Chaos in a long-term experiment with a plankton community

Elisa Benincà, Jef Huisman, Reinhard Heerkloss, Klaus D. Jöhnk, Pedro Branco, Egbert H. Van Nes, Marten Scheffer & Stephen P. Ellner

Contents

This file contains the following Supplementary Information:

1. Materials and methods of the mesocosm experiment

In spring 1989, the mesocosm was filled with a 10 cm sediment layer from the Darss-Zingst estuary (southern Baltic Sea, 54° 26' N, 12° 42' E). After preincubation for one week to stabilize the sediment, the mesocosm was filled with 90 litres of water from the same location, which had been filtered through a 200 μm gauze. This inoculum

provided all species in the food web. During the first weeks, several plankton species were lost, as they were not able to survive the laboratory conditions. The sediment layer served as a source and refuge for resting stages and buffered the nutrient cycles.

The mesocosm was placed in a 15 ^{o}C climate room, and heated by an aquarium thermostat (Rena Cal Excel aquarium heater, 100 Watt, Aquarium Pharmaceuticals, Chalfont, PA, USA) to maintain the mesocosm temperature at 20 ^{o}C. The mesocosm was illuminated from above, by neon fluorescent lamps providing an incident irradiation of 50 µmol photons m^{-2} s^{-1} (16 hours/8 hours light-dark cycle). The mesocosm walls were not transparent. Salinity was maintained at 9‰, reflecting the salinity of the Darrs-Zingst estuary. The mesocosm was constantly aerated by bubbling with compressed air.

Nutrients were measured weekly after filtration of 20-mL samples through glass fiber filters (Whatman GF/F, 0.7 µm). Concentrations of soluble reactive phosphorous and dissolved inorganic nitrogen (nitrate, nitrite and ammonium) were analyzed according to standard methods (Rohde & Nehring 1979, Grasshoff et al. 1983) using a flow-injection autoanalyzer (Alpkem RFA-300, Alpkem, Wilsonville, OR, USA).

Species abundances were measured twice a week. Picophytoplankton, nanophytoplankton, and protozoa were counted alive, immediately after sampling, in a Kolkwitz plankton chamber under fluorescence light using an interference contrast microscope (Olympus research microscope BH-2). Bacteria were counted using fluorescence microscopy, in samples fixed with 2% glutaraldehyde and stained with DAPI (Porter & Feig 1980). Zooplankton, detritivores, and filamentous diatoms were sampled by scooping 10 litres from the mesocosm using a 2-litre beaker. The water was sieved through a 50 µm net to retain the plankton, and the filtrate was returned to the mesocosm. The sieved material was washed off into 20 ml aged biotope water and fixed with neutralized formaldehyde to a final concentration of 4%. For the abundant species, 100 to 200 individuals were counted. For less abundant species (i.e., with less than 100 individuals per sample), the total number of individuals per sample were counted. It was

difficult to take a representative sample of the cyclopoid copepods. The adults and later copepodite stages moved very fast, and escaped the scooping procedure by the beaker. Therefore, only the nauplii of the cyclopoid copepods were counted. The abundances of the species were converted into biomass using geometric equivalents of the body volumes. A list of the geometric conversion factors of the different species is provided in Heerkloss et al. (1991).

Attached algae were brushed from the walls of the mesocosm once a month. During the entire experiment, small quantities of biotope water filtered (0.4 μm pore size) from the Darss-Zingst estuary were added to compensate for water losses due to sampling. In addition, small quantities of distilled water were added to compensate for water loss due to evaporation.

2. Earlier analyses of the same time series

Part of the same time series has been presented in earlier publications (Heerkloss & Klinkenberg 1993, 1998; Dippner et al. 2002). The papers of Heerkloss & Klinkenberg (1993, 1998) present graphs of the time series, and suggest from visual inspection of the irregular ups and downs of the species that this food web might display chaotic dynamics. However, a nonlinear analysis of the data is not presented in their papers.

In contrast, Dippner et al. (2002) present a nonlinear analysis of the mesocosm data. However, they could not detect chaos in these time series. Why did Dippner et al. (2002) reach a conclusion that is completely different from our findings?

There were several differences between the approach of Dippner et al. (2002) and our approach.

First, Dippner et al. (2002) analyzed a shorter time series. They used the mesocosm data obtained until May 11, 1995 (day 1425 in our time series).

Second, Dippner et al. (2002) did not transform the time series to obtain stationary data with homogenized variances. As a consequence, their time series showed sharp spikes which may have hampered the interpretation of their results.

Third, Dippner et al. (2002) applied a different analysis. They used a graphical method known as recurrence quantification analysis (RQA). RQA is based on the analysis of recurrence plots, which were introduced by Eckmann et al. (1987). A recent review of recurrence plots is provided by Marwan et al. (2007). In essence, recurrence plots visualize the times at which a trajectory visits roughly the same area in phase space. The recurrence of the trajectory to similar states, after some time of divergence, is one of the key features of deterministic dynamical systems. To draw a recurrence plot, one needs to define when trajectories are considered as 'nearby' (i.e., when they pass through roughly the same area). For this purpose, a 'neighbourhood' of radius r is defined in phase space and two trajectories are assigned as 'nearby' if they both pass through the same neighbourhood. Unfortunately, in retrospect, the size of the neighbourhood in Dippner et al. was chosen much too large. For calanoid copepods, for instance, Dippner et al. (2002, p.33) used a neighbourhood with radius of $r = 5$ mg l^{-1}. However, this radius is about 30% of the maximum biomass measured in the calanoid time series. Moreover, population abundances of the calanoids remained below 5 mg l^{-1} for long stretches of time. In fact, the complete population dynamics of the calanoids from t=400 to 600 days (index=110 to 170), and also from t=850 to 1250 days (index=250 to 350) vanished within the same neighbourhood (this yielded the large black areas in their recurrence plots; compare Figure 1C and Figure 5 in Dippner et al. 2002). Thus, in retrospect, many of the interesting ups and downs in the population dynamics of the calanoids remained undetected in their recurrence quantification analysis, and, hence, their resolution was too coarse to detect rapid chaotic fluctuations with a predictability of only 15-30 days. The same comment applies to their analysis of the other phytoplankton and zooplankton species.

3. Transformation of the time series

We transformed the original time series, shown in Fig. 1b-g of the main text, to obtain stationary time series with equidistant data and homogeneous units of measurement. The transformation steps are illustrated for the bacteria (Fig. S1).

First, the time series were interpolated using cubic hermite interpolation, to obtain data with equidistant time intervals of 3.35 days (Fig. S1a).

Next, because the original time series showed many sharp spikes, the time series were rescaled using a fourth-root power transformation (Fig. S1b). The sharp spikes bias "direct method" estimates of the Lyapunov exponent, because nearby pairs of reconstructed state vectors mostly occurred in the troughs between spikes. The average rate of subsequent trajectory divergence from these pairs is therefore an estimate of the local Lyapunov exponent in the troughs, which may be very different from the global Lyapunov exponent. By making spikes and troughs more nearly symmetric, the power transformation resulted in a much more even spread of nearby state vector pairs across the full range of the data for all functional groups in the food web. The transformation is also useful for fitting nonlinear models of the deterministic skeleton (used for nonlinear predictability and indirect method estimates of the Lyapunov exponent), which was done by least squares and therefore is most efficient when error variances are stabilized. Fourth-root transformation is intermediate between the square-root transformation that would approximately stabilize the measurement error variance in count data from random subsamples, and the log transformation that is usually recommended for stabilizing process noise variance due to stochastic variation in birth and death rates.

The time series were then detrended using a Gaussian kernel with a bandwidth of 300 days (red line in Fig. S1b), to obtain stationary time series. Most species did not show long-term trends, except for the bacteria, detritivores (ostracods and harpacticoid copepods), dissolved inorganic nitrogen and soluble reactive phosphorus. One possible

explanation for these trends in the microbial loop could be the slow accumulation of refractory organic material in the mesocosm, but we have not measured this component.

Finally, the time series were linearly rescaled to have zero mean and a standard deviation of 1 (Fig. S1c).

The time series of cyclopoid copepods, protozoa, filamentous diatoms, harpacticoid copepods and ostracods contained long sequences of zero values. This does not imply that these species were absent from the food web during these periods, but that their concentrations were below the detection limit. Time series dominated by many zeros can bias the statistical analysis. Therefore, these time series were shortened to remove long sequences of zero values, before the data transformation.

The transformed data of all species in the food web are shown in Figure S2.

4. Spectral analysis

We applied spectral analysis to obtain a better understanding of the predominant periodicities in the species fluctuations. The discrete Fourier transform X for a time series of length N, with observations $x_0,\ldots,x_{N-1}$, is:

$$X(k) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j \, e^{-2\pi i j k / N}, \qquad (1)$$

where $k = 0,\ldots,N-1$ is the frequency index. The power spectrum is defined as:

$$P(k) = X(k)\, X^*(k), \qquad (2)$$

where $X^*(k)$ is the complex conjugate of $X(k)$.

