

# Quantifying the effects of dynamical noise on the predictability of a simple ecosystem model

Barbara A. Bailey<sup>1,\*†</sup>, Scott C. Doney<sup>2</sup> and Ivan D. Lima<sup>2</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign, Champaign, IL 61820, U.S.A.

<sup>2</sup>Woods Hole Oceanographic Institution, Department of Marine Chemistry and Geochemistry MS 25, Woods Hole, MA 02543, USA

## SUMMARY

The need to understand the effects of anthropogenic perturbations on ocean biology has renewed interest in ecosystem and biogeochemical models in recent years. We develop a nonlinear time series approach to quantify the effects of different types of noise on ecosystem dynamics. Different types of noise can alter the local predictability of the system, induce qualitative regime shifts in model dynamics, and destroy (create) internal nonlinear oscillations and chaos. The ecosystem model we use in our article is a model of plankton dynamics and nitrogen cycling. It is a compartmental model (NPZD) consisting of compartments for nitrogen (N), phytoplankton (P), zooplankton (Z), and detritus (D). The flows or intercompartmental exchanges are modeled as a nonlinear system of four first-order differential equations. The types of noise of interest are both independent and correlated. Because the noise is an integral part of the system's dynamics, a nonlinear time series approach is used to quantify the dynamics and predictability of the system. This involves fitting neural network models and estimating dynamical system quantities of interest such as global and local Lyapunov exponents, along with measures of uncertainty for these estimates. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: Lyapunov exponents; nonlinear time series; neural network models

## 1. INTRODUCTION

Marine ecosystems play numerous and important roles in the earth system, ranging from biogeochemical cycling and carbon storage to biodiversity and fisheries. The behavior of these biological systems is of significant societal concern, in part because of the increasing rates of human environmental perturbations either directly via, for example, excess nutrient release and management or indirectly through global climate change. Our present inability to predict the ocean response to and feedbacks on these perturbations has sparked renewed interest in numerical simulation of ocean biology and biogeochemistry (Doney, 1999).

Current ocean ecosystem models typically consist of a small number (4–12) of highly aggregated compartments with biomass and biological flows formulated in terms of Nitrogen (Fasham *et al.*, 1990). This class of simulation models, which has been used extensively in oceanography (Steele,

---

\*Correspondence to: Barbara A. Bailey, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, U.S.A.

†E-mail: babailey@stat.uiuc.edu

1974) and theoretical ecology (May, 1973; Case, 2000), can exhibit a variety of nonlinear, intrinsic dynamics including limit cycles and chaos (Hastings and Powell, 1991; Popova *et al.*, 1997; Ryabchenko *et al.*, 1997). Such models, however, are still only incomplete representations of the richness and complexity of actual systems. One approach is to increase the number of model compartments, but this brings other difficulties associated with the development of realistic parameterizations, model–data evaluation, and simulation analysis. Marine ecosystems are also strongly influenced by physical environmental forcing (e.g. storms, mesoscale turbulence), and the relative importance of intrinsic and extrinsic variability is still a matter of some debate (Ascioti *et al.*, 1993). Here we investigate the effects of different types of noise (dynamical and environmental) on the behavior and predictability of a simple ecosystem model.

The ecosystem model we use is one developed by Lima *et al.* (2002) that, while simple, accounts for most of the fundamental aquatic ecosystem processes—plankton dynamics, trophic interactions, and nutrient limitation and recycling. The model consists of a nonlinear system of four first-order differential equations for the flows or intercompartmental exchanges among nutrients, phytoplankton, zooplankton, and detritus. The coupled set of equations will be integrated numerically in time, with noise added at each time-step either to the state variables or model parameters.

Because the noise becomes an integral part of the system, the data therefore can no longer be analyzed by deterministic methods but requires the statistical methods of nonlinear time series modeling. The global and local Lyapunov exponents will be used to quantify the dynamics of the system of interest. These two quantities will be used to compare the system's long-term predictability and short-term behavior. A primary focus of the analysis will be on the potential for noise to induce qualitative shifts in model dynamics, to alter the local predictability of the system, and to destroy (create) internal nonlinear oscillations and chaos.

