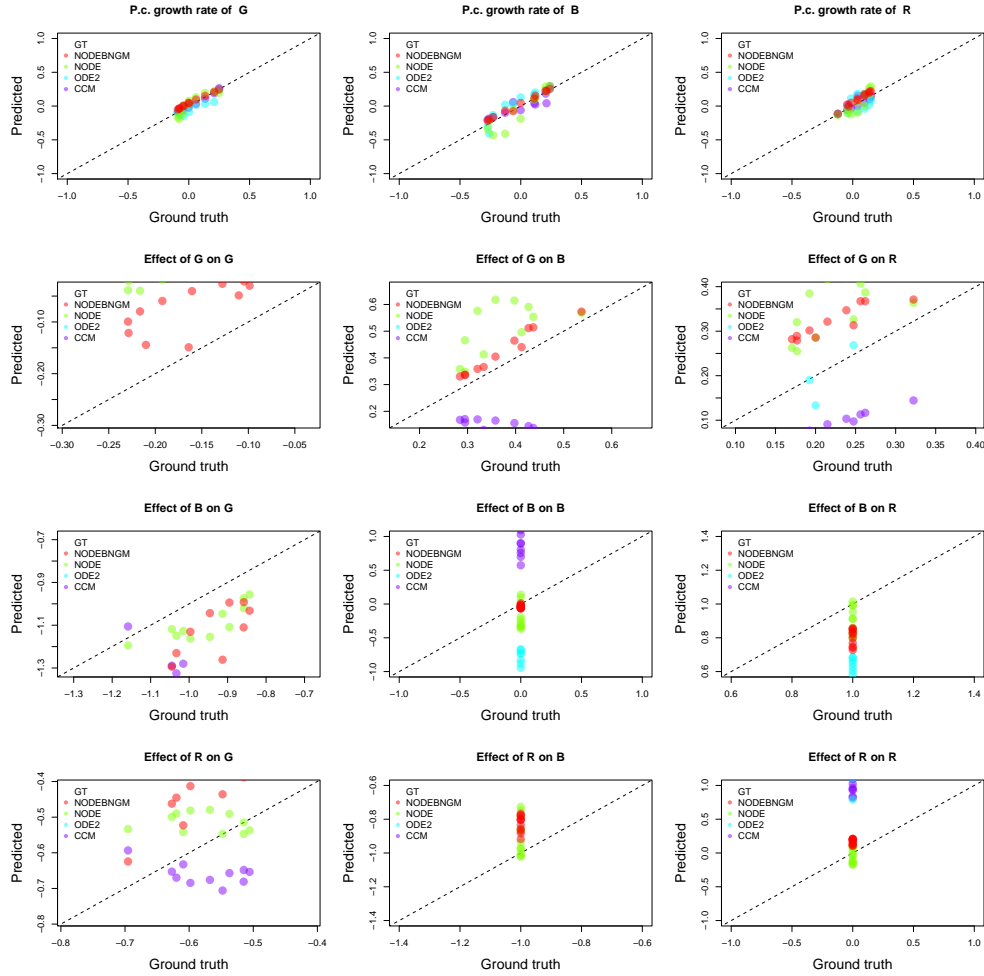


Figure S2: Test set accuracy of predicted per-capita growth rates and effects estimated by NODEBNGM, standard NODEs, ODE2, and CCM. The NODEBNGM method (nonparametric) involves fitting a NODE system by Bayesian neural gradient matching (BNGM). The NODE method (nonparametric) involves fitting a NODE system with an ODE solver. The ODE2 method (parametric) involves fitting an ODE system with polynomial functions of species densities with an ODE solver. The CCM method (nonparametric) involves computing locally linear approximations of the state space. For each method, we trained 30 models on the two first thirds of the artificial time series where ground truth is known (Fig. 2). For each plot, the x-axis corresponds to the ground truth, known from the equations that generated the artificial time series, and the y-axis corresponds to the prediction of the best model. Effects are computed as the sensitivity (i.e. derivative) of the per-capita growth rate with respect to each species density G, B, and R, namely the prey, intermediate and top predator.

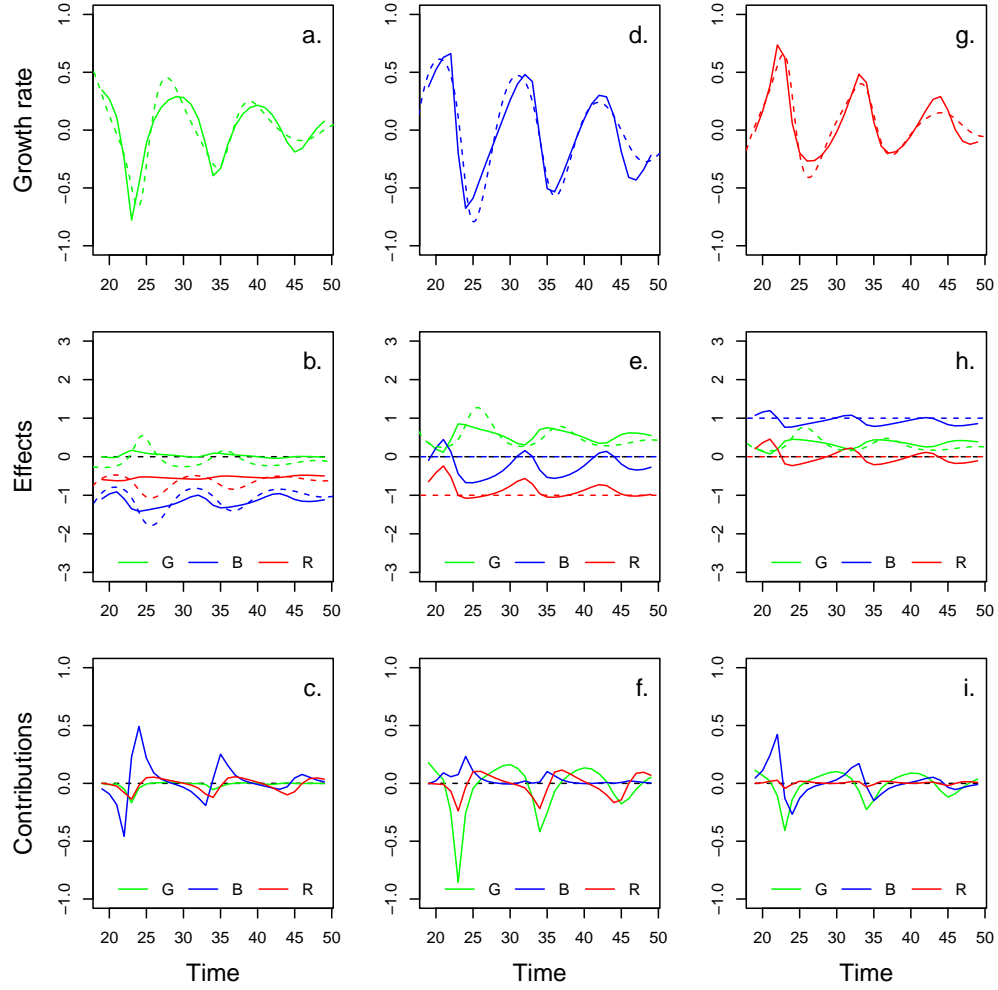


Figure S3: Effects and contributions inferred by standard NODEs. The standard NODE model nonparametrically approximates the per-capita growth rate of the 3 species with an ANN featuring a single layer, 3 input nodes, 10 hidden nodes, 3 outputs. 30 models are fitted to the two first third of the time series using BFGS and a Runge-Kutta ODE solver. The graphs present the predictions obtained for the model that maximises posterior density of the network parameters given the time series. We estimate the direction of ecological interactions (effects) by computing the derivative of the per-capita growth rate approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions) by multiplying the interpolated dynamics of each population with its effects.

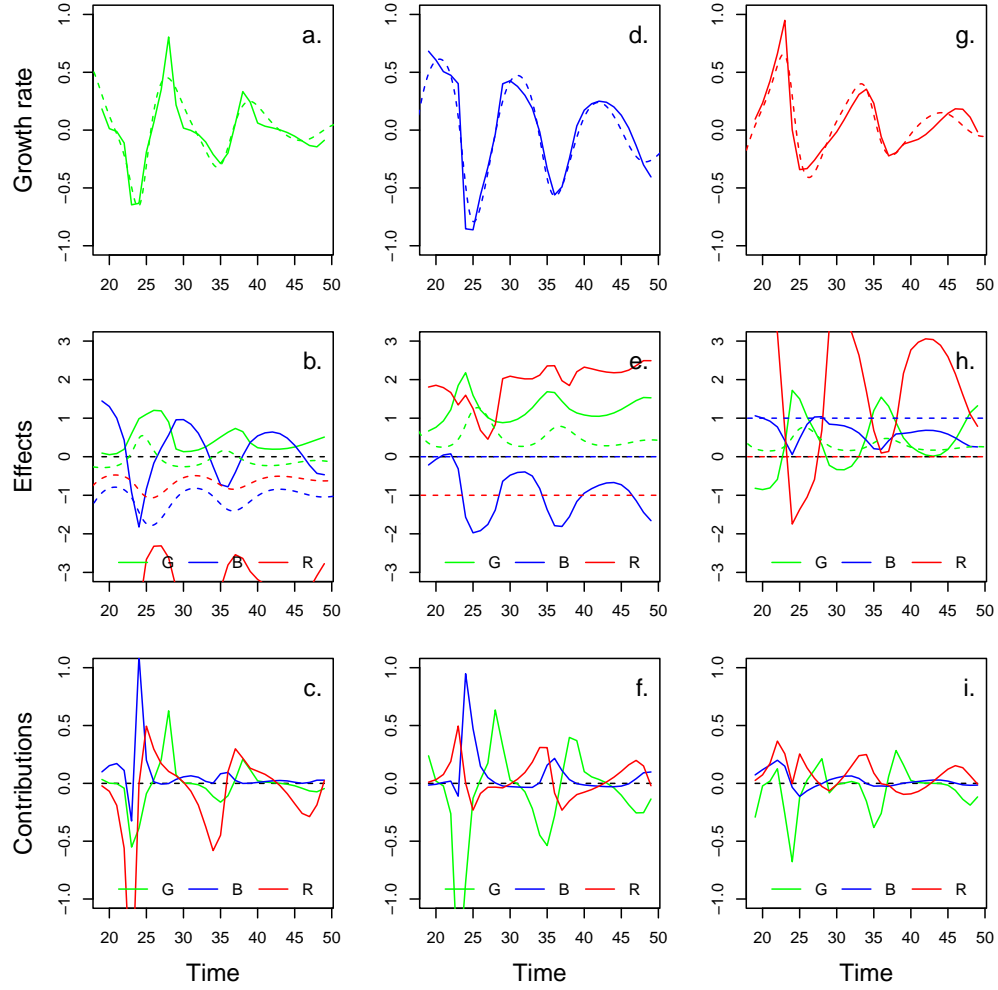


Figure S4: Effects and contributions inferred by parametric ODE. The ODE2 model parametrically approximates the per-capita growth rate of the 3 species with second order polynomial functions. 30 models are fitted to the two first third of the time series using BFGS and a Runge-Kutta ODE solver. The graphs present the predictions obtained for the model that maximises posterior density of the network parameters given the time series. We estimate the direction of ecological interactions (effects) by computing the derivative of the per-capita growth rate approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions) by multiplying the interpolated dynamics of each population with its effects.