We present both raw power spectra (Fig. S3) and smoothed power spectra (Fig. S4). The raw periodogram is not a consistent estimators of the spectral density, as its variance does not converge to zero when increasing the length of a time series (Percival

& Walden 1993). Consistent estimators can be derived by smoothing the raw periodogram. However, smoothing might introduce substantial bias at frequencies near spectral peaks by spreading and flattening the signal. Thus, there is a tradeoff between bias and variance. We smoothed the power spectrum using the modified Welch periodogram (Welch 1967). This method splits the time series in overlapping segments, called Hamming windows, and calculates the periodogram for each window separately. The Welch periodogram is then obtained by averaging the resulting periodograms. This yields a smooth periodogram, which is a consistent and asymptotically unbiased estimator of the spectral density. Visual inspection of the raw power spectra and Welch periodograms indicated that we obtained good results using 5 Hamming windows with 50% overlap.

All species in the food web showed reddened power spectra (i.e., decreasing power with increasing frequency; Fig. S3, S4), indicating some persistence in the data. Therefore, the power spectra of the species were compared with the power spectrum of red noise. A simple model for red noise is the univariate lag-1 autoregressive [AR(1)] process (e.g., Torrence & Compo 1998):

$$x_t = \alpha x_{t-1} + z_t, \qquad (3)$$

where α is the lag-1 autocorrelation calculated from the time series under investigation, x_0=0, and z_t is taken from Gaussian white noise. Following Gilman et al. (1963), the power spectrum of red noise calculated from Eq.3 is:

$$P(k) = \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos(2\pi k/N)}, \qquad (4)$$

where $k=0,\ldots,N$-1 is the frequency index.

Both the raw spectra and Welch periodograms show that picophytoplankton, rotifers, and calanoid copepods fluctuated with periodicities of ~30 days, and its possible harmonics at ~60 and ~120 days (Fig. S3, S4). This is consistent with earlier

studies reporting periodicities of ~30 days for phytoplankton-zooplankton oscillations (Scheffer & Rinaldi 2000, Fussmann et al. 2000).

The raw spectra further suggest that ostracods and harpacticoid copepods, which are connected to bacterial activity, may have fluctuated with periodicities of ~15 days (Fig. S3). Other periodicities in the raw spectra can be observed at ~25 days (bacteria, dissolved inorganic nitrogen, ostracods, harpacticoid copepods), ~75 days (dissolved inorganic nitrogen), ~150 days (soluble reactive phosphorus, ostracods), and ~225 days (bacteria), which points at intriguing linkages between nutrients and the microbial loop at a range of commensurate frequencies (Fig. S3). However, many of these periodicities are evident only in the raw spectra and less in the Welch periodogram. For this reason we cannot tell with certainty whether these periodicities of the nutrients and microbial loop are real features of the food web.

5. Predictability

5.1. Neural network

To investigate the predictability of the food-web dynamics, we employed a neural network model for each species in the food web. The neural network model assumes that the population dynamics of the focal species is a (complex nonlinear) function of the population abundances of this focal species and the species that have a direct link to this focal species (cf. Eq.1 in the Methods section):

$$N_{i,t+T} = f_{i,T}(N_{i,t}, N_{1,t}, N_{2,t}, ..., N_{m,t}) \qquad (5)$$

where $N_{i,t}$ is the population abundance (or nutrient concentration) of species i at time t, the subscripts 1 to m indicate all species directly linked to species i, T is the prediction time (i.e., the number of days that we want to predict in advance), and $f_{i,T}$ is an unknown function estimated by the neural network model.

The architecture of the neural network model is shown in Figure S5. The input of the network is received by input units, which monitor the population abundances of the focal species and the species linked to this focal species. The input values are passed on to a layer of hidden units. Each connection between input unit j and hidden unit k (indicated by an arrow in Fig. S5) performs a linear transformation determined by the connection strength γ_{kj}. Hence, the total input for hidden unit k at time t is given by:

$$in_k = \gamma_{ki} N_{i,t} + \sum_{j=1}^{m} \gamma_{kj} N_{j,t} + \alpha_k \qquad (6)$$

where the first term on the right-hand side describes the neuron's activity due to input from the focal species i, the second term describes the neuron's activity due to input from the connected species, and the third term, α_k, describes the neuron's intrinsic activity level. In other words, each hidden unit receives its own input, depending on its connection strengths to the input units and its intrinsic activity level.

The hidden unit performs a non-linear transformation on its total input, defined by the activation function ψ. This activation function is the same for all hidden units. Typically, the activation function is a sigmoid function, which approaches zero if the total input is very negative, while it approaches 1 if the total input is very positive. Following Nychka et al. (1992), we used the following activation function:

$$\psi(in_k) = \frac{in_k (1 + |in_k / 2|)}{2 + |in_k| + in_k^2 / 2} \qquad (7)$$

The activation signals from the hidden units are collected by a single output unit, which performs a linear transformation on the activation signals to present the output of the neural network. The network output can therefore be written as:

$$Out = \beta_0 + \sum_{k=1}^{n} \beta_k \psi(in_k) \qquad (8)$$

where β_0 is the intrinsic output level, the β_k's are the weights given to the activation signals from different hidden units, and n is the total number of hidden units. This

output represents the model prediction, by the neural network, of the population abundance of the focal species.

The total number of hidden units, n, should be sufficiently large to avoid inaccurate predictions, but should not be too large either as a large number of hidden units increases the computation time substantially. We selected the total number of hidden units by minimizing a generalized cross-validation statistic (GCV_2) with 2-fold overweighting of model degrees of freedom to avoid overfitting (Nychka et al. 1992). This yielded an estimate of 4 hidden units for bacteria, nanophytoplankton and protozoa, 5 hidden units for cyclopoid and calanoid copepods, rotifers, picophytoplankton, phosphorus, nitrogen, ostracods and harpacticoid copepods, and 6 hidden units for filamentous diatoms.

The parameters (α_k, β_i, γ_{ij}) of the neural network models were estimated using the software package LENNS (Ellner et al. 1992, Nychka et al. 1992), which fits neural networks to data using a least-squares approach.

5.2 Testing for differences between the nonlinear and linear model

We tested whether the nonlinear neural network model yielded significantly higher predictabilities than the corresponding linear model. For this purpose, we calculated the Pearson product-moment correlation coefficient (r) between predicted and observed values, for both the nonlinear model (r_1) and the linear model (r_2). These correlation coefficients are simply the square roots of the coefficients of determination (R^2) shown in Figure 2 of the main text.

For each prediction time, we tested the null hypothesis that the linear model and nonlinear model yielded the same predictability (i.e., $r_1 = r_2$) against the alternative hypothesis that the nonlinear model yielded better predictions than the linear model

(i.e., $r_1 > r_2$). The two correlation coefficients were both transformed with the Fisher z-transformation (Sokal & Rohlf 1995):

$$z_i = \frac{1}{2}\ln\left(\frac{1+r_i}{1-r_i}\right) \tag{9}$$

The sampling distribution of the z statistic is known to be approximately normal, with a standard error of

$$\sigma_z = \frac{1}{\sqrt{N-3}} \tag{10}$$

where N is the number of observations. We note that, in our case, N is the same for both correlation coefficients, since they are calculated for the same time series. Accordingly, we calculated the difference

$$\Delta z = \frac{z_1 - z_2}{\sqrt{2/(N-3)}} \tag{11}$$

Under the null hypothesis, the sampling distribution of Δz has a standard normal distribution with mean of 0 and standard deviation of 1. Hence, in view of the alternative hypothesis, we rejected the null hypothesis if the calculated value of Δz exceeded the 95th percentile of the standard normal distribution.

The results show that, already after one time step, the nonlinear model yielded significantly better predictions than the linear model for all species (Table S1).

6. Methods for estimating Lyapunov exponents

Numerous methods have been proposed and studied for estimating Lyapunov exponents from time series data. Essentially, these methods can be classified into two types of approaches, *direct methods* and *indirect methods* (the latter are also called *Jacobian methods*); we applied both approaches.

6.1 Direct method by time-delay embedding

Direct methods descend from Guckenheimer (1982) and Wolf et al. (1985). The data are searched to find nearby pairs of state vectors (or reconstructed state vectors). In other words, the data are searched for different points in the time series at which all species abundances in the food web are in a similar state. The rate of trajectory divergence at subsequent times, averaged over many such pairs, is an estimate of the dominant Lyapunov exponent λ.

All calculations to estimate the Lyapunov exponent by the direct method were performed using the software of the Tisean package (Hegger et al. 1999). We chose the procedure of Rosenstein et al. (1993), because it was specifically developed and tested for short, noisy time series (which ours are by the standards of theoretical physics). This method uses attractor reconstruction by time-delay embedding (Takens 1981, Strogatz 1994, Kantz & Schreiber 1997), so that separate estimates could be obtained for each species in the food web, providing an additional check on the robustness of our conclusions.