An improved understanding of relationships between internal (biological) and external (physically forced) variability in the marine environment may lead to the development of a new class of models that provide a more robust response across seasons, biomes, and geographic areas. The understanding of a system's response to different types of noise and the development of a class of stochastic models is proposed not as a replacement, but as a complement to the need for the development of more complex deterministic biogeochemical models.

In this article we present a nonlinear time series approach to quantifying the predictability of noisy nonlinear systems. The biogeochemical model used for the study is introduced in Section 2. The data generation process along with the different types of noise added to the system is described in Section 3. Section 4 describes the estimation of the dynamical system quantities used to quantify the predictability of the system. Finally, Section 5 is a discussion of the results and future work.

## 2. ECOSYSTEM MODEL

The ecosystem model framework we use in our investigation consists of a system of nonlinear first-order differential equations in which the compartments are expressed as their equivalent scalar concentrations of nitrogen (Fasham *et al.*, 1990), a key limiting nutrient in aquatic systems. The model represents plankton trophic interactions and nutrient cycling. A complete description of the biological model is given in Lima *et al.* (2002). Thus, here we restrict ourselves to a brief discussion of its main characteristics. The NPZD model has four basic compartments: nutrients ( $N$ ), phytoplankton ( $P$ ), zooplankton ( $Z$ ) and detritus ( $D$ ) (mMol nitrogen  $m^{-3}$ ). The flows or intercompartmental exchanges (mMol nitrogen  $m^{-3} d^{-1}$ ) are modeled as a nonlinear system of four first-order differential equations:

$$\frac{dN}{dt} = a(1 - m)gG(P, D)Z - U(I, N)P + eD \quad (1)$$

$$\frac{dP}{dt} = U(I, N)P - gG(P, D)Z - sP \quad (2)$$

$$\frac{dZ}{dt} = amgG(P, D)Z - dZ^2 \quad (3)$$

$$\frac{dD}{dt} = (1 - a)gG(P, D)Z - gG(P, D)Z + sP + dZ^2 - eD \quad (4)$$

where  $a$ ,  $g$ ,  $s$ ,  $m$ ,  $d$ , and  $e$  are model parameters. Phytoplankton growth ( $U(I, N)$ ) is controlled by light ( $I$ ) and nutrient concentration ( $N$ ) and losses through grazing ( $-gG(P, D)Z$ ) and natural mortality ( $-sP$ ). Zooplankton growth is a function of total food availability ( $G(P, D)$ )—phytoplankton and detritus. In the zooplankton predation functional response, the preference terms for the different food sources change as a function of the relative proportion of the available food. The mathematical forms of the phytoplankton growth function and the zooplankton predation functional response are described in detail in Lima *et al.* (2002).

The total amount of nitrogen in the system  $N_o = N + P + Z + D$ , which is constant for unperturbed conditions, and the light availability  $I$  are the primary factors controlling the dynamics of the system. Nitrogen is regarded as a main limiting nutrient of primary production and therefore the biomass in marine ecosystems; the model could be equivalently formulated in terms of phosphorus for freshwater systems. The dynamics of different formulations of the model described above have been investigated with respect to variations in light intensity and the total amount of nitrogen (Lima *et al.*, 2002). In general, the model has asymptotically stable equilibrium points at low total nitrogen concentrations (oligotrophic environment) and high to moderate light intensities and oscillates in a limit cycle at low light intensities and high total nitrogen concentrations. Equilibrium values and the boundaries of the oscillatory regime can be seen in Figure 1.

The parameter values and definitions for the model are given in Table 1. Values for most of the model parameters are available from the literature (e.g. Eppley, 1980; Cullen, 1990; Platt *et al.*, 1992; Harrison, 1992; Hansen *et al.*, 1997; Geider *et al.*, 1997). The parameters for which there are few observational data (e.g. phytoplankton natural mortality rate) can be varied to match nutrient and phytoplankton measurements from observation stations, or chosen for consistency with previous modeling studies (Fasham *et al.*, 1990; Doney *et al.*, 1996). For a more complete discussion of the choice of values for the parameters see Lima *et al.* (2002).