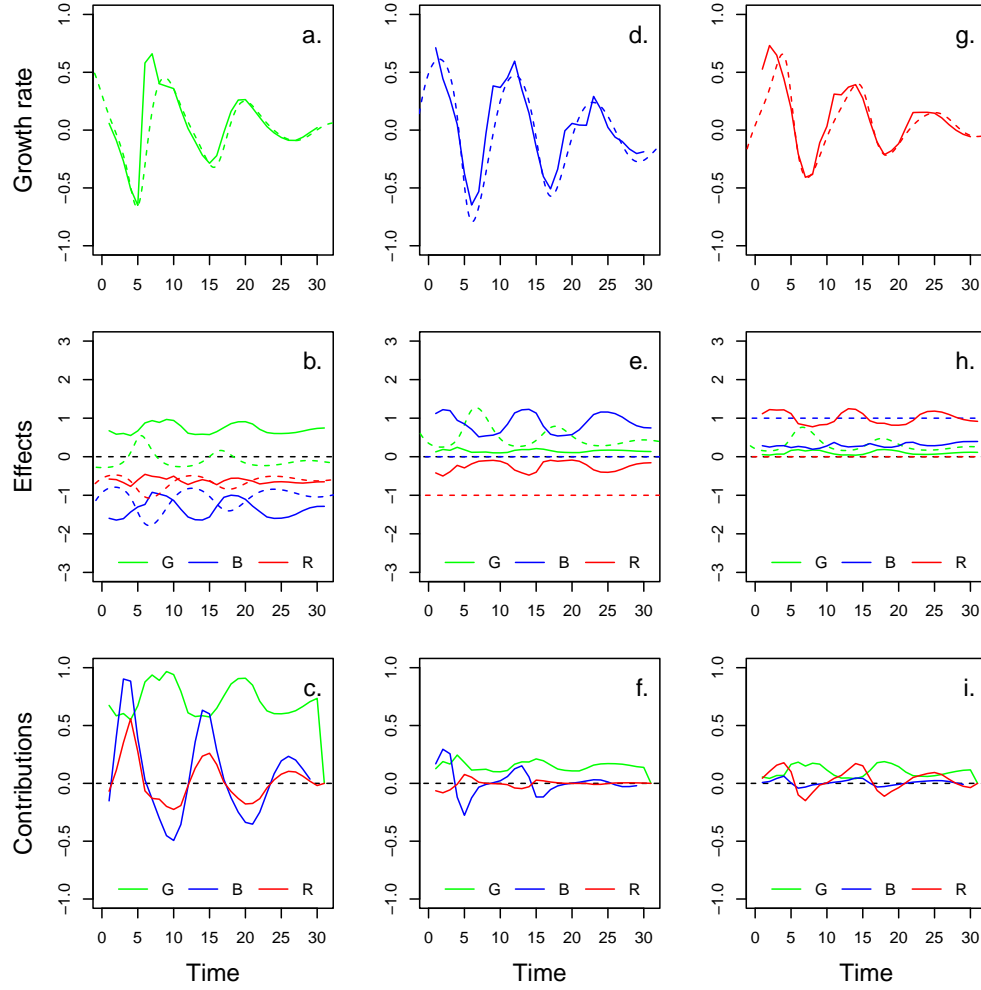


Figure S5: Effects and contributions inferred by CCM. The CCM method nonparametrically approximates the state space from which it derives the sensitivity of population dynamics to a change in the density of the each species. We use the rEDM implementation and derived our code from the three species example provided in the package tutorial (v1.13.1, Sugihara et al. 2012). We calculated the dynamics and per capita growth rate using finite differences, as the standard library does not readily provide estimates. The effects correspond to the s-map coefficients. Finally, we compute the strength of ecological interactions (contributions) by multiplying the interpolated dynamics of each population with its effects.

C Complementary results case study 2 replicate A

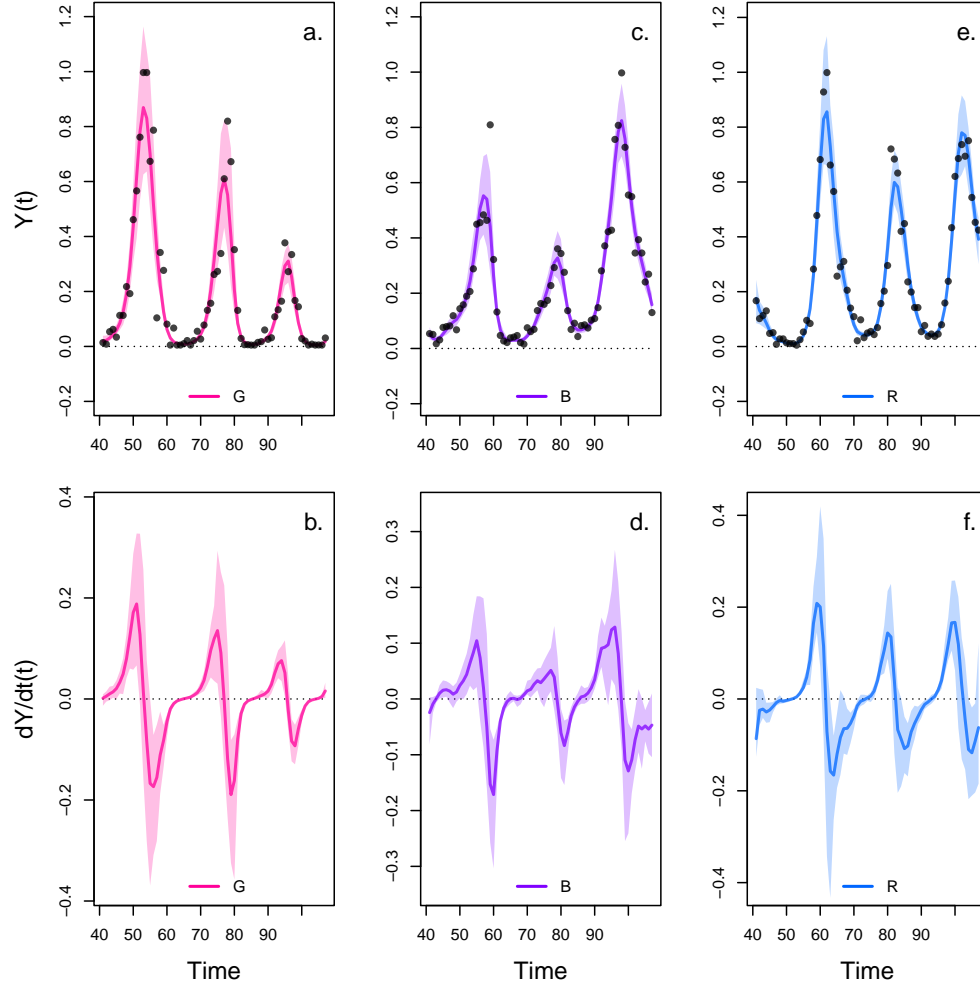


Figure S6: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate A. Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from digitising the time series in Hiltunen et al. 2013.

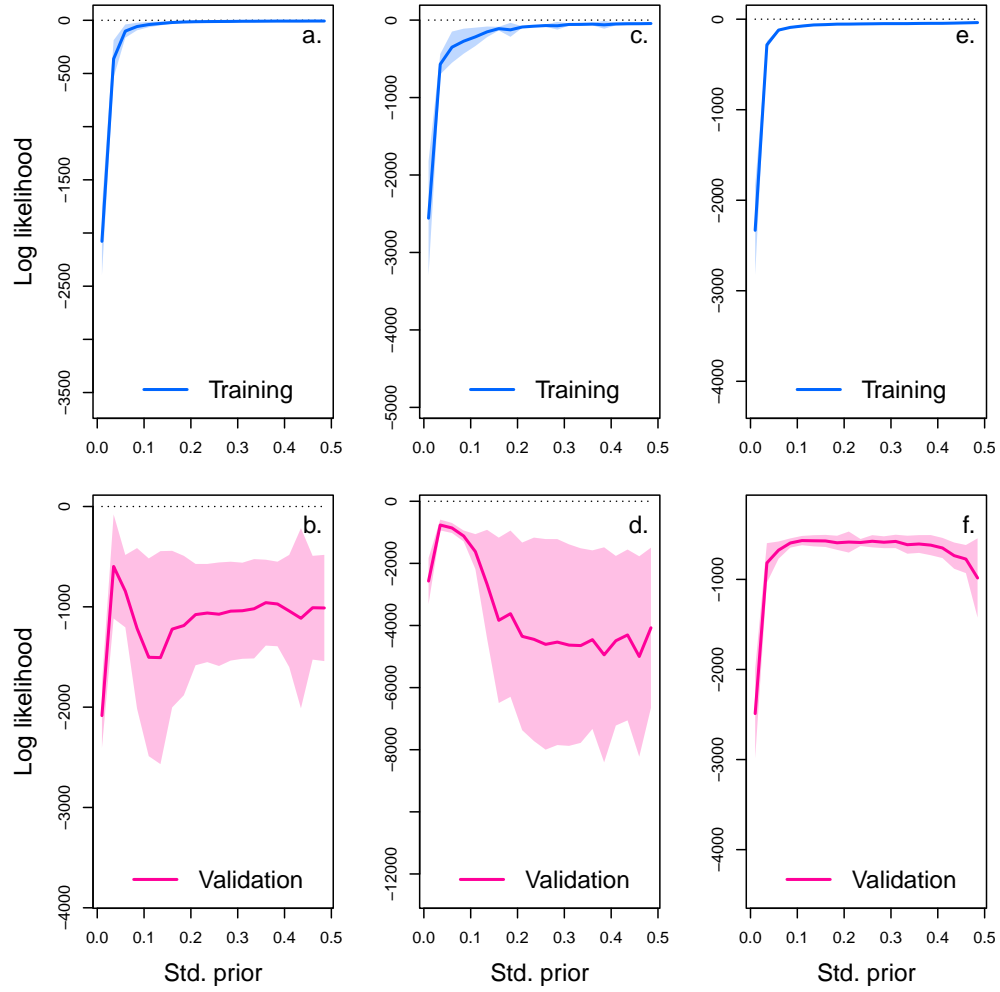


Figure S7: Cross-validation plot of the NODE analysis of replicate A. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the validation set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

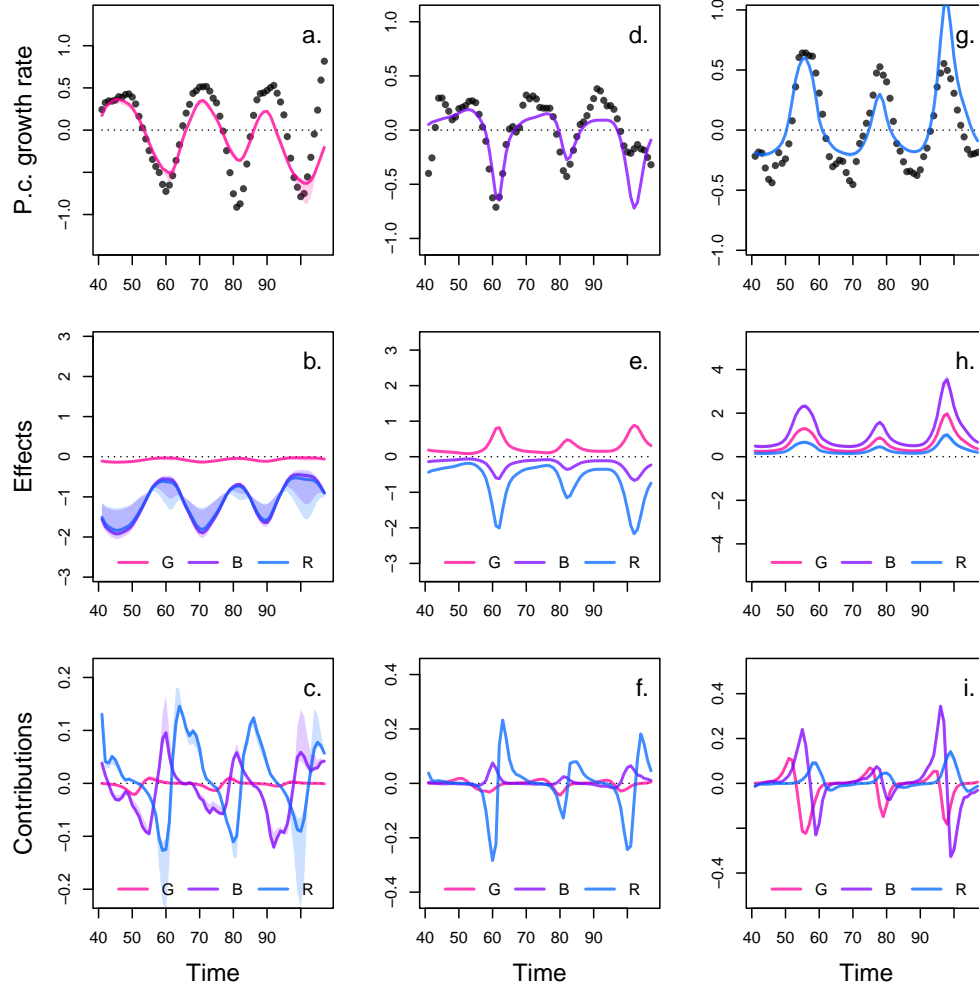


Figure S8: Drivers of dynamics of algae, flagellate, and rotifer in replicate A. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