Time-delay embedding requires a suitable choice of time delay and embedding dimension. Since all data are from the same dynamical system, we chose a single value of the time delay and embedding dimension representative for the entire food web. Rosenstein et al. (1993) suggested that a good choice of time delay is the time lag where the autocorrelation function drops to a fraction 1-1/e (i.e., 63%) of its initial value. Following this criterion, we estimated time delays ranging from 1 to 4 time steps, depending on the species (where 1 time step equals 3.35 days). Turchin (2003) suggested a time delay of 1 time step for organisms with generation times less than the unit time interval. In our food web, several species have a generation time equal or less than ~3.35 days (e.g., bacteria, picophytoplankton, nanophytoplankton, filamentous diatoms, ostracods, rotifers). Hence, a time delay of 1 time step would seem suitable. A time delay of 1 also gave robust results in terms of the linear scaling region in the

exponential divergence which is used to calculate the Lyapunov exponent. Accordingly, we chose a time delay of 1 time step.

Theory suggests that the embedding dimension, *m*, can be estimated by the value where the number of false nearest neighbours drops to zero (Kantz & Schreiber 1997). Following this criterion, we estimated embedding dimensions ranging from *m*=4 to *m*=9, depending on the species. At embedding dimensions of 6 and higher, the initial slope of the exponential divergence became independent of embedding dimension (Fig. S6). Hence, we chose an embedding dimension of *m*=6.

Estimates of trajectory divergence, and therefore of the Lyapunov exponent, may be distorted if nearby state vectors (in state space) are also near in time. For example, the time-delay state vector consisting of samples 101-106 is near to that consisting of samples 102-107, and their future histories never diverge very far. The Theiler window offers a classical and effective solution to this problem (Theiler 1986, Kantz & Schreiber 1997). A Theiler window removes temporally nearby data points from the set of pairs used to estimate trajectory divergence. We estimated a suitable size of the Theiler window by visual inspection of space time separation plots (Provenzale et al. 1992). Space time separation plots show how the temporal distance between pairs of data points affects their spatial distance on the reconstructed attractor (Fig. S7). The Theiler window should be sufficiently large to exclude those data points for which the spatial distance on the attractor is affected by their temporal distance. For our species, the effect of temporal distance on spatial distance vanished when data points were separated by more than 10 to 20 time steps (Fig. S7). To be on the safe side, we therefore introduced a Theiler window of 50 time steps (~170 days) for each species in the food web.

6.2 Jacobian method: a neural network food web model

Jacobian methods descend from Eckmann et al. (1986). These are based on the development of a deterministic model of the underlying dynamics of the system. This deterministic model will henceforth be called the '*deterministic skeleton*'. The deterministic skeleton is differentiated to estimate the Jacobian matrices. The Lyapunov exponent is then defined in terms of the sequence of Jacobian matrices of the deterministic skeleton, evaluated at the time series of observed or reconstructed state vectors. Thus, Jacobian methods require the preliminary step of estimating the deterministic skeleton, either locally or by fitting a global map to the time series. For theoretical properties of Jacobian methods, see McCaffrey et al. (1992), Bailey et al. (1997), Lu and Smith (1997).

The length of our time series and the high dimensionality of the system (i.e., the relatively large number of interlinked species in the food web) favour the use of a neural network regression model to estimate the deterministic skeleton, as discussed by Ellner and Turchin (1995). However, our analysis here incorporates some subsequent developments. In particular, follow-up studies have shown that semi-mechanistic (also called semi-parametric) models should be preferred over state space reconstruction in time-delay coordinates (Ellner et al. 1998, Smith et al. 2000). "Semi-mechanistic" means that the structure of the model is based on biological knowledge about the system (when that is available), as are any process rate equations for which independent data are available, while nonparametric methods are used to fit aspects about which little is known. Here we used the food web structure (Fig. 1a) to dictate the structure of the model, exactly as in the estimates of nonlinear predictability. Hence, functional relationships in the model include only those species that have a direct link in the food web to the focal species. So for example, the deterministic skeleton for rotifers is assumed a priori to have the form

$$\text{rotifers}(t+1) = f_{\text{rot}}[\text{rotifers}(t),\text{cyclopoids}(t),\text{pico}(t),\text{nano}(t),\text{bacteria}(t)]$$

where t is the time measured in time steps of 3.35 days. The omission of other dependencies (e.g., absence of nutrients on the right-hand side of the last equation) leads to structural zeros in the fitted Jacobians, which avoids "fitting the noise" or incorporating functional relationships in the model that are absent in the real system. Another advantage of adopting the food web structure is that it eliminates the potential problem of spurious exponents in Jacobian-based estimates using state space reconstruction (Sauer et al. 1998).

As in the direct method estimates, the forecasting interval (one sample time) is chosen on mechanistic grounds (generation time) rather than a statistical rule-of-thumb. However, the skeleton maps f were estimated by fitting a neural network regression model with no a priori limits on the number of hidden units, so that any model shape and complexity can be fitted, if there is evidence for it in the data.

Hence, the final issue is selecting the complexity (number of hidden units) of the skeleton map for each species in the food web. This has no simple resolution. To avoid overfitting we need a conservative criterion. Ellner and Turchin (1995), following Nychka et al. (1992), used GCV_2, a modification of the Generalized Cross Validation criterion in which model degrees of freedom are overweighted by a factor of 2. Subsequent work suggested that twofold overweighting may be excessive. With short, noisy time series Kendall (2001) found that GCV_2 model selection had a substantial risk of drastically underfitting, leading to spurious strongly negative estimates of λ. For longer and less noisy data sets, McCaffrey et al. (1992) and Nychka et al. (1992) obtained good results using the Bayesian Information Criterion (BIC) with neural network models. Assuming Gaussian errors and using the maximum likelihood estimate of the error variance (i.e. the mean squared residual error, MSE), the BIC criterion is equivalent to chosing the model that minimizes $\log(MSE)+P\log(N)/N$, where P is the number of model parameters and N is the sample size. GCV_2 is equivalent to selecting the model that minimizes $\log(MSE)-2\log(1-2P/N)$. For our time series length, the BIC criterion is more conservative.

Based on these considerations, we applied both the GCV_2 and BIC criteria to estimate the number of hidden units. However, in our case, both criteria gave rather similar results. Depending on the species, we obtained skeleton maps *f* with up to 8 hidden units, which was sufficient for model selection under all criteria.

All calculations to estimate the Lyapunov exponent by the Jacobian method were performed using the LENNS software (Ellner et al. 1992, Nychka et al. 1992). An R (R Development Core Team 2006) version of LENNS for the Windows operating system is available on request from S.P.E. or can be downloaded from www.eeb.cornell.edu/ellner/software.

To place confidence limits on the estimate of λ we used bootstrapping in particular the "resampling errors" approach to bootstrapping regression models (Davison and Hinkley 1997, section 6.2.3), as follows. For each functional group *i* in the food web, let $\hat{f}_i(X_i)$ denote the fitted deterministic skeleton map, where X_i is the vector of all functional groups linked with species *i*, including itself. From this we obtain a time series of forecasting errors $e_i(t) = x_i(t) - \hat{f}_i(X_i(t-1))$. Each bootstrap sample for functional group *i* was generated by first sampling with replacement from $\{e_i(t)\}_{t=2}^{N}$ to generate a series $\{e_i^*(t)\}_{t=2}^{N}$. These reshuffled error terms were subsequently used to create a new time series consisting of fictitious "one step ahead" data

$$x_i^*(t) = \hat{f}_i(X_i(t-1)) + e_i^*(t) \qquad (12)$$

For each such "data set" we then refitted the neural network model (with x_i^* as the response variable, and the real data series $X_i(t\text{-}1)$ as the predictors), including selection of model complexity by the BIC criterion. Due to the high computational time required for fitting neural network models reliably, we limited the complexity of the refitted networks to at most 1 more hidden unit than the number selected by BIC for the real data. Because more complex models have a tendency to lead to more strongly positive estimates of λ, this limitation is conservative for our purposes (assessing the strength of

evidence that $\lambda>0$). Obtaining 1000 bootstrap replicates required about one month on a current desktop PC.

Based on these 1000 bootstraps, a one-sided confidence interval at the 95% confidence level yielded a lower bound of $\lambda=0.03$ day^{-1}. This confirms that the Lyapunov exponent was significantly positive. We also report the two-sided confidence interval at the 95% confidence level, which yielded $0.025 < \lambda < 0.109$ day^{-1}.

We also assessed the robustness of the Jacobian estimates informally in several different ways (these were done individually, not in all possible combinations).

- We modified the food web by completely removing the detritivores (ostracods and harpacticoid copepods), and subsequent refitting of the neural network model to the remaining time series.

- We eliminated the first 500 days of the time series, during which the temperature was slightly higher than 20 ^{o}C.

- Instead of a neural network we used a generalized additive model (GAM) with spline ridge functions (Wood 2006, package **mgcv** package in R). The model for functional group *i* included a univariate self-limitation spline term $s(x_i)$, a univariate spline term $s(x_j)$ for each group *j* linked to group *i*, and a bivariate spline term $s(x_i, x_j)$ for each group *j* linked directly to group *i* (i.e., indirect links via the microbial loop were not represented by bivariate spline terms). A spline GAM can be fitted quickly and reliably, including selecting model complexity by GCV criteria, but it cannot include higher-order interactions (for example, all possible 3-way interactions between rotifers and all groups linked to them) due to the large number of basis functions required for such a model. We fitted the GAM to the shortened time series mentioned just above.