### 3. GENERATING DATA

Equations (1–4) can be solved by numerically integrating the system to obtain  $N_t$ ,  $P_t$ ,  $Z_t$ , and  $D_t$ . This solution of the discretized system is the data generation process. The dynamical noise of interest is added at each integration time step. The investigation of different types of noise on the system behavior will be generalized as the following model in vector form:

$$Y_t = f(Y_{t-1}; \eta + g_t) + \varepsilon_t \quad (5)$$

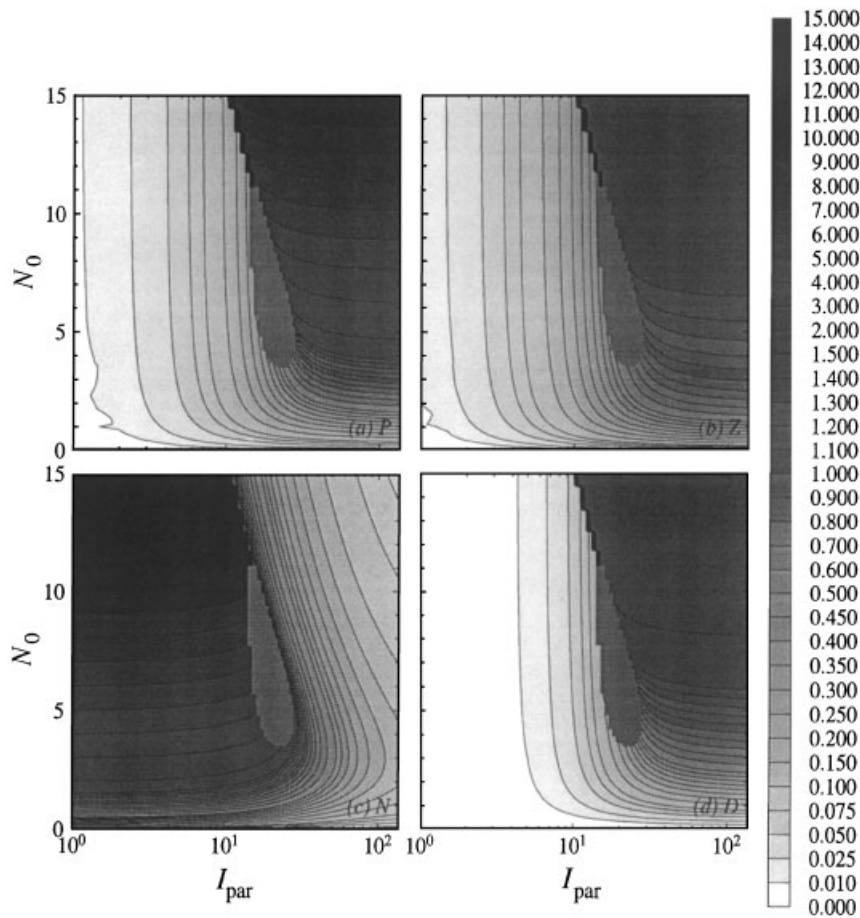


Figure 1. Equilibrium values of the different compartments in the NPZD model with fixed preferences as function of total nitrogen in the system ( $N_0$ ) and light (photosynthetically available radiation,  $I$ ). (a) phytoplankton, (b) zooplankton, (c) dissolved nutrients, (d) detritus. Units are  $\text{W m}^{-2}$  ( $I$ ) and  $\text{mMol N m}^{-3}$  ( $N_0$ ). Light variation (horizontal axis) is presented in logarithmic scale and shaded gray area indicates where system oscillates in a limit cycle

where  $Y_t$  is the state vector  $(N_t, P_t, Z_t, D_t)$ ,  $\eta$  is the vector of model parameters and  $f$  is the dynamical operator for the discretized set of equations. The types of noise of interest are  $\varepsilon_t$ , which are added to each of the state variables of  $Y_t$ , and  $g_t$ , which are added to the model parameters.