D Complementary results case study 2 replicate B

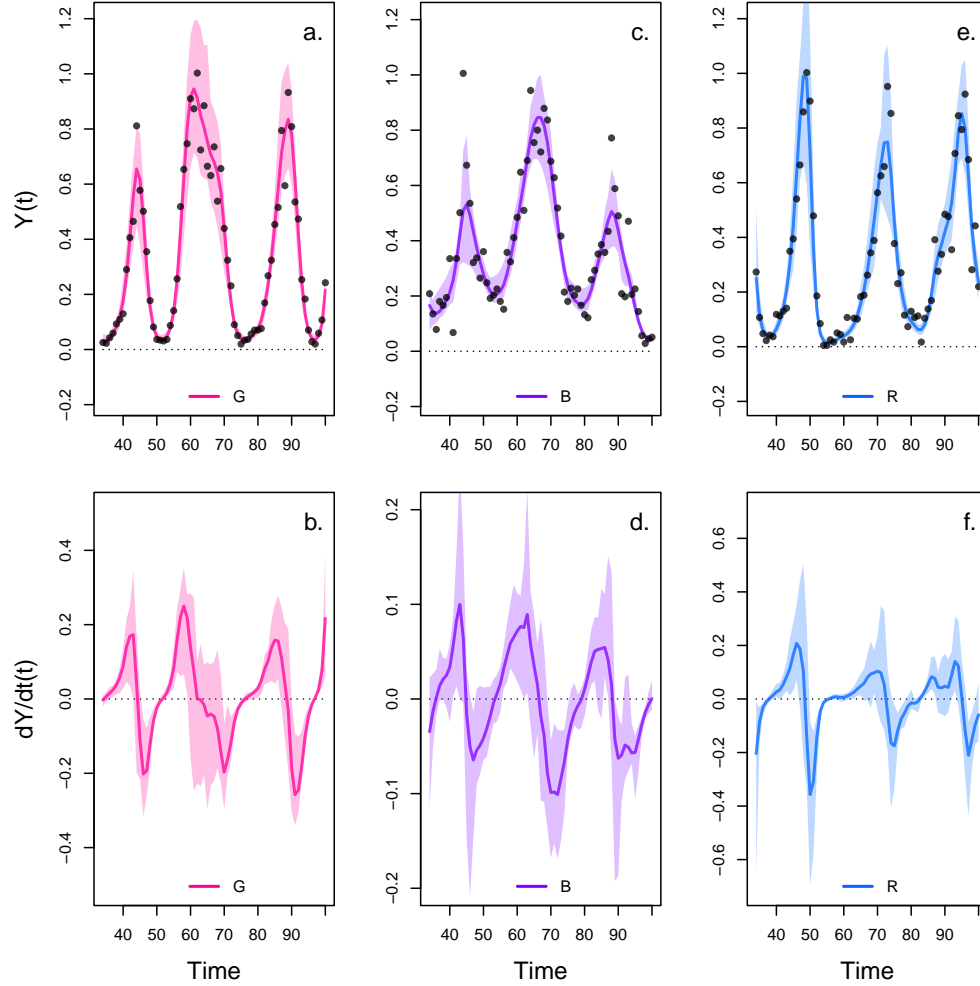


Figure S9: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate B. Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from digitising the time series in Hiltunen et al. 2013.

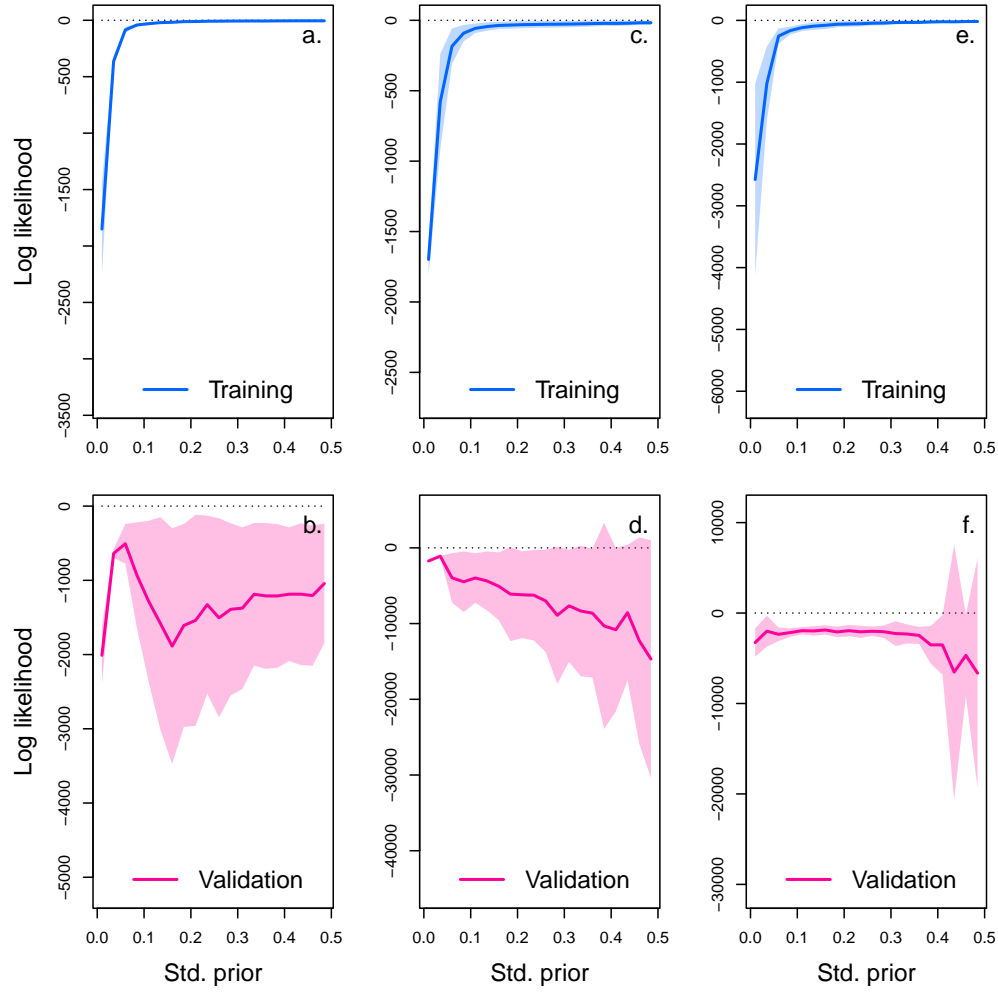


Figure S10: Cross-validation plot of the NODE analysis of replicate B. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the validation set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

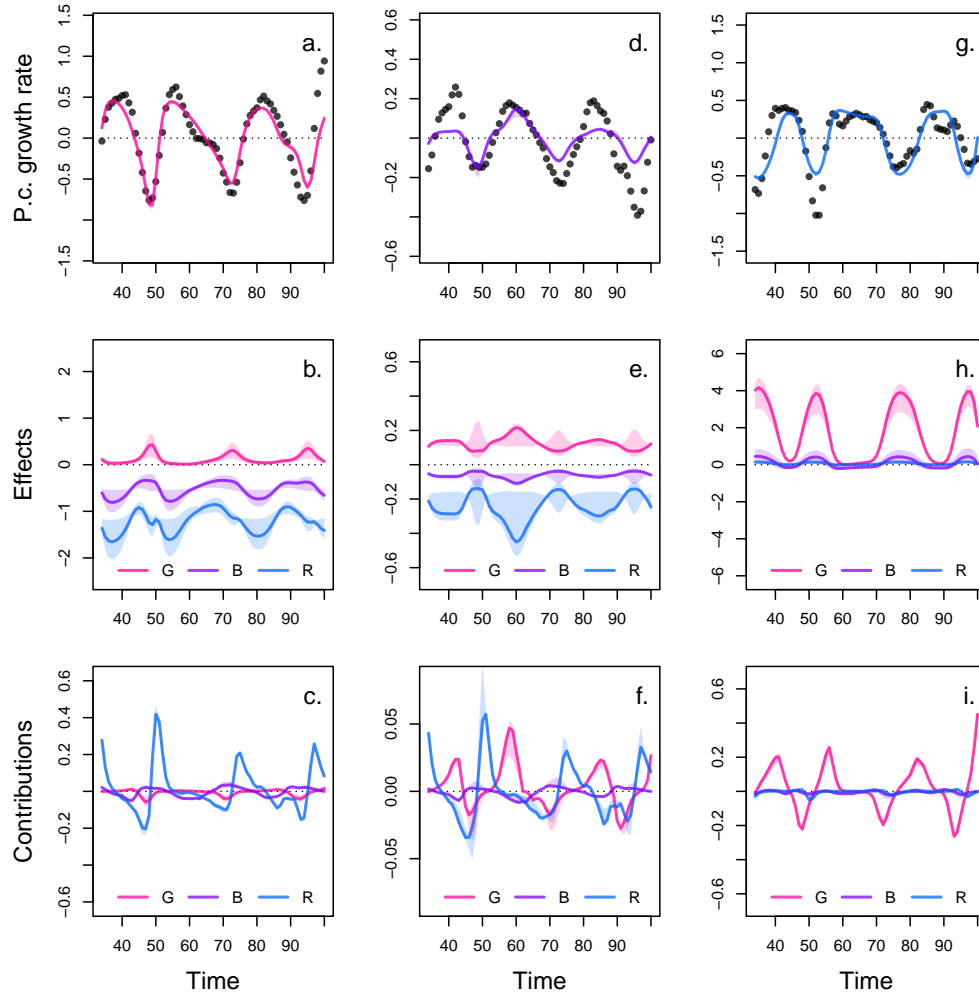


Figure S11: Drivers of dynamics of algae, flagellate, and rotifer in replicate B. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

E Complementary results case study 2 replicate C

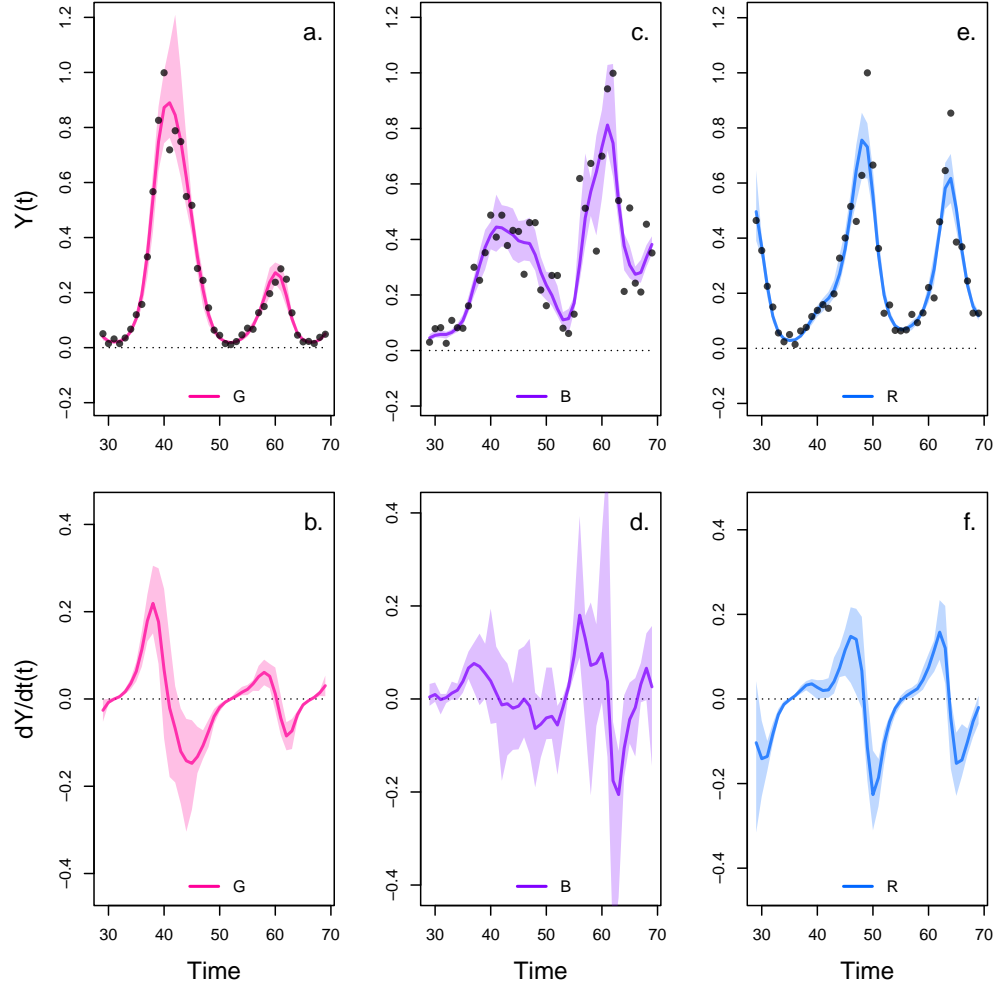


Figure S12: Interpolation of state and dynamics of algae, flagellate, and rotifer density in replicate B. Graph a., c., and e. display the neural interpolation of the population density of algae (G), flagellate (B), and rotifer (R), respectively (obtained with Eq. 7). Graph b., d., and f. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from digitising the time series in Hiltunen et al. 2013.