All of these modifications produced similar results, in particular they all yielded a positive value for the dominant Lyapunov exponent. For the GAM model, we obtained

a Lyapunov exponent of λ=0.08 day^{-1}, and an application of our bootstrap procedure (using 200 replicates) resulted in a positive lower bound of the Lyapunov exponent at the 95% confidence level of 0.024 day^{-1}. Analyses of the shorter time series (without the first 500 days) produced a slightly larger estimate of the dominant Lyapunov exponent, of λ=0.097 day^{-1}, with a positive lower bound at the 95% confidence level of 0.02 day^{-1}.

7. Temperature fluctuations

The mesocosm was placed in a 15 ^{o}C climate room, and heated by an aquarium thermostat to maintain a constant water temperature. During the entire period of investigation, for 2,319 days, the temperature of the mesocosm was ~20˚C (mean = 20.28 ˚C, s.d. =1.07 ˚C, n = 688). Nevertheless, some temperature fluctuations were unavoidable (Fig. S8a), either by failure of the climate room or by failure of the aquarium thermostat. For instance, accidental failures of the thermostat resulted in a fast temperature increase on April 3, 1992 (day 292), June 17, 1993 (day 732), May 18, 1995 (day 1432), and September 4, 1997 (day 2272).

One might argue that the chaotic behavior of the food web could have been driven by the temperature fluctuations. There are actually two hypotheses: (1) the temperature itself fluctuated chaotically, and/or (2) the temperature was not chaotic, but temperature variability pushed the species dynamics into a chaotic regime.

To investigate these hypotheses, we carried out several analyses. For this purpose, the temperature data were transformed in exactly the same way as the species in the food web (i.e., interpolation, fourth-root power transformation, detrending, and normalization of the time series). This yielded a stationary time series with equidistant data and homogeneous units of measurements (Fig. S8b).

First, we investigated whether the observed species variability was associated with temperature variability, by calculation of the product-moment correlation coefficient between the species abundances and temperature. This revealed that bacteria and rotifers showed significantly positive correlations with temperature, whereas the other species did not show a significant relationship (Table S2). Visual inspection of the data suggested that the positive correlations of bacteria and rotifers could be attributed to the slightly elevated temperatures from day 300 to day 475 (Fig. S8a; mean = 21.55 ^{o}C; s.d. = 0.96 ^{o}C, n = 47). Indeed, when we shortened the time series by removal of the first 500 days, none of the species in the food web showed significant correlations with temperature anymore (Table S2). Hence, we conclude that the temperature fluctuations may have had some effect on the species abundances, but that this effect was relatively minor and mainly concentrated in the first 500 days of the time series.

Second, we investigated the predictability of the temperature fluctuations by developing a neural network model using the same methodology as for the species in the food web. However, we could of course not exploit the food web structure to predict temperature, and instead we predicted the temperature by time-delay embedding of the time series. We used a time delay of d=1 time step, which corresponded to the value where the autocorrelation function dropped to 1-1/e. We used an embedding dimension of m=3, corresponding to the first minimum of the GCV_2 statistics as a function of the embedding dimension. The results showed that the predictability of temperature was already low for a prediction time of only one time step (R^2=0.14), and was further reduced from the second time step onwards (R^2<0.08). Hence, the temperature signal was noise dominated, with only a very weak deterministic component.

Third, we estimated the Lyapunov exponent using both direct and indirect methods. In the direct method, the Lyapunov exponent was estimated using attractor reconstruction by time-delay embedding, following exactly the same approach as for the species. We used a time delay of d=1 time step, an embedding dimension of m=3, and a Theiler window of 70 time steps (about 235 days). This yielded a strongly positive Lyapunov exponent (λ=0.151 day^{-1}). Does this imply that the temperature fluctuations were driven by chaotic dynamics? Not necessarily. The direct method may yield positive Lyapunov exponents when the dynamics are chaotic, but also when the dynamics are not chaotic but dominated by noise (Ellner and Turchin 1995). Moreover, the predictability results above had indicated that the temperature fluctuations were noise dominated.

The indirect method is specifically designed for noisy data sets; it avoids the noise signal by investigation of the deterministic skeleton underlying the time series. For this purpose, we used again a neural network model to estimate the deterministic skeleton, and subsequently calculated the Lyapunov exponent by evaluation of the Jacobian matrices. The methodology was as described previously (see the earlier section on the Jacobian method). However, because we could not exploit the food web structure, we used time-delay embedding with a time delay of d=1 time step and an embedding dimension of m=3 to generate the neural network model. This yielded a strongly negative Lyapunov exponent (λ= –0.091 day^{-1}). In fact, a negative Lyapunov exponent should have been expected, because the temperature was regulated by a thermostat system (i.e., the temperature trajectories should all converge to the same point attractor at 20 ^{o}C). The contrasting results from the direct and indirect method emphasize once more that assessment of the chaotic nature of noisy time series (i.e.,

many ecological time series) requires investigation of the underlying deterministic skeleton.

Thus, we conclude that the temperature fluctuations were not driven by chaotic dynamics, but reflected a stable thermostat system disturbed by noise.

Finally, could the temperature fluctuations have pushed the species dynamics into a chaotic regime? In principle, this could have been the case. However, the species abundances showed at best only a very weak relationship with temperature (Table S2). Moreover, the positive Lyapunov exponent of the entire food web, estimated by the indirect method (see main text), points out that the underlying population dynamics were chaotic. Hence, we conclude that the chaotic nature of this food web was not driven by external forcing, but by the food web interactions themselves.

References

S1. Rohde, K. H. & Nehring, D. Ausgewählte Methoden zur Bestimmung von Inhaltsstoffen im Meer- und Brackwasser. *Geod. Geoph. Veröff. R IV* **27**, l-68 (1979).

S2. Grasshoff, K., Ehrhardt, M. & Kremling, K., eds. *Methods of Seawater Analysis*, 2nd edition (Verlag Chemie, Weinheim, 1983).

S3. Porter, K. G. & Feig, Y. S. The use of DAPI for identifying and counting aquatic microflora. *Limnol. Oceanogr.* **25**, 943-948 (1980).

S4. Heerkloss, R., Schnese, W. & Adamkiewicz-Chojnacka, B. Influence of eutrophication on seasonal variation of zooplankton biomass in shallow coastal lagoons of the Southern Baltic. *Acta Ichthyologica et Piscatoria* **21**, 67–76 (1991).

S5. Heerkloss, R. & Klinkenberg, G. Chaotic dynamics of a plankton community in a species-depleted mesocosmos. *Verh. Internat. Verein. Limnol.* **25**, 995-1000 (1993).

S6. Heerkloss, R. & Klinkenberg, G. A long-term series of a planktonic foodweb: a case of chaotic dynamics. *Verh. Internat. Verein. Limnol.* **26**, 1952-1956 (1998).

S7. Dippner, J. W., Heerkloss, R. & Zbilut, J. P. Recurrence quantification analysis as a tool for characterization of non-linear mesocosm dynamics. *Mar. Ecol. Prog. Ser.* **242**, 29-37 (2002).

S8. Eckmann, J. P., Kamphorst, S. O. & Ruelle, D. Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973-977 (1987).

S9. Marwan, N., Romano, M. C., Thiel, M. & Kurths, J. Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237-329 (2007).

S10. Percival, D. B. & Walden, A. T. *Spectral Analysis for Physical Applications*. (Cambridge Univ. Press, Cambridge, 1993).

S11. Welch, P. D. The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.*, vol. **AU-15**, 70-73 (1967).

S12. Torrence, C. & Compo, G. P. A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.* **79**, 61-78 (1998).

S13. Gilman, D. L., Fuglister, F. J. & Mitchell Jr., J. M. On the power spectrum of "red noise". *J. Atmos. Sci.* **20**, 182-184 (1963).

S14. Scheffer, M. & Rinaldi, S. Minimal models of top-down control of phytoplankton. *Freshwater Biol.* **45**, 265-283 (2000).

S15. Fussmann, G. F., Ellner, S. P., Shertzer, K. W. & Hairston, N. G. Jr. Crossing the Hopf bifurcation in a live predator-prey system. *Science* **290**, 1358-1360 (2000).

S16. Nychka, D., Ellner, S., Gallant, A. R. & McCaffrey, D. Finding chaos in noisy systems. *J. Roy. Stat. Soc. B Met.* **54**, 399-426 (1992).

S17. Ellner, S. P., Nychka, D. W. & Gallant, A. R. LENNS, a program to estimate the dominant Lyapunov exponent of noisy nonlinear systems from time series data. *Institute of Statistics Mimeo Series # 2235 (BMA series # 39)*, Statistics Department, North Carolina State University, Raleigh NC 27695-8203 (1992).

S18. Sokal, R. R. & Rohlf, F. J. *Biometry*, 3rd edition (Freeman, New York, 1995).

S19. Guckenheimer, J. Noise in chaotic systems. *Nature* **298**, 358-361 (1982).