The model noise term  $\varepsilon_t$  is assumed to be an independent, Gaussian random variable with mean zero and variance  $\sigma_\varepsilon^2$ . It may be more realistic to adjust the noise so that it is proportional to the magnitude of each state variable and so this case will also be considered.

The parameter noise  $g_t$  added to  $\eta$  is independent and can be correlated over time, to represent, for example, lower frequency variations in external forcing. The process  $g_t$  used in each integration time step is defined by the equation

$$g_t = \phi g_{t-1} + a_t \quad (6)$$

Table 1. Parameter values and definitions for biological model

Parameter	Value (units)	Definition
$I$	variable ( $\text{W m}^{-2}$ )	Irradiance (light intensity)
$\theta$	0.45	PAR fraction of total irradiance
$\alpha$	$0.025 \text{ (mMol N (mg Chl)}^{-1} \text{ day}^{-1} \text{ m}^2 \text{ W}^{-1})$	Initial slope of the $P - I$ curve
$u_{\max}$	$2.0 \text{ (day}^{-1})$	Phytoplankton maximum growth rate
$g$	$1.0 \text{ (day}^{-1})$	Zooplankton maximum growth rate
$s$	$0.01 \text{ (day}^{-1})$	Phytoplankton mortality term
$d$	$0.1 \text{ ([mMol N m}^{-3} \text{ day]}^{-1})$	Zooplankton mortality
$e$	$0.25 \text{ (day}^{-1})$	Remineralization rate for detritus
$a$	0.8	Zooplankton assimilation efficiency
$m$	0.25	Zooplankton metabolic efficiency
$K_P$	$0.6 \text{ (mMol N m}^{-3})$	Half-saturation constant for phytoplankton
$K_Z$	$1.0 \text{ (mMol N m}^{-3})$	Half-saturation constant for zooplankton
$[chl : N]$	$1.0 \text{ (mg Chl (mMol N)}^{-1})$	Chlorophyll to nitrogen ratio
$\phi_P$	0.8	Preference of zooplankton for phytoplankton
$\phi_D$	0.2	Preference of zooplankton for detritus

where  $\phi$  is the AR(1) model parameter and  $a_t$  is assumed to be an independent Gaussian random variable with mean zero and variance  $\sigma_a^2$ . For this study,  $\eta$  is the light dependent growth parameter  $I$ .

A third form of noise, measurement or sampling noise, could be added to the state variables after the fact but without altering the model trajectory. Sampling noise would complicate somewhat the neural net model fitting (Section 4), increasing the resulting model residuals, but would not alter the underlying dynamics of the ecosystem model. Treatment of sampling error is essential for the analysis of actual observational data, which we reserve for a follow-on paper, focusing here on the idealized model system.

Figure 2 is a set of time series plots. The top panel is the unperturbed (no noise) solution to the system. For this study, the light parameter ( $I = 16.5 \text{ W/m}^2$ ) and the total amount of nitrogen ( $N_0 = 9 \text{ mMol m}^{-3}$ ) are chosen in the range where the system is unstable and oscillates in a limit cycle. This can be seen in the regions where there are irregular contour lines in Figure 1. These conditions are not atypical for the subsurface chlorophyll maximum near the base of the euphotic zone.

The next four panels are time series of a sample size of 1000 days generated by adding different types of noise to the unperturbed conditions of Figure 2(a). In each case, the noise is constant over a single day of model integration, one day being the discretization for (5) and (6); the ecosystem model itself is solved with an adaptive time-step, fourth-order Runge-Kutta scheme (Press *et al.*, 1992). The magnitude of noise is chosen to induce qualitative changes in the series. The effects of increasing the magnitude of noise are also considered. The same sequence of Gaussian i.i.d. noise is applied to all of the cases. The datasets are described below.