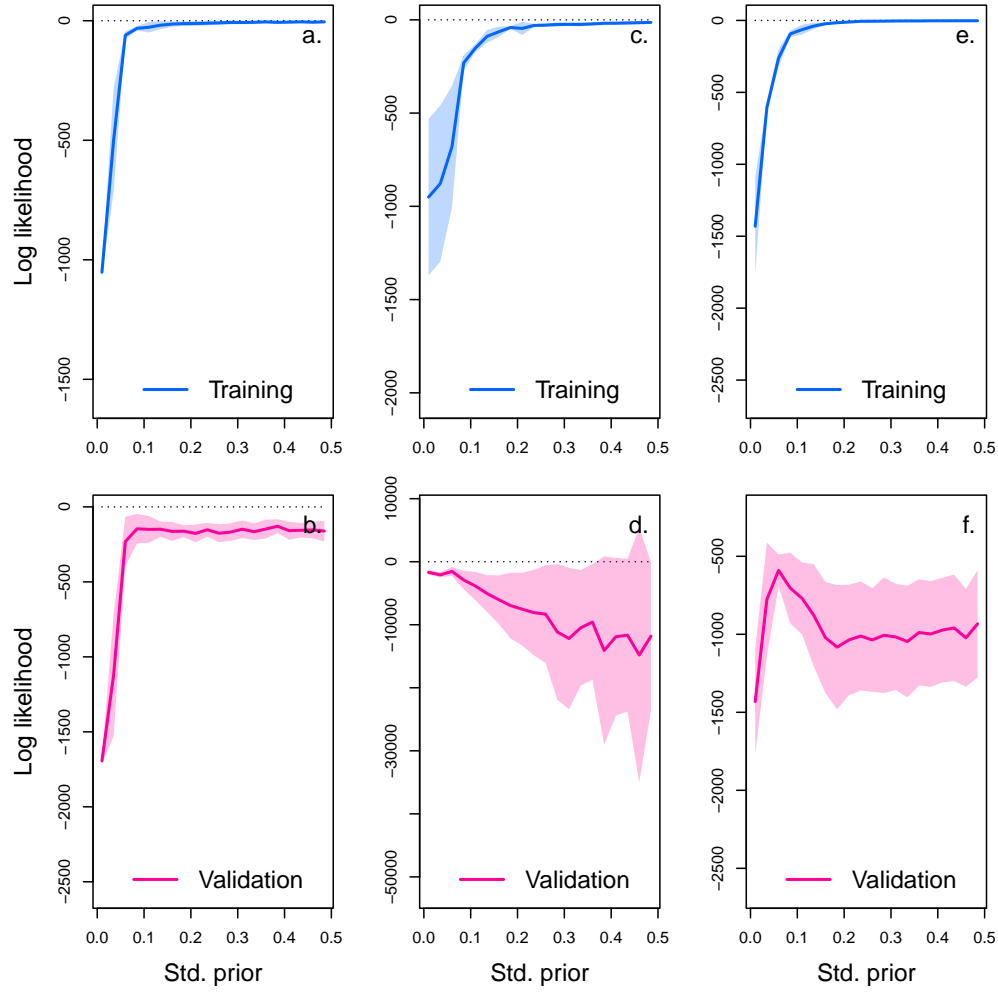


Figure S13: Cross-validation plot of the NODE analysis of replicate C. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.5) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set for increasing value of standard deviation (from 0.05 to 0.5 by 0.05 increments), and evaluated on the validation set. Graph a., c., and e. show the log likelihood of the NODE system fitted by BNGM to the train set of algae, flagellate, and rotifer, respectively. Graph b., d., and f. show the log likelihood of the fitted NODE, evaluated on the corresponding validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

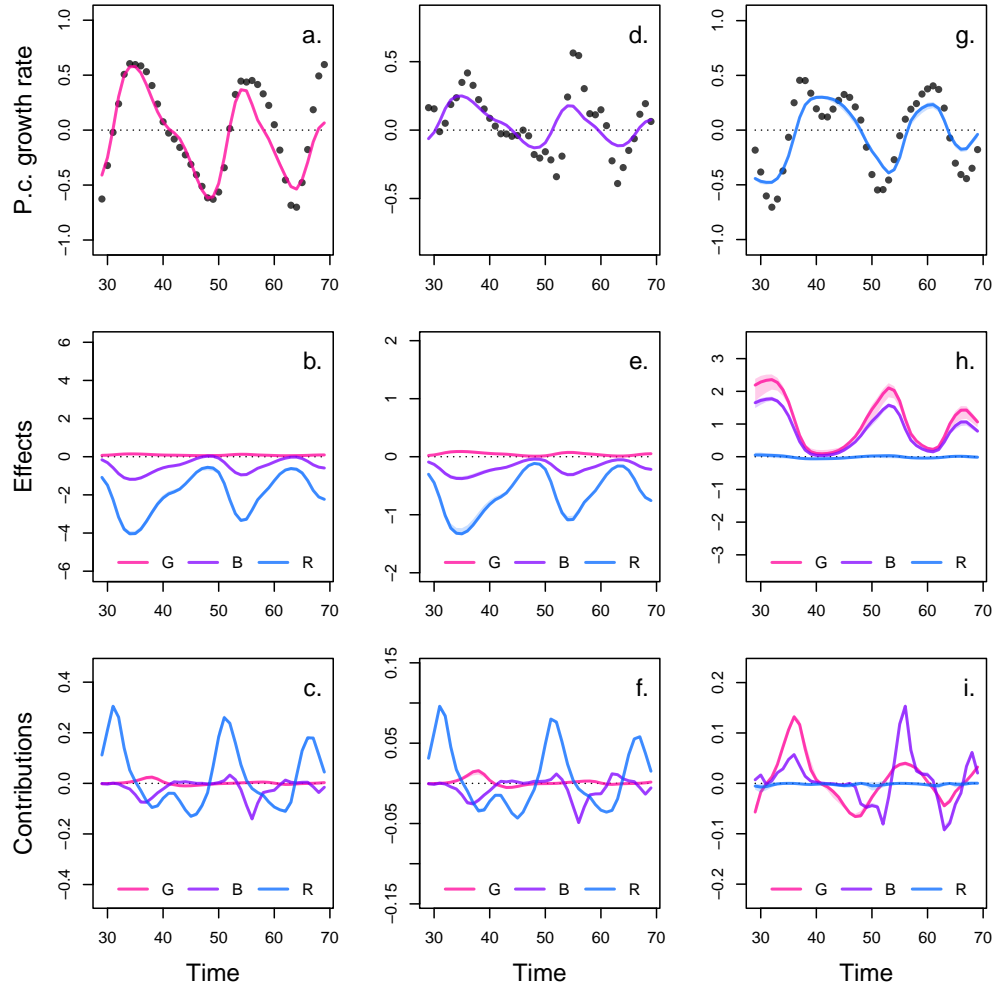


Figure S14: Drivers of dynamics of algae, flagellate, and rotifer in replicate C. This figure displays the NODE nonparametric approximations of the per-capita growth rate of algae (a., b., c.), flagellate (d., e., f.), and rotifer (g., h., i.). We obtain the NODE approximations (a., d., g., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e., h.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f., i.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions. The replicated time series were obtained by digitising the time series in Hiltunen et al. (2013).

F Complementary results case study 3

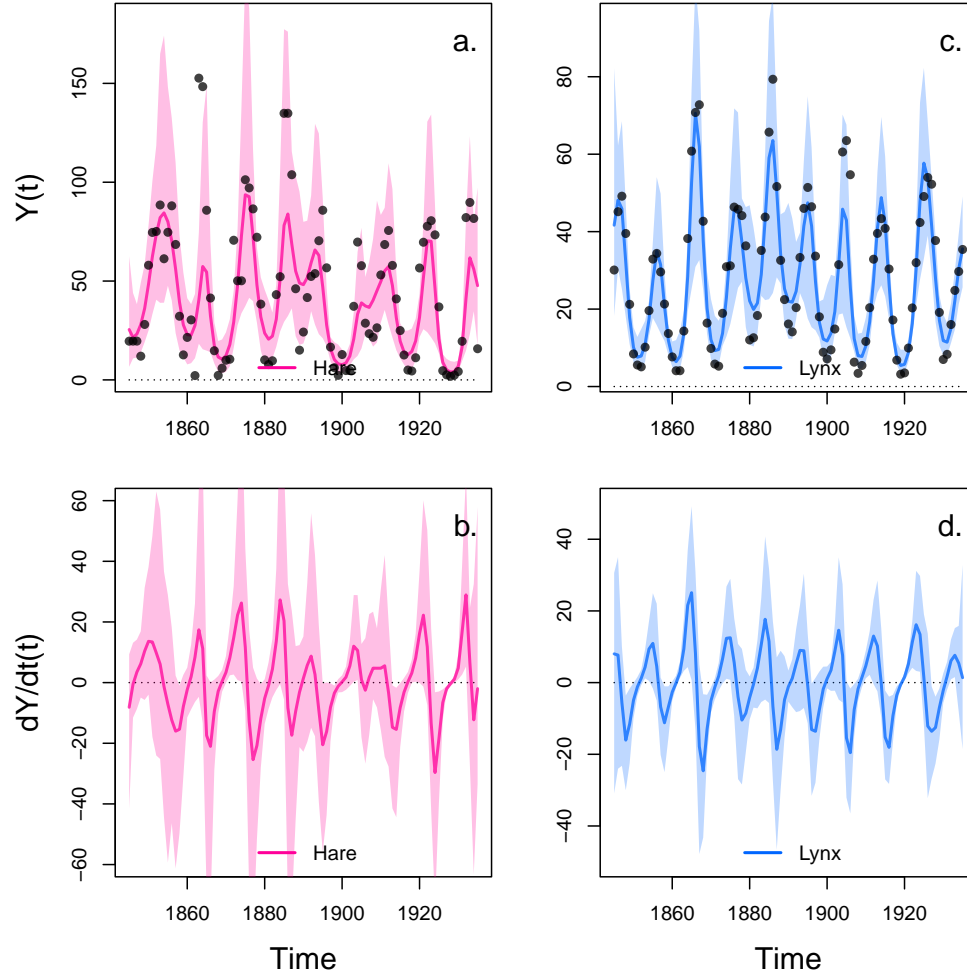


Figure S15: Interpolation of state and dynamics of hare and lynx. Graph a. and c. display the neural interpolation of the population density of hare and lynx respectively (obtained with Eq. 7). Graph b. and d. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7) (Pearce et al. 2018). Time series are obtained from Bonnaffé, Sheldon, and Coulson 2021, originally from Odum and Barrett 1972.