S20. Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. Determining Lyapunov exponents from a time-series. *Phys. D* **16**, 285-317 (1985).

S21. Hegger, R., Kantz, H. & Schreiber, T. Practical implementation of nonlinear time series methods: the Tisean package. *Chaos* **9**, 413-435 (1999).

S22. Rosenstein, M. T., Collins, J. J. & De Luca, C. J. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D* **65**, 117-134 (1993).

S23. Takens, F. *Detecting Strange Attractors in Turbulence* (Lecture Notes in Math. 898, Springer, New York, 1981).

S24. Strogatz, S. H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Perseus, Cambridge, MA, 1994).

S25. Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

S26. Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton University Press, Princeton, 2003).

S27. Theiler, J. Spurious dimension from correlation algorithms applied to limited time-series data. Phys. Rev. A **34**, 2427-2432 (1986).

S28. Provenzale, A., Smith, L. A., Vio, R. & Murante, G. Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D* **58,** 31-49 (1992).

S29. Eckmann, J. P., Kamphorst, S. O., Ruelle, D. & Ciliberto, S. Liapunov exponents from time-series. *Phys. Rev. A* **34**, 4971-4979 (1986).

S30. McCaffrey, D. F., Ellner, S., Gallant, A. R. & Nychka, D. W. Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *J. Amer. Stat. Assoc.* **87**, 682-695 (1992).

S31. Bailey, B. A., Ellner, S. & Nychka, D. W. Chaos with confidence: asymptotics and applications of local Lyapunov exponents. In *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences* (eds Cutler C. D. & Kaplan D. T.) 115-133 (Fields Institute Communications, vol **110**, American Mathematical Society, Providence RI, 1997).

S32. Lu, Z.-Q. & Smith, R. L. Estimating local Lyapunov exponents. In *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences* (eds Cutler C. D. & Kaplan D. T.) 135-151 (Fields Institute Communications, vol **110**, American Mathematical Society, Providence RI, 1997).

S33. Ellner, S. P. & Turchin, P. Chaos in a noisy world: new methods and evidence from time-series analysis. *Am. Nat.* **145**, 343-375 (1995).

S34. Ellner, S. P. *et al.* Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *Am. Nat.* **151**, 425-440 (1998).

S35. Smith, R. H. *et al.* Blowflies as a case study in non-linear population dynamics. In *Chaos in Real Data: Analysis of Non-linear Dynamics from Short Ecological Time Series* (eds Perry, J. N., Smith, R. H., Woiwod, I. P. & Morse, D. R.) 137-172 (Kluwer Academic Publishers, Dordrecht, 2000).

S36. Sauer, T. D., Tempkin, J. A. & Yorke, J. A. Spurious Lyapunov exponents in attractor reconstruction. *Phys. Rev. Lett.* **81**, 4341-4344 (1998).

S37. Kendall, B. E. Cycles, chaos, and noise in predator-prey dynamics. *Chaos Solitons Fract.* **12**, 321-332 (2001).

S38. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and their Application* (Cambridge University Press, Cambridge, 1997).

S39. Wood, S. N. *Generalized Additive Models: An Introduction with R* (Chapman and Hall/CRC, Boca Raton, Florida, 2006).

S40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B.* **57**, 289–300 (1995).

Table S1. Statistical evaluation of differences between the predictability of the nonlinear neural network model and the predictability of the best-fitting linear model. Data entries show P values, for each species and each prediction time (measured in time steps, where one time step equals 3.35 days). If $P < 0.05$, the nonlinear model had a significantly higher predictability than the linear model. n.s. = not significant.

Species	Prediction time			
	1	2	3	> 3
Bacteria	<0.02	<0.001	<0.001	<0.001
Harpaticoids	<0.001	<0.001	<0.001	<0.001
Ostracods	<0.001	<0.001	<0.001	<0.001
Nitrogen	n.s	<0.02	<0.001	<0.001
Phosphorus	n.s.	<0.005	<0.001	<0.001
Picophytoplankton	<0.05	<0.001	<0.001	<0.001
Nanophytoplankton	n.s	<0.01	<0.001	<0.001
Filamentous diatoms	<0.02	<0.001	<0.001	<0.001
Rotifers	<0.05	<0.001	<0.001	<0.001
Protozoa	n.s	<0.01	<0.005	<0.005
Calanoids	n.s.	<0.01	<0.001	<0.001
Cyclopoids	n.s.	<0.05	<0.05	<0.02

Table S2. Correlations between species abundances and temperature. Table entries show the product-moment correlation coefficients, after transformation of the data to stationary time series (see Methods section). In the shortened time series, the first 500 days were removed from the data set. Significance tests were corrected for multiple hypothesis testing by calculation of adjusted p-values using the false discovery rate (Benjamini & Hochberg 1995). Significant correlations are indicated in bold: *** = $P < 0.001$, ** = $P < 0.01$; * = $P < 0.05$. Filamentous diatoms and cyclopoid copepods were not included in the correlation analysis, because their time series contained too many zeros.

Species	Correlation coefficient	
	complete time series	shortened time series
Bacteria	**0.11***	0.11
Harpacticoids	0.09	0.09
Ostracods	0.10	0.11
N	0.02	0.03
P	0.07	-0.11
Picophytoplankton	-0.12	-0.04
Nanophytoplankton	-0.07	-0.07
Rotifers	**0.11***	0.08
Protozoa	-0.01	0.00
Calanoids	-0.11	0.05

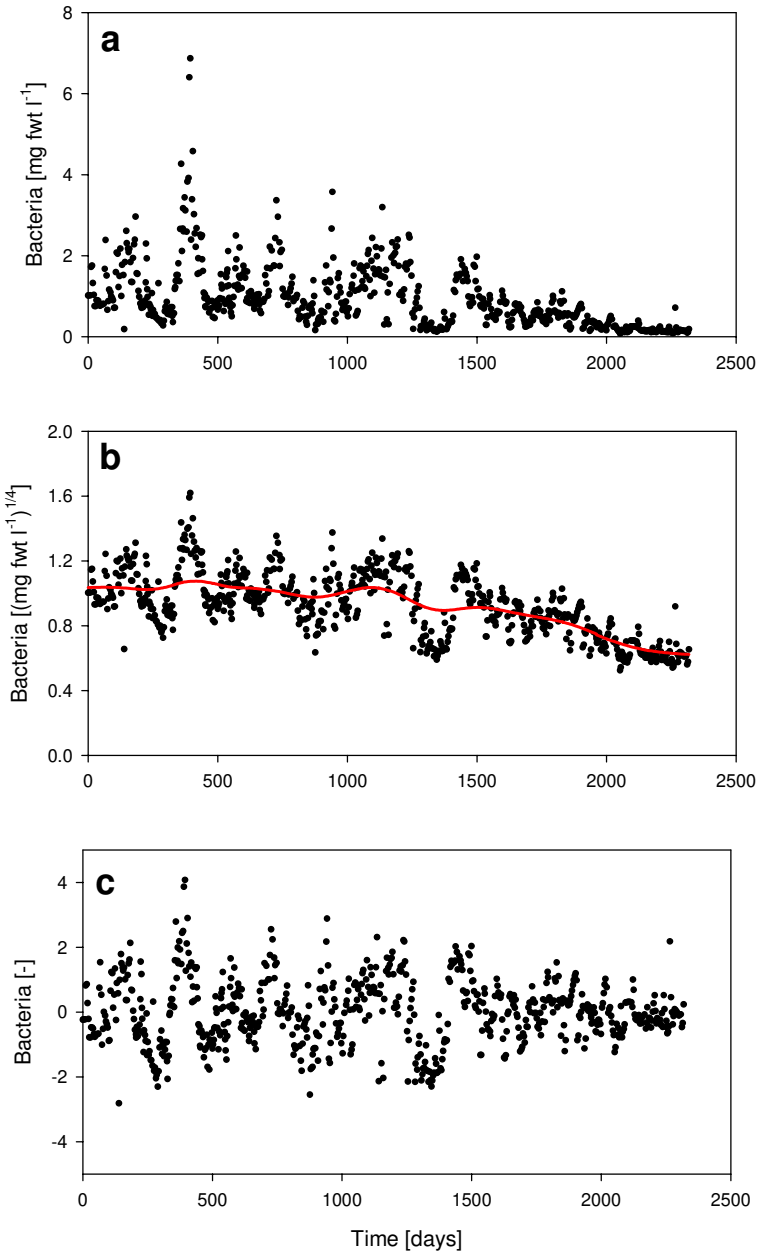

Figure S1. Data treatment of the time series of bacteria. **a,** First, the time series was interpolated to obtain equidistant intervals of 3.35 days. **b,** Next, the time series was transformed by a fourth-root power transformation to suppress large spikes in the data, and the trend (red line) was calculated by a Gaussian kernel window with a bandwidth of 300 days. **c,** Finally, the time series was detrended, and subsequently normalized to obtain a stationary time series with mean zero and standard deviation of 1.