#### Data:

1. Noise is added to each state variable ( $N, P, Z, D$ ). The noise,  $\varepsilon_t$  is i.i.d.  $N(0, \sigma_\varepsilon^2)$  with  $\sigma_\varepsilon = 0.002$ .
2. Noise is added to each state variable. The noise is proportional to the magnitude of the state variable. The noise is  $\varepsilon_t Y_{t-1}$ , where  $\varepsilon_t$  is i.i.d.  $N(0, \sigma_\varepsilon^2)$  with  $\sigma_\varepsilon = 0.002$ .
3. Noise is added to the parameter  $I$ . The noise,  $g_t = a_t$  ( $\phi = 0$ ), is i.i.d.  $N(0, \sigma_a^2)$  with  $\sigma_a = 0.025$ .
4. Noise is added to the parameter  $I$ . The correlated noise is  $g_t = \phi g_{t-1} + a_t$  with  $\phi = 0.7$  and  $a_t$  is i.i.d.  $N(0, \sigma_a^2)$  with  $\sigma_a = 0.0178$  ( $\sigma_g = 0.025$ ).
5. Identical to data case 1 except that  $\sigma_\varepsilon = 0.010$ .

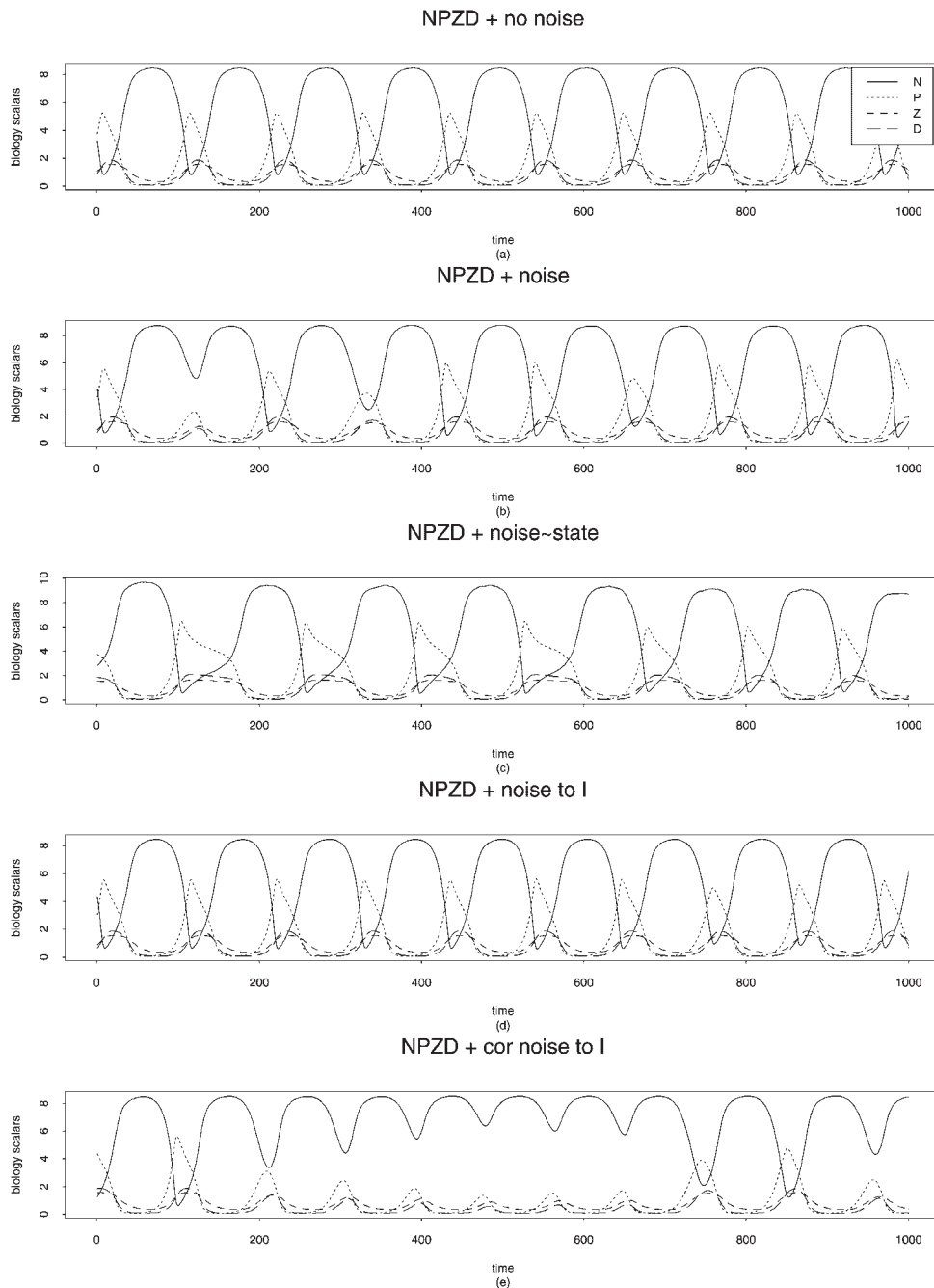


Figure 2. (a) Time series of NPZD generated data with no noise. (b) Time series of NPZD generated data with i.i.d. noise added to the state variables. (c) Time series of NPZD generated data with i.i.d. noise added to the state variables, which is proportional to the state variables. (d) Time series of NPZD generated data with i.i.d. noise added to the light parameter  $I$ . (e) Time series of NPZD generated data with correlated noise added to the light parameter  $I$

Note that the manner in which the noise is added to the state variables in data cases 1 and 2 does not ensure conservation of  $N_0$ . Some of the observed behavior, particularly in the  $\sigma_\varepsilon = 0.01$  case, is due to a shift of  $N_0$  outside the limit cycle regime (Figure 1). Perturbations to the system total nitrogen reflect a number of realistic processes including turbulent mixing, vertical migration (zooplankton), and particle sinking/remineralization. Alternatively, a focus strictly on the effect of model noise on the internal partitioning of nitrogen among the compartments could be constructed by setting  $\varepsilon_t$  for nutrients equal to the negative of the sum of the noise for the other three compartments.