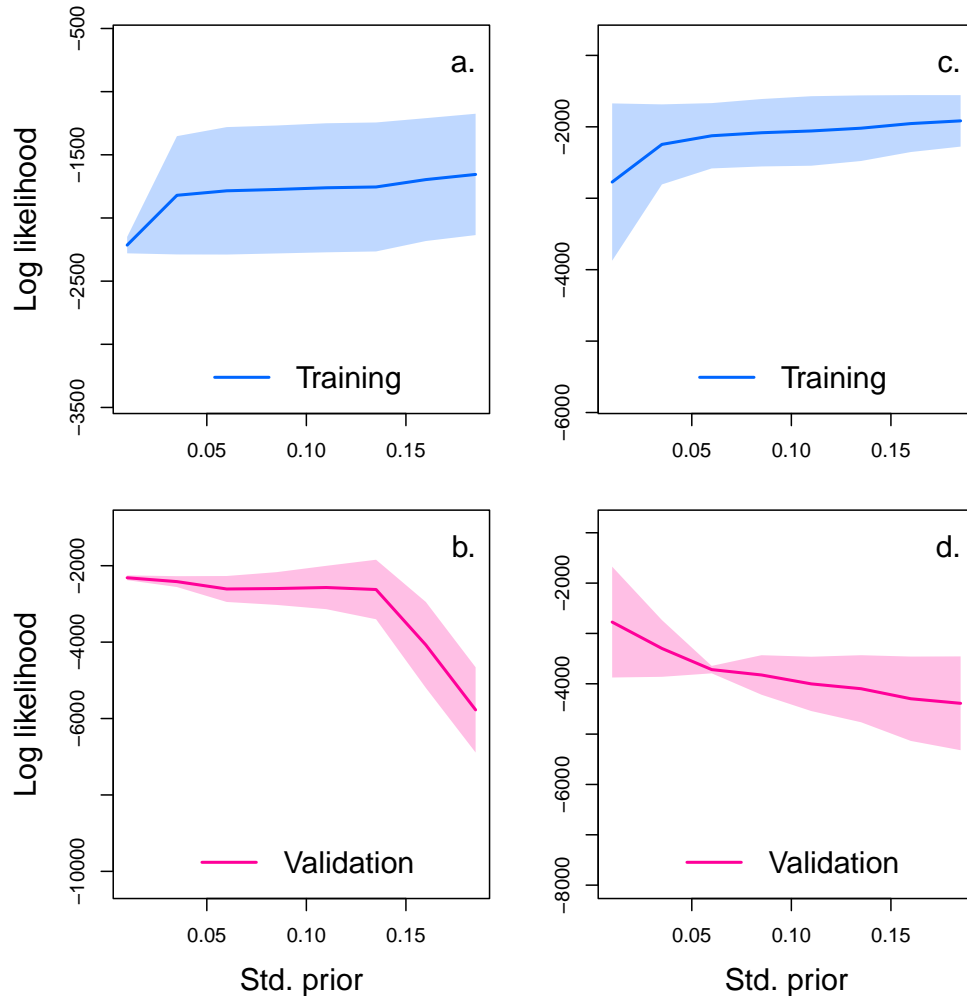


Figure S16: Cross-validation plot of the NODE analysis of the hare-lynx system. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values (towards 0.2) correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set for increasing value of standard deviation (from 0.0 to 0.2 by 0.025 increments), and evaluated on the validation set. Graph a. and c. show the log likelihood of the NODE system fitted by BNGM to the train set of hare and lynx, respectively. Graph b. and d. show the log likelihood of the fitted NODE, evaluated on the corresponding validation set. The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

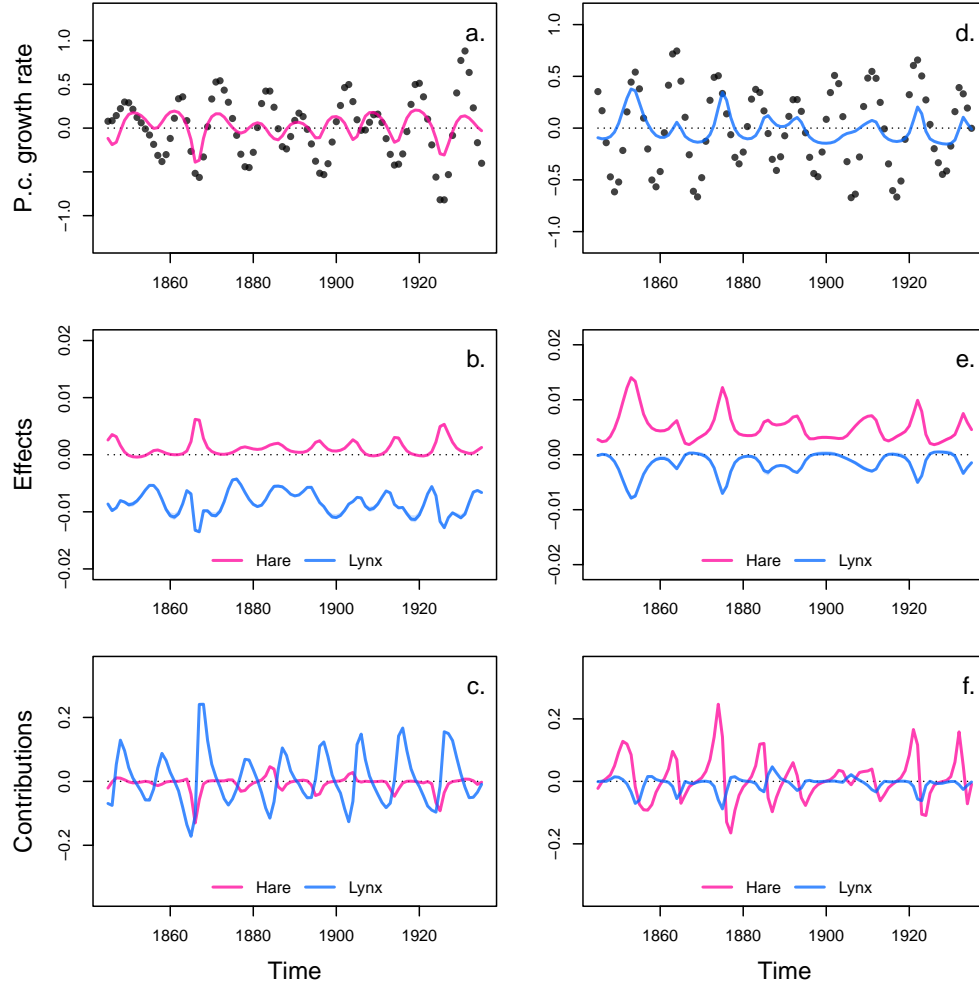


Figure S17: Drivers of dynamics of hare and lynx in the Odum and Barrett pelt count time series. This figure displays the NODE nonparametric approximations of the per-capita growth rate of hare (a., b., c.), and lynx (d., e., f.). We obtain the NODE approximations (a., d., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, b., e.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, c., f.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

G Complementary results case study 4

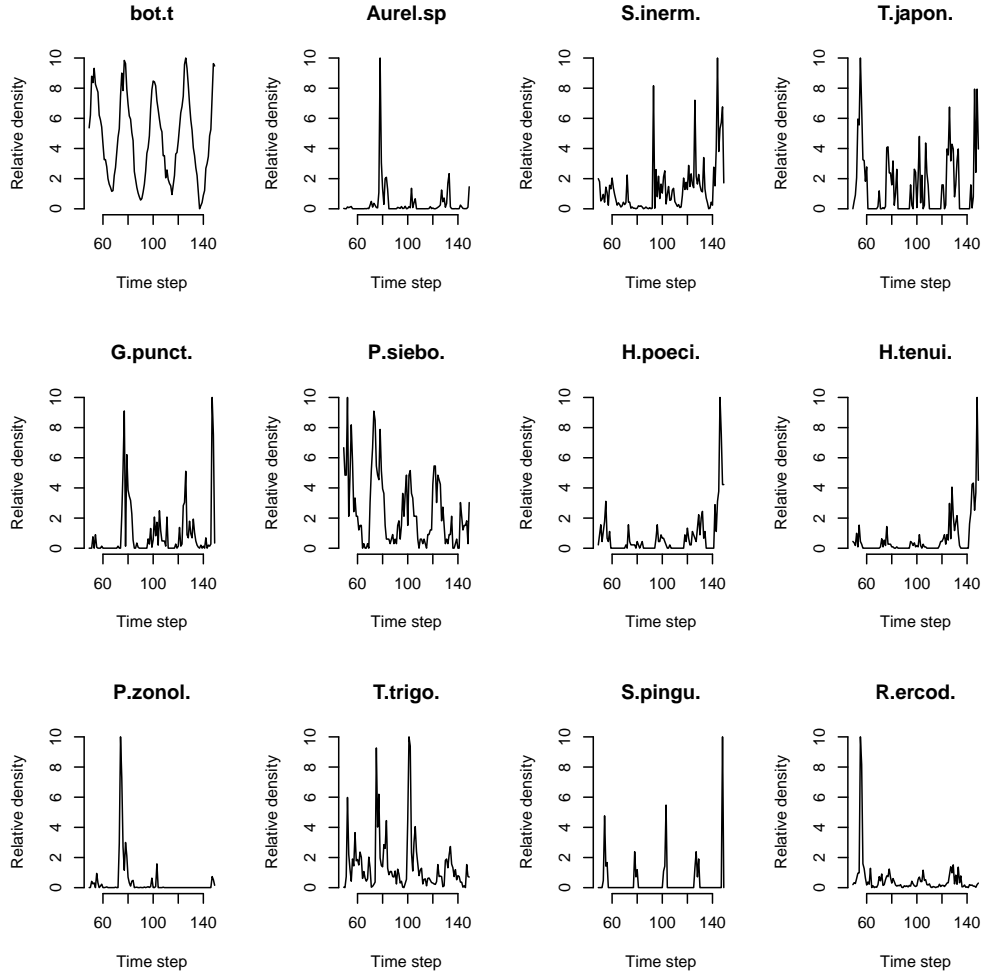


Figure S18: Time series of the aquatic community in Ushio et al. 2018. The time series was collected for 12-years on a fortnightly basis, for 15 dominant species in the Maizuru bay in Japan. We focus on the 11 species and the 100 time steps with the least sparse abundance records. Bot.t corresponds to water temperature near the bottom. The main species are *Aurelia sp.*, *Sebastes inermis*, *Trachurus japonicus*, *Girella punctata*, *Pseudolabrus sieboldi*, *Halichoeres poecilopterus*, *Halichoeres tenuispinnis*, *Pterogobius zonoleucus*, *Tridentiger trigonocephalus*, *Sphyrna pinguis*, and *Rudarius ercodes*. Variables are normalised with respect to minimum and maximum values.

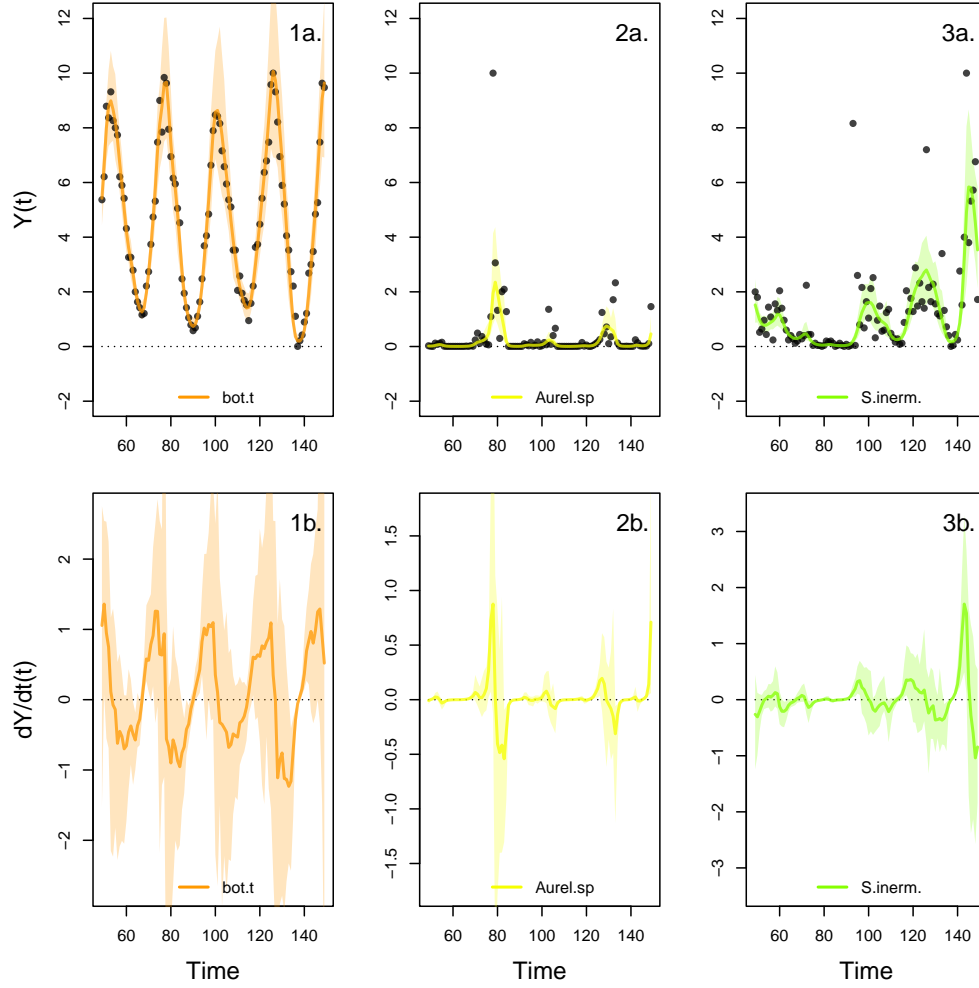


Figure S19: Interpolation of state and dynamics of temperature and species abundance in the Maizuru bay community. Graphs a. display the neural interpolations of the population density (obtained with Eq. 7), apart from graph 1a. which corresponds to sea bottom temperature. Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7).