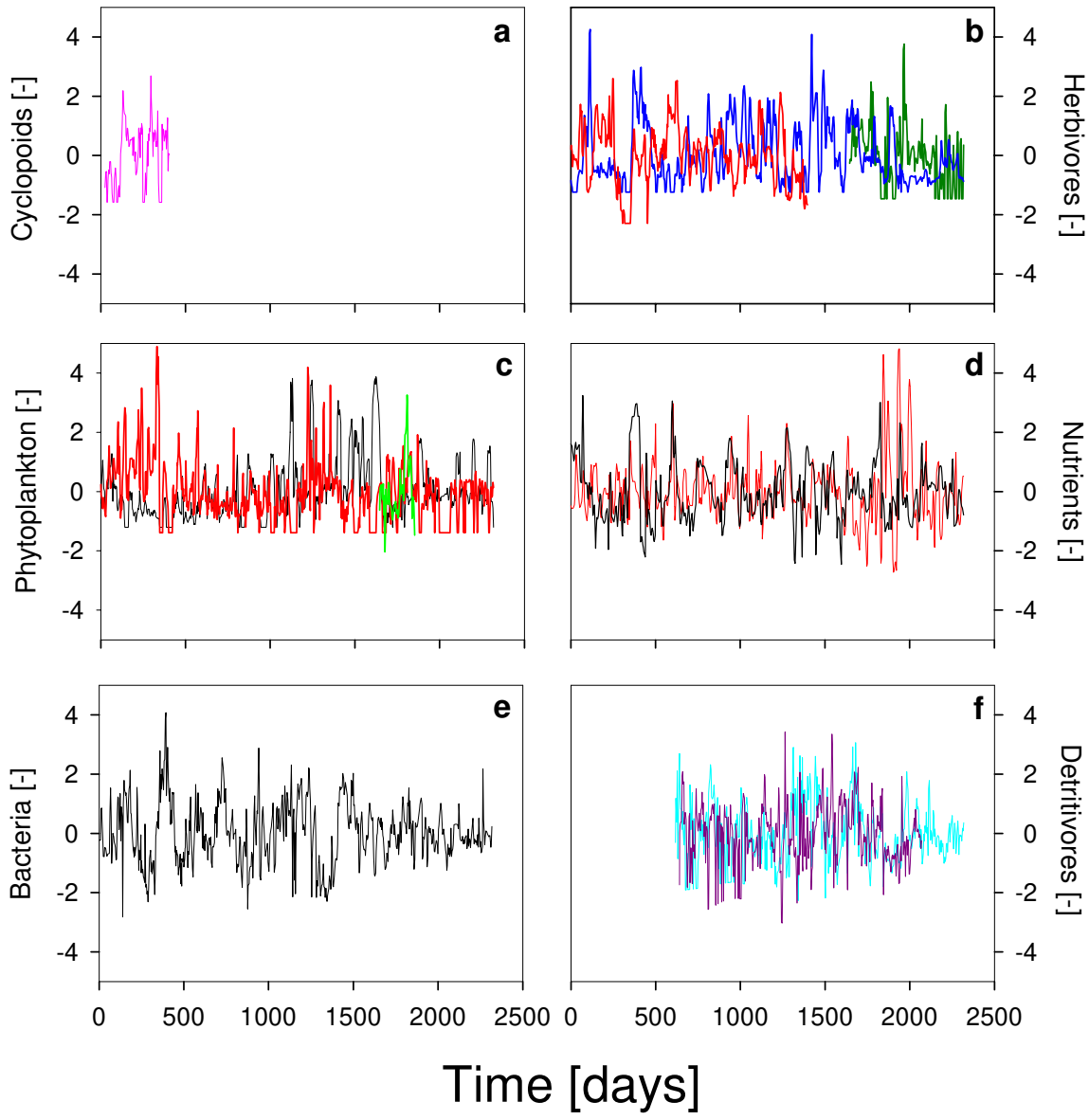

Figure S2. Stationary time series of the functional groups in the food web, after data treatment. **a,** Cyclopoid copepods**; b**, calanoid copepods (red), rotifers (blue), and protozoa (dark green); **c**, picophytoplankton (black), nanophytoplankton (red), and filamentous diatoms (green); **d**, dissolved inorganic nitrogen (red) and soluble reactive phosphorus (black); **e**, heterotrophic bacteria; **f**, harpacticoid copepods (violet) and ostracods (light blue).

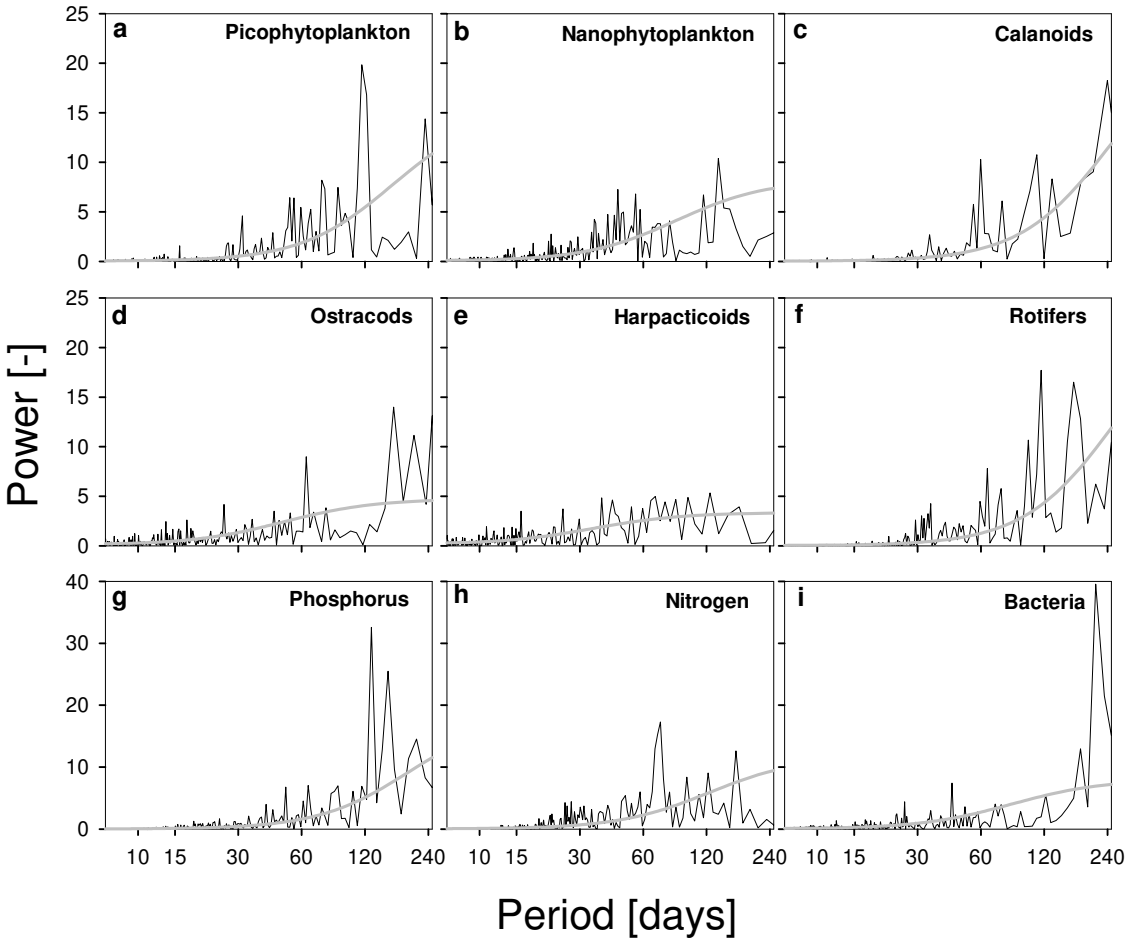

Figure S3. Raw power spectra of the species. **a**, Picophytoplankton; **b**, nanophytoplankton; **c**, calanoid copepods; **d**, ostracods; **e**, harpacticoid copepods; **f**, rotifers; **g**, soluble reactive phosphorus; **h,** dissolved inorganic nitrogen; **i**, bacteria. For comparison, the grey line shows the red-noise spectrum calculated from an AR1-process. Note the different scale of the y-axes in panels **g-i**.