#### 4. ESTIMATING DYNAMICAL SYSTEMS QUANTITIES

In the systems described by (5), the noise is an integral part of the system's dynamics, affecting how the state variables change over time. The data generated by numerically integrating the differential equations over time with added noise can no longer be analyzed by deterministic methods but require the statistical methods of nonlinear time series modeling.

The nonlinear time series modeling of the data generated in Section 3 not only makes the estimation of interesting dynamical system quantities possible, but it also allows the construction of confidence intervals or a measure of uncertainty (Bailey, 1996).

The nonlinear time series models for the data generated in Section 3 are:

$$\begin{aligned} N_t &= F_1(N_{t-1}, P_{t-1}, Z_{t-1}, D_{t-1}; \theta_1) + e_{1,t} \\ P_t &= F_2(N_{t-1}, P_{t-1}, Z_{t-1}, D_{t-1}; \theta_2) + e_{2,t} \\ Z_t &= F_3(N_{t-1}, P_{t-1}, Z_{t-1}, D_{t-1}; \theta_3) + e_{3,t} \\ D_t &= F_4(N_{t-1}, P_{t-1}, Z_{t-1}, D_{t-1}; \theta_4) + e_{4,t} \end{aligned} \quad (7)$$

The estimation of each of the maps  $F_i$  of (7) and their corresponding nonlinear time series model parameters  $\theta_i$  will be necessary in order to estimate quantities of interest. This means that  $F_1, F_2, F_3$ , and  $F_4$  will be estimated for each of the data sets generated in Section 3.

This system in (7) can be written in vector form:

$$X_t = F(X_{t-1}; \theta) + e_t \quad (8)$$

where  $X_t$  is the state vector  $(N_t, P_t, Z_t, D_t)$ ,  $\theta$  is the vector  $(\theta_1, \theta_2, \theta_3, \theta_4)$ , and  $e_t$  are assumed to be independent Gaussian random variables with mean zero and variance  $\sigma_x^2$ . Note that for time series realizations generated by (5), the time series models of (7) are strictly true only for data generated in data case 1, where i.i.d. noise is added to each state variable. It will be necessary to assume that the underlying model is of the form of (8) to ensure the asymptotic normality of the nonlinear autoregression parameters and corresponding confidence intervals (Bailey, 1996). Checking the model assumptions will be an important part of model diagnostics.