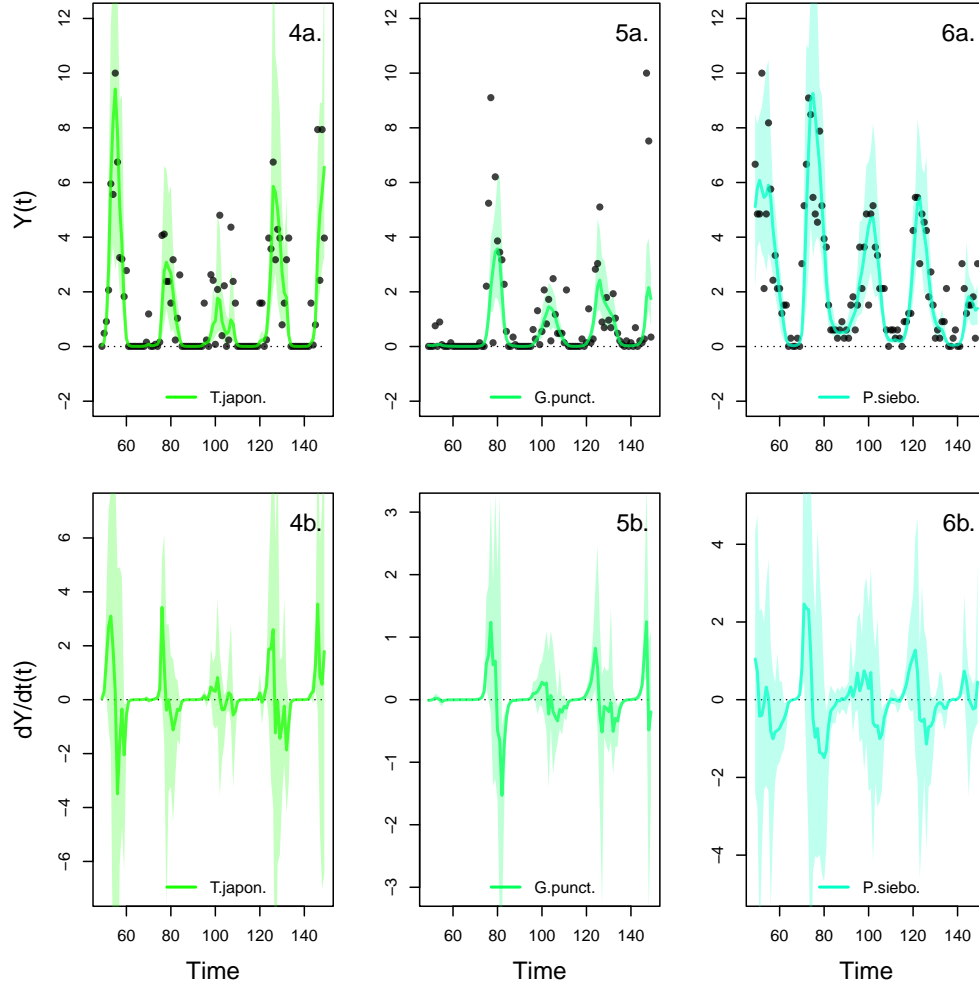


Figure S20: Interpolation of state and dynamics of species abundance in the Maizuru bay community. Graphs a. display the neural interpolations of the population density (obtained with Eq. 7). Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7).

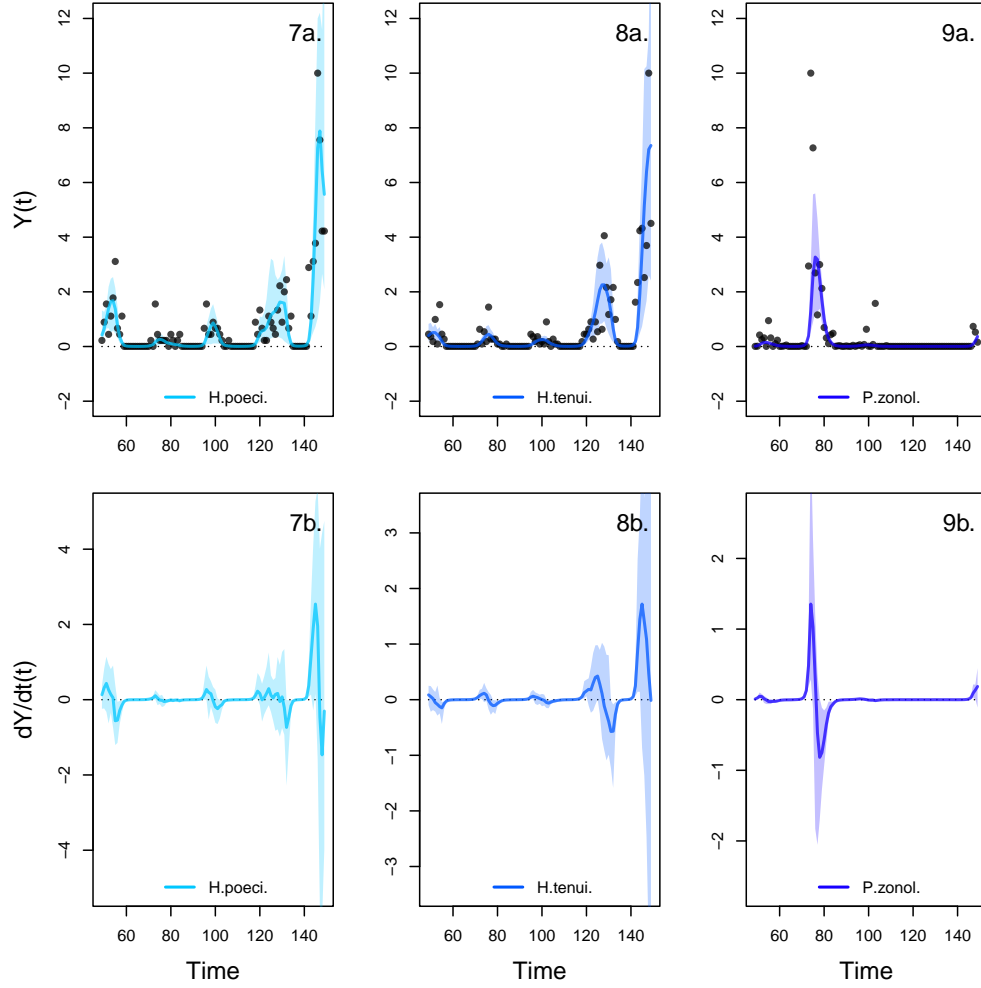


Figure S21: Interpolation of state and dynamics of species abundance in the Maizuru bay community. Graphs a. display the neural interpolations of the population density (obtained with Eq. 7). Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7).

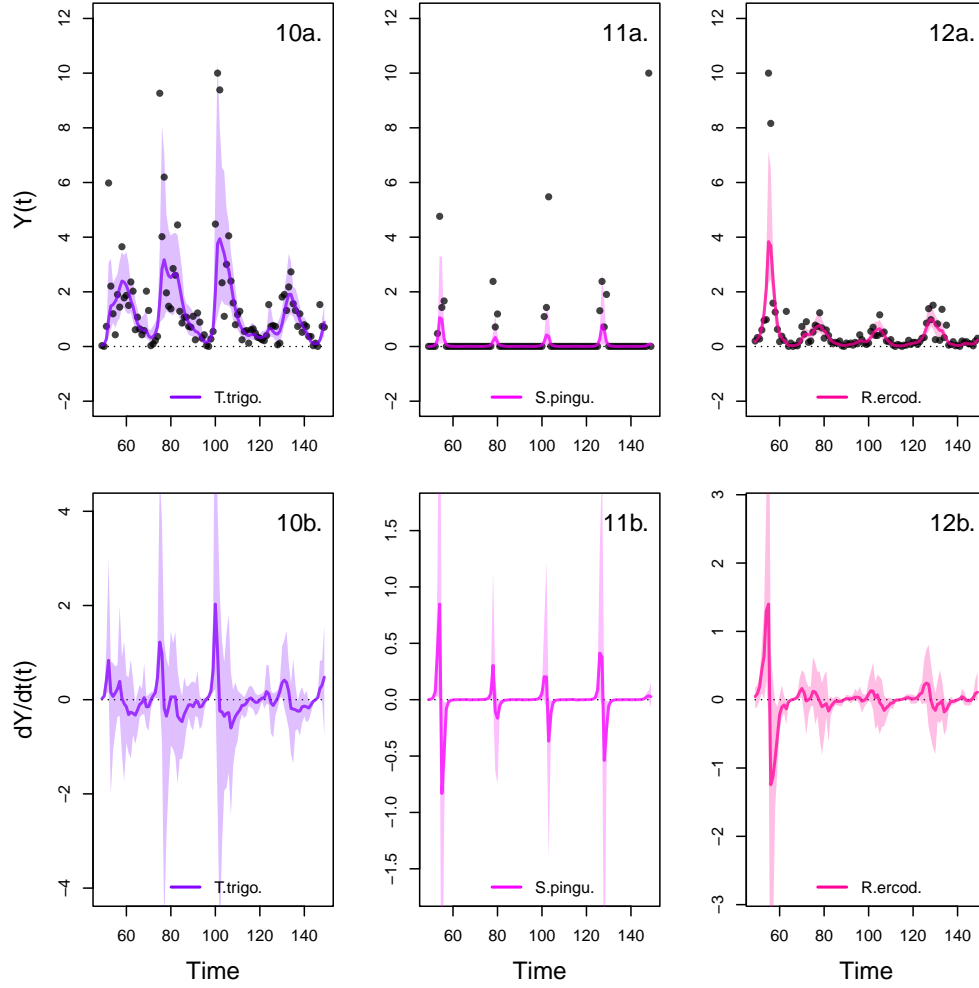


Figure S22: Interpolation of state and dynamics of species abundance in the Maizuru bay community. Graphs a. display the neural interpolations of the population density (obtained with Eq. 7). Graphs b. show the corresponding interpolated dynamics, obtained by differentiating the interpolation of the states with respect to time (Eq. 5). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log marginal posterior distribution (Eq. 7).