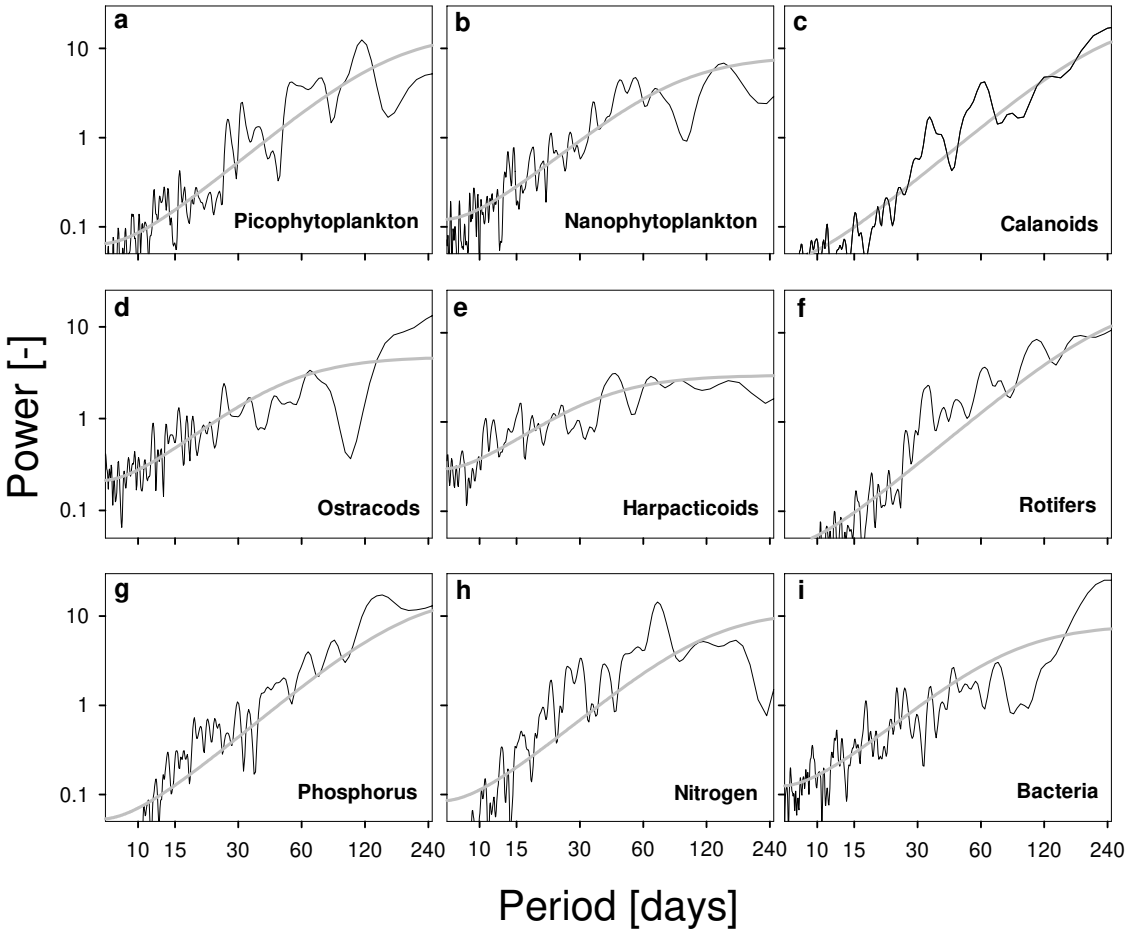

Figure S4. Welch periodogram of the species. **a**, Picophytoplankton; **b**, nanophytoplankton; **c**, calanoid copepods; **d**, ostracods; **e**, harpacticoid copepods; **f**, rotifers; **g**, soluble reactive phosphorus; **h,** dissolved inorganic nitrogen; **i**, bacteria. For comparison, the grey line shows the red-noise spectrum calculated from an AR1-process. Please note that the power in this graph is plotted on a log scale.

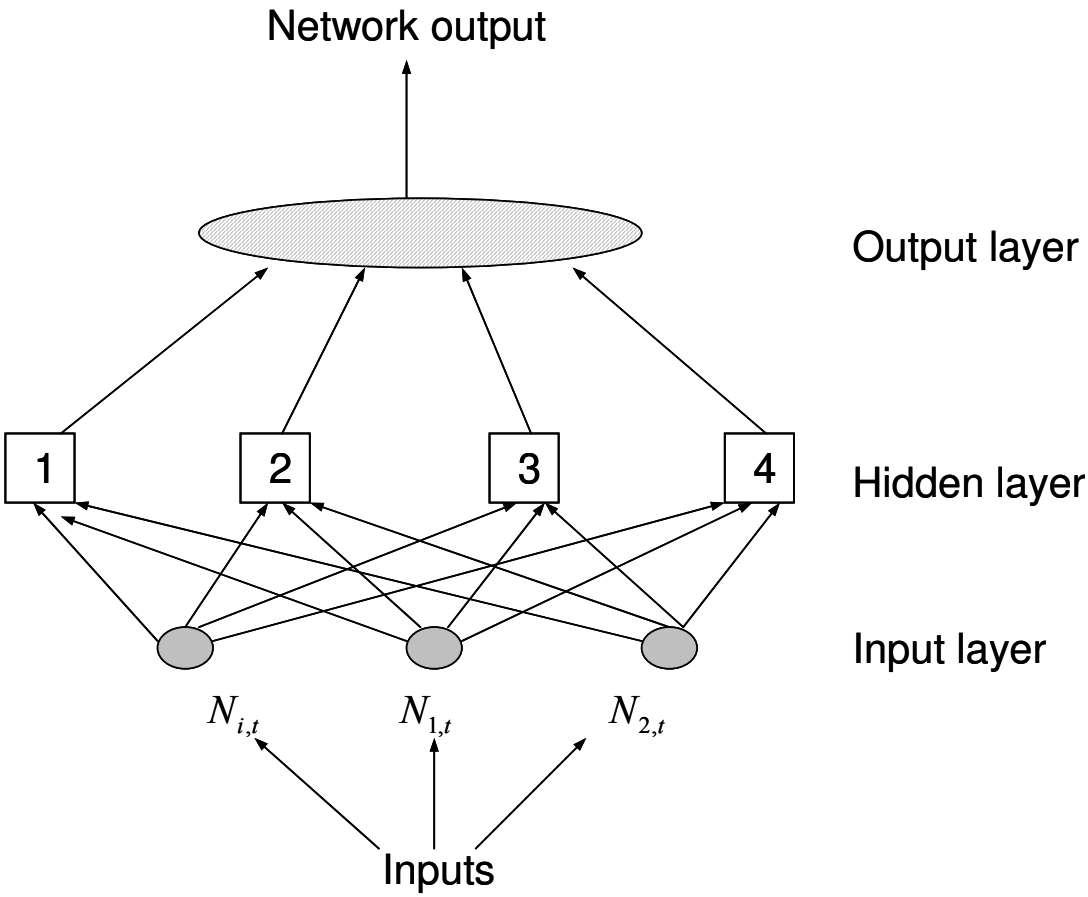

Figure S5. Architecture of the neural network model. In this example, the input layer consists of 3 input units representing the population abundances of focal species *i* and two connected species at time *t*, the hidden layer consists of 4 hidden units, and the output layer predicts the new population abundance of focal species *i* at time $t+T$.

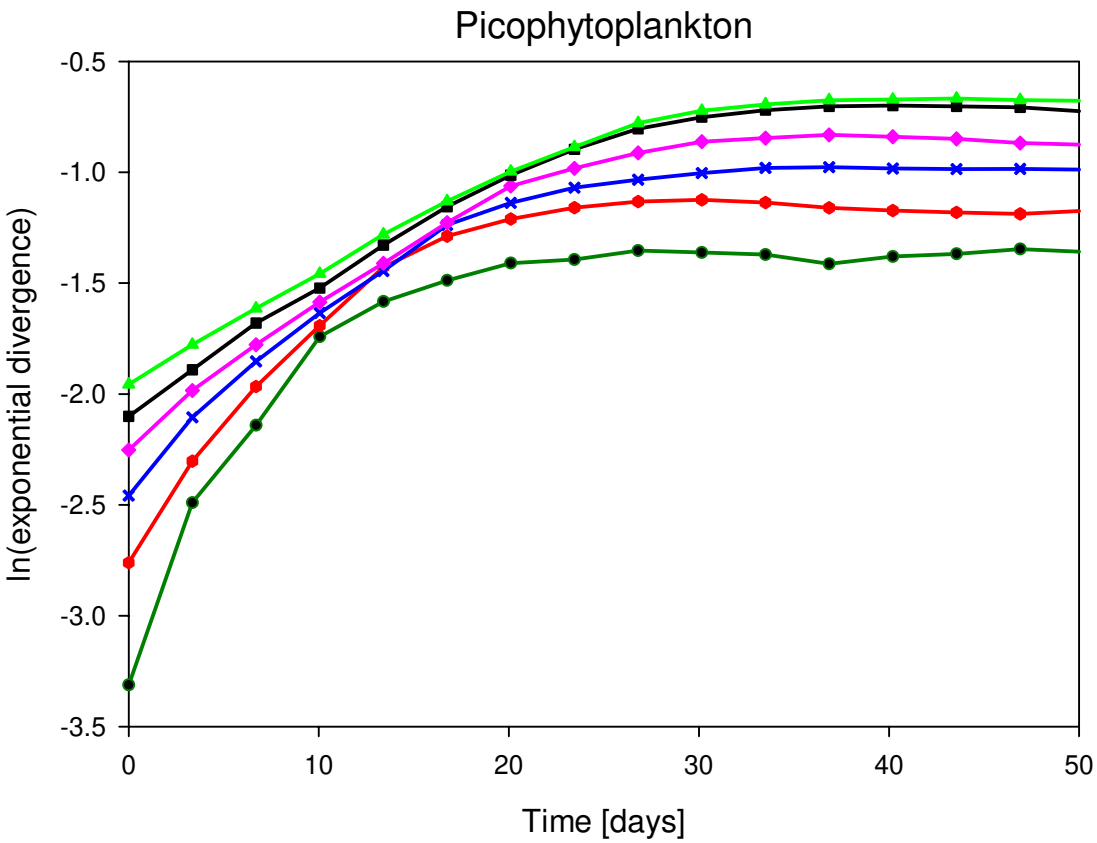

Figure S6. Exponential divergence of the trajectories of picophytoplankton as a function of time, calculated with different embedding dimensions (m=3, dark green; m=4, red; m=5, blue; m=6, pink; m=7, black; m=8, light green). Exponential divergence is plotted on a natural-logarithmic scale. Robust estimates of the Lyapunov exponent require that the initial slope of the exponential divergence is independent of the exact value of the embedding dimension. This requirement is fulfilled with an embedding dimension of $m \geq 6$.