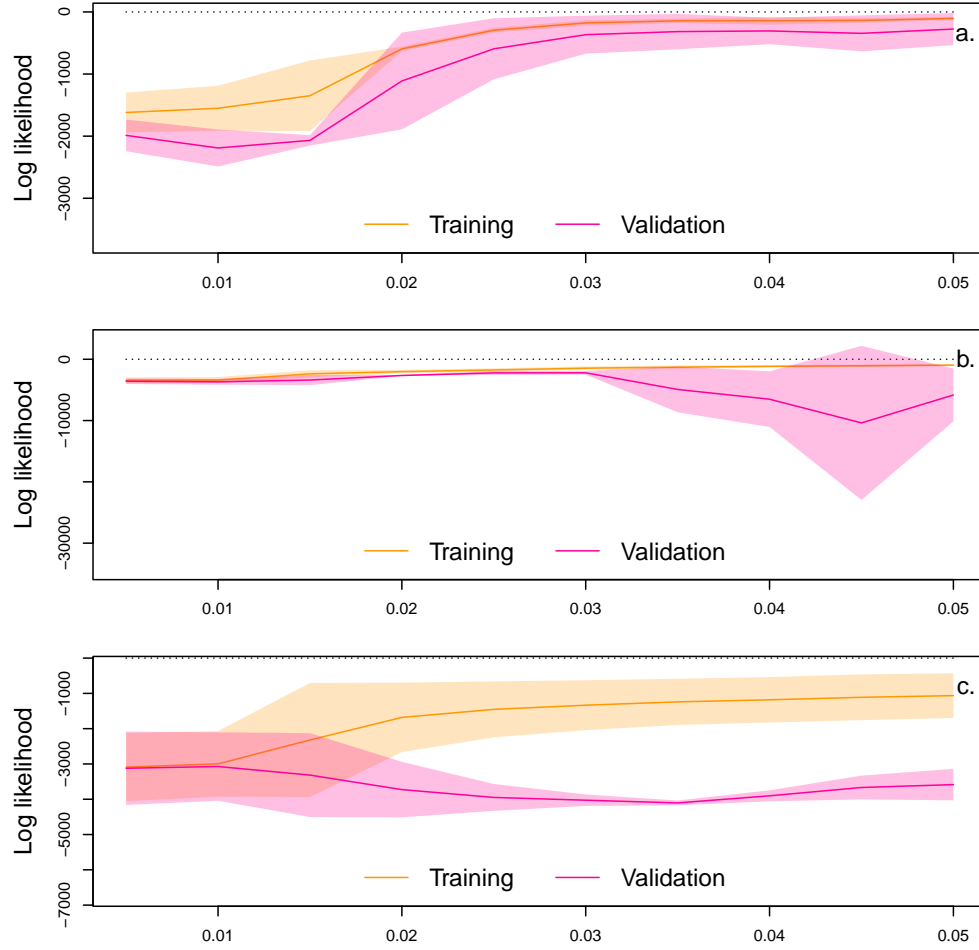


Figure S23: Cross-validation plot of the NODE analysis of the Maizuru bay community. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

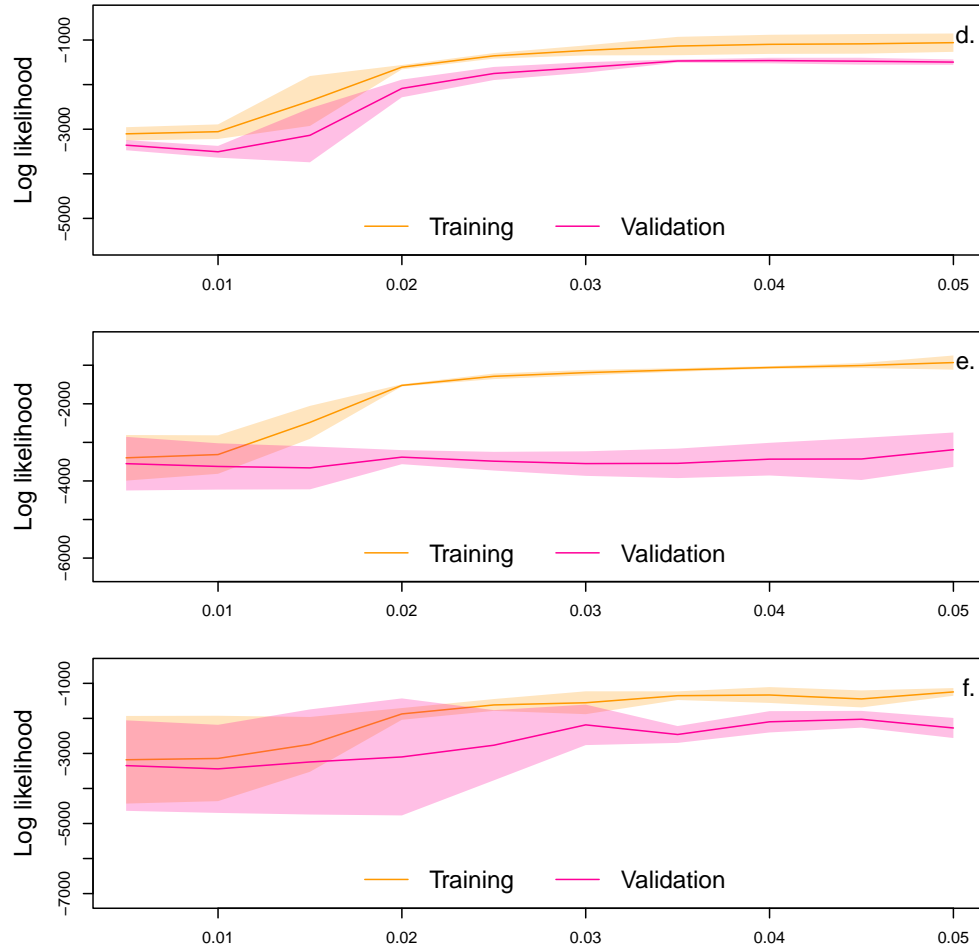


Figure S24: Cross-validation plot of the NODE analysis of the Maizuru bay community. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

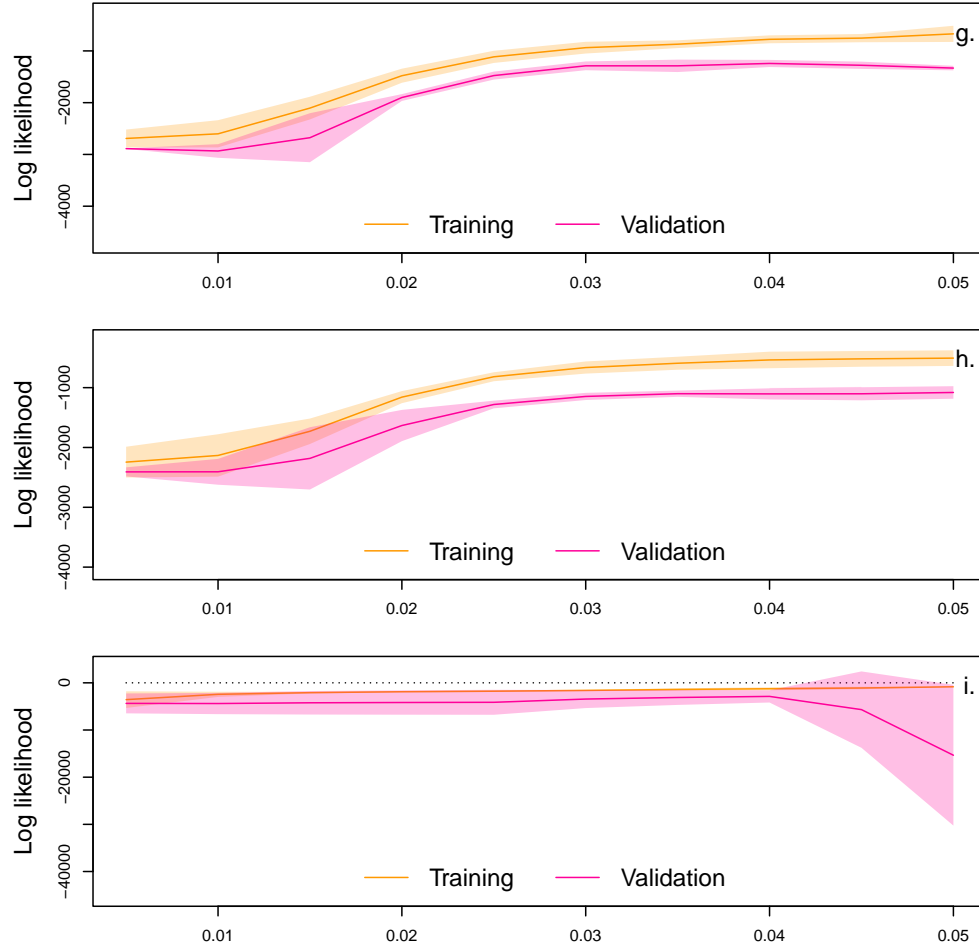


Figure S25: Cross-validation plot of the NODE analysis of the Maizuru bay community. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

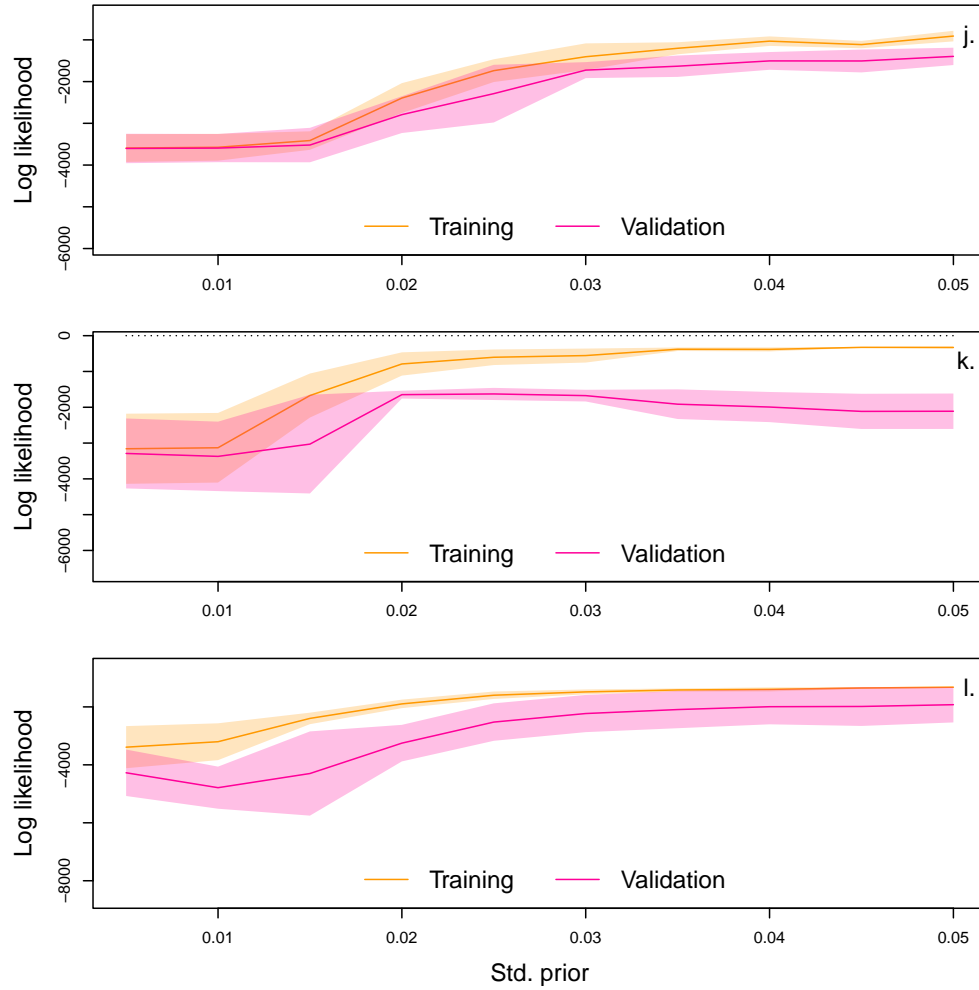


Figure S26: Cross-validation plot of the NODE analysis of the Maizuru bay community. The x-axis of the graphs correspond to the standard deviation of the prior distribution of the NODE parameters, which constrains the nonlinearity of the nonparametric approximation of the NODEs. Small values of standard deviation correspond to a linear model, while higher values correspond to a highly nonlinear model. Time series are split in three thirds to create a train, validation, and test set. The model is fitted to the train set (i.e. first third) for increasing value of standard deviation (from 0.005 to 0.05 by 0.005 increments), and evaluated on the validation set. The operation is repeated by swapping the training and validation set. The graphs show the log likelihood of the NODE system fitted by BNGM to the train set (in orange), and evaluated on the corresponding validation set (in red). The shaded areas represent the 90% confidence interval on estimates, obtained by anchored ensembling of the log posterior distribution (Eq. 8) (Pearce et al. 2018).