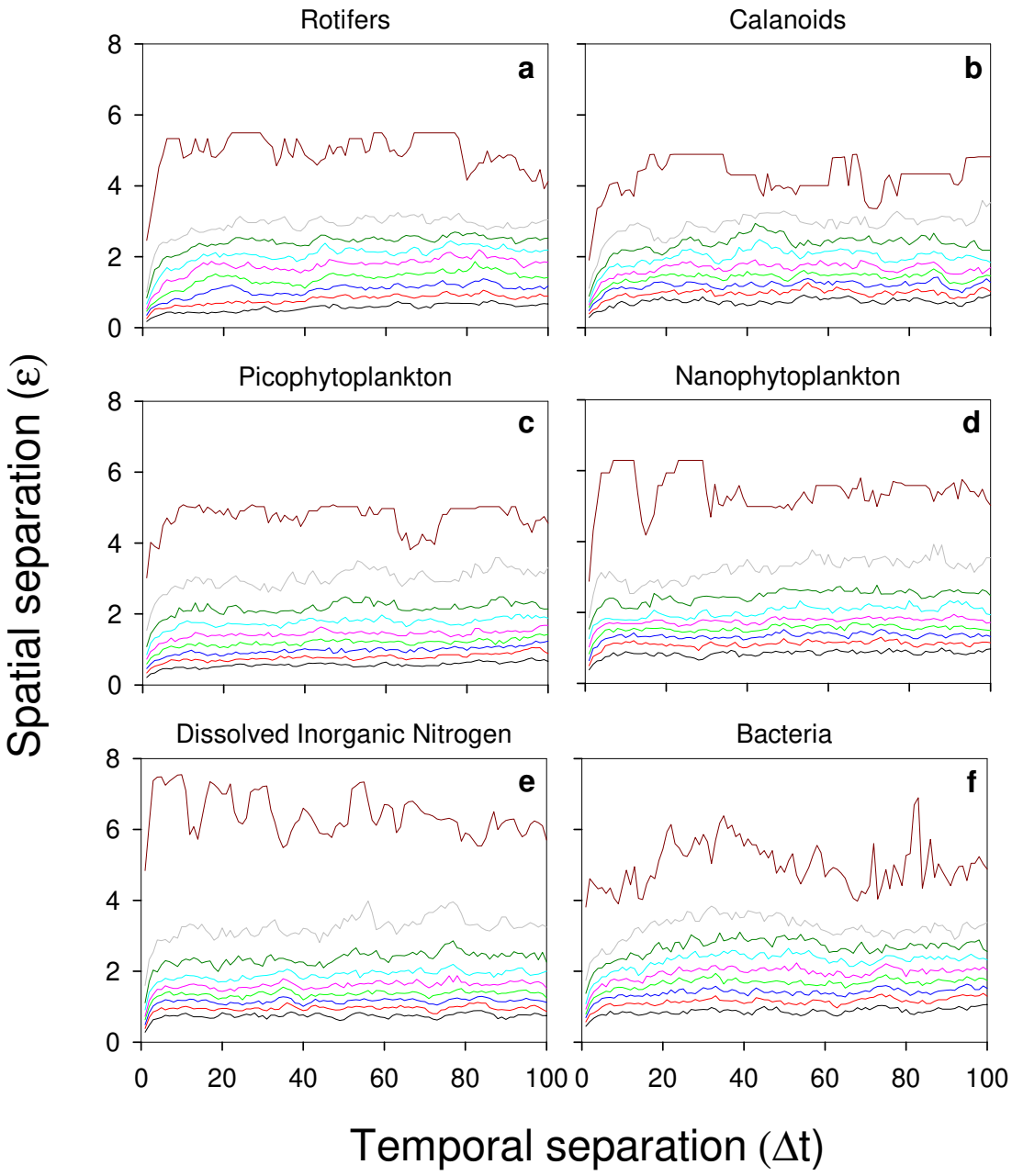

Figure S7. Space time separation plots of the functional groups. **a,** Rotifers; **b,** calanoid copepods; **c,** picophytoplankton**; d,** nanophytoplankton; **e,** dissolved inorganic nitrogen; **f**, bacteria. The plots indicate how the temporal distance between pairs of data points from the time series affects their spatial distance on the reconstructed attractor. Contour lines are shown at the spatial distance ε where for a given temporal separation Δt (in time steps) a fraction of 1/10, 2/10, … (lines from below) of pairs are found.

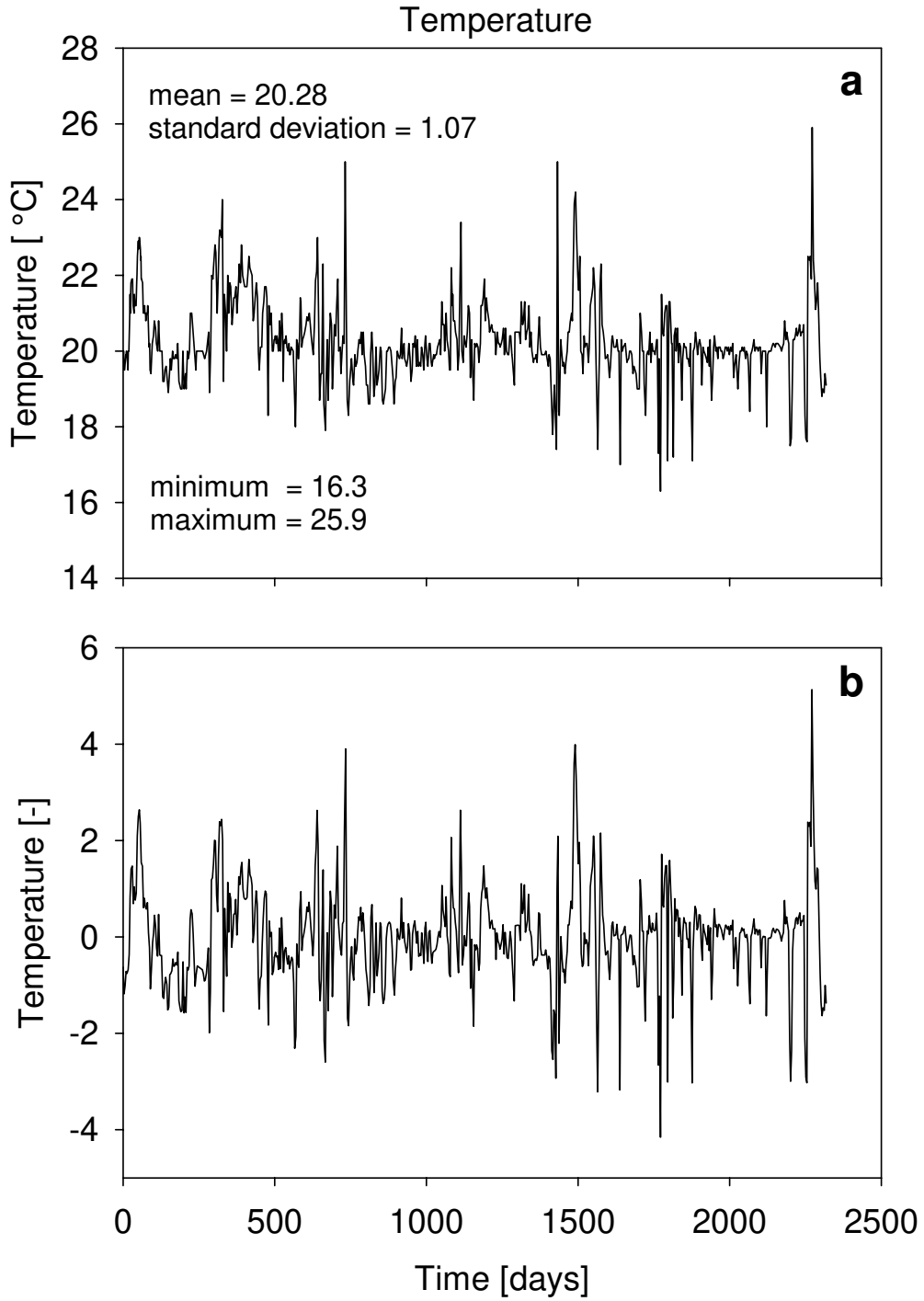

Figure S8. Time series of the temperature in the mesocosm experiment. **a**, Original time series. **b**, Stationary time series after data transformation.