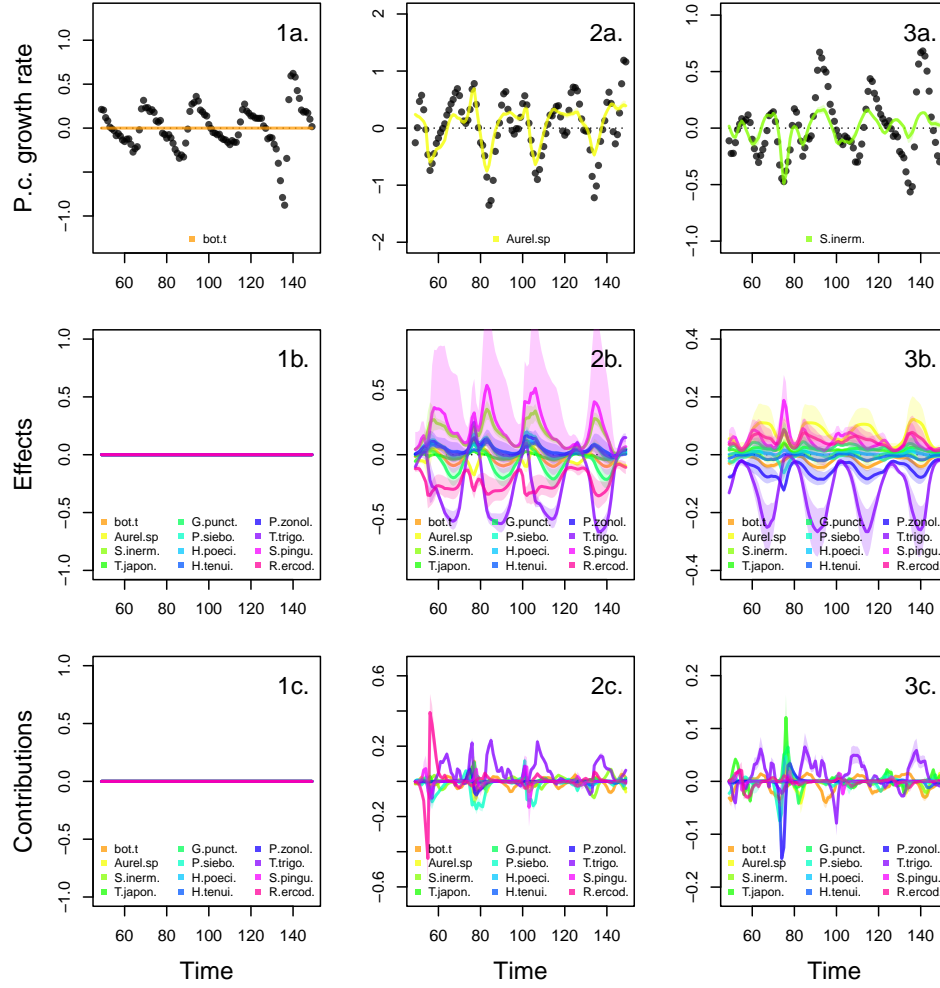


Figure S27: Drivers of dynamics of species abundance in the Maizuru bay community. This figure displays the NODE nonparametric approximations of the per-capita growth rates (2-3a.). We obtain the NODE approximations (2-3a., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, 2-3b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 2-3c.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

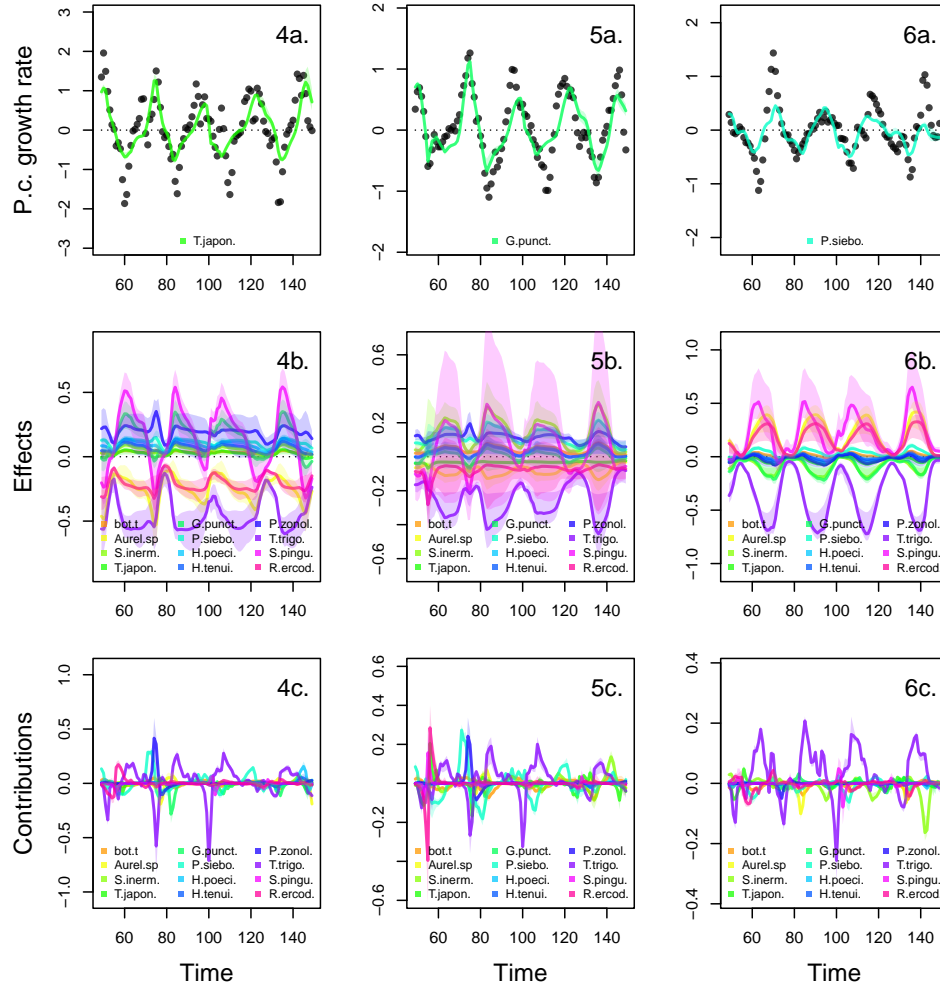


Figure S28: Drivers of dynamics of species abundance in the Maizuru bay community. This figure displays the NODE nonparametric approximations of the per-capita growth rates (4-6a.). We obtain the NODE approximations (4-6a., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, 4-6b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 4-6c.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

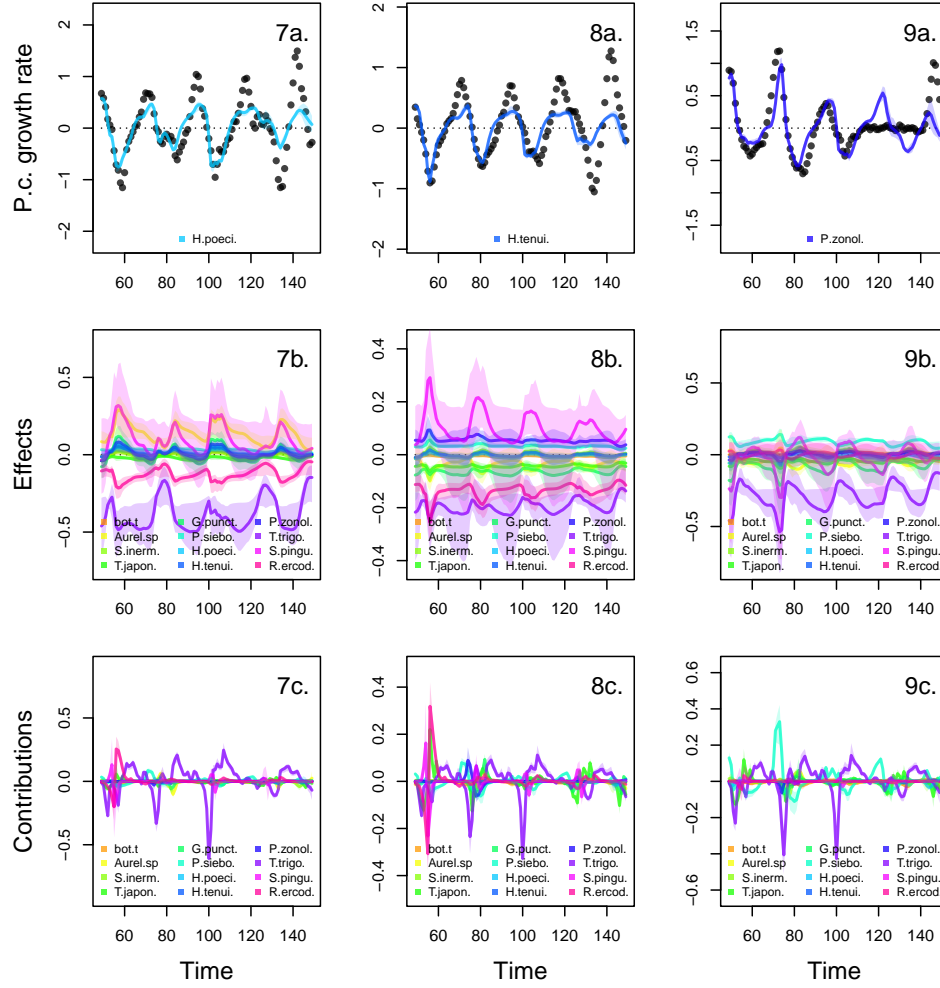


Figure S29: Drivers of dynamics of species abundance in the Maizuru bay community. This figure displays the NODE nonparametric approximations of the per-capita growth rates (7-9a.). We obtain the NODE approximations (7-9a., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, 7-9b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 7-9c.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.

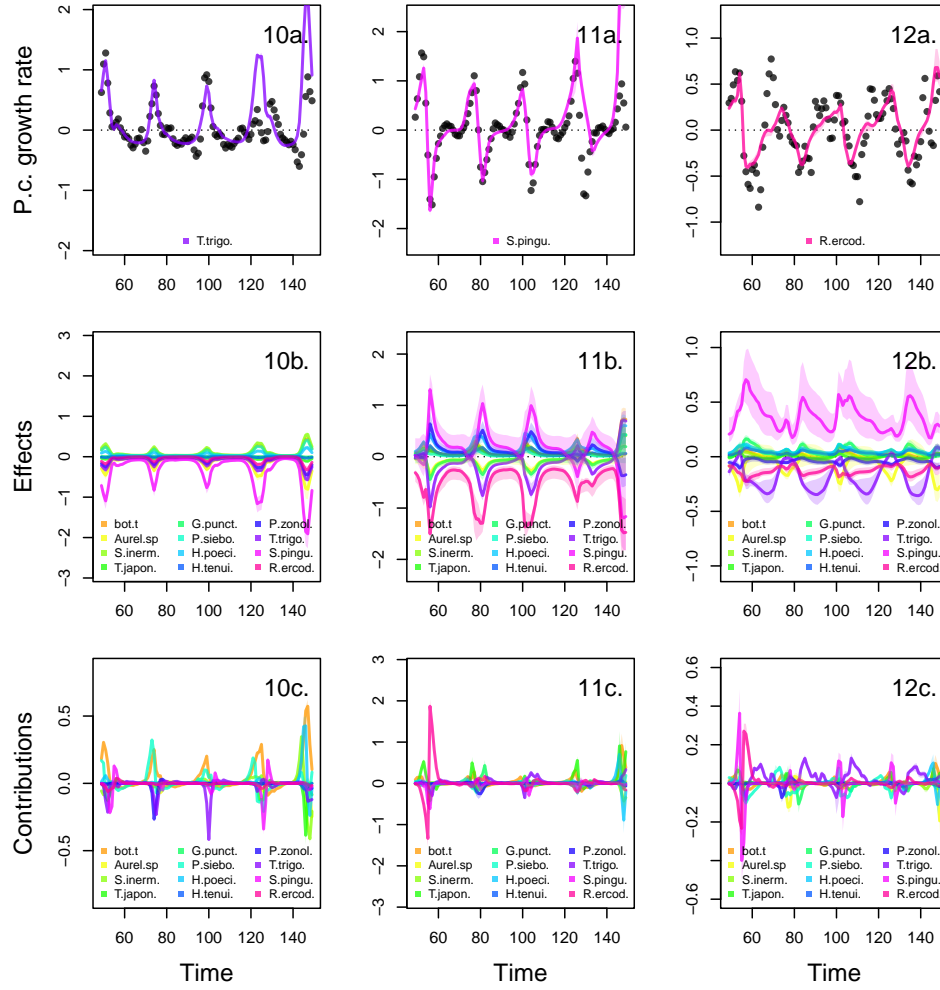


Figure S30: Drivers of dynamics of species abundance in the Maizuru bay community. This figure displays the NODE nonparametric approximations of the per-capita growth rates (10-12a.). We obtain the NODE approximations (10-12a., solid line) by fitting the interpolated per-capita growth rates (black dots) with ANNs that take population densities as input. We then estimate the direction of ecological interactions (effects, 10-12b.) by computing the derivative of the NODE approximations with respect to each density. Finally, we compute the strength of ecological interactions (contributions, 10-12c.) by multiplying the interpolated dynamics of each population with its effects. The shaded area shows the 90% confidence interval, obtained by approximately sampling the posterior distributions.