The estimated global and local Lyapunov exponents will be used to quantify the dynamics of the systems of interest. The global Lyapunov exponent will describe the system's long-term predictability, and the distribution of local Lyapunov exponents will describe the system's short-term behavior. These two quantities are described as follows.

Consider a discrete dynamical system perturbed by noise. Assume that  $X_t, e_t \in \mathfrak{R}^d$  and that the time evolution of the  $\{X_t\}$  process follows the map of (8). Assume that  $F$  is continuously differentiable. The local Lyapunov exponent (LLE) is defined by making an infinitesimal perturbation of  $X_t$  at time  $t$  and following forward the perturbed and unperturbed trajectories or sample paths. The LLE measures the log-difference in the norms of the two trajectories after  $n$  time steps. For small perturbations, the linear approximation will be defined by the derivatives of the map  $F$  (Eckmann and Ruelle, 1985). Then

$$\lambda_n(t) = \frac{1}{n} \ln \|J_{n+t-1} J_{n+t-2} \cdots J_t U_0\| \quad (9)$$

where  $J_t$  is the Jacobian matrix of  $F$  evaluated at  $X_t$ ,  $U_0$  is a unit vector and  $\|\cdot\|$  is the Euclidean vector norm. Since  $\lambda_n(t)$  is a function of time, the LLE depends on the trajectory and can be thought of as an ‘ $n$ -step ahead’ local Lyapunov exponent process.

If  $U_0$  is chosen at random with respect to the uniform measure on the unit sphere, then with probability 1 as  $n \rightarrow \infty$ ,  $\lambda_n(t)$  converges to the global Lyapunov exponent,  $\lambda$  because  $U_0$  has zero probability of falling into the subspace corresponding to subdominant exponents. In practice, however, we usually take  $U_0 = (1, 0, 0, \dots, 0)$ . The global Lyapunov exponent measures the long-term average rate at which nearby trajectories of a system converge (diverge). For example, a positive LLE5 or 5-step-ahead LLE means that the system exhibits short-term (in 5 time steps or days) trajectory divergence or exponential divergence of nearby trajectories. A negative LLE5 means that there is short-term trajectory convergence.

The calculation of global and local Lyapunov exponents uses (9) and requires an estimate of the Jacobian matrix  $J_t$ . The map for each  $F_t$  in (7) is estimated as a feed-forward neural network with a single layer of hidden units (Ellner *et al.*, 1992). The form of the model is

$$F(X) = \beta_0 + \sum_{i=1}^k \beta_i \varphi(X^T \gamma_i + \mu_i) \quad (10)$$

where  $\varphi(u) = e^u / (1 + e^u)$ . The parameter vector estimates  $\theta_i = (\beta, \gamma, \mu)$  are found by least squares. The complexity of the model or the number of hidden units,  $k$ , is chosen based on generalized cross-validation (GCV). The estimated Jacobian is then the derivative of the estimated  $F$ . There are many other models to use to estimate  $F_t$  (for example, thin plate smoothing splines, projection pursuit regression), but the neural network has been shown to be a robust method where increasing the dimension of  $X$  does not degrade the accuracy of  $\lambda$  (Nychka *et al.*, 1992).

Again, it is important to note that the noise assumption for (8) of independent Gaussian random variables with mean zero and variance  $\sigma_X^2$  is for the purpose of asymptotic properties of model parameters and the construction of confidence intervals. Recall that data generated by the noise added to the system of (5) can be, for example, correlated. It is assumed that the procedure of fitting a neural network model to each of equations (7) is flexible enough to accommodate data in which noise is not independently distributed. That is, after fitting a neural network model to the data, the residuals from each fit should be approximately independent Gaussian random variables with mean zero and variance  $\sigma_X^2$ , regardless of how the data are generated. The analysis described in Section 5 shows that this is a reasonable assumption.

The LLEs for each fixed ‘step-ahead’ are a distribution of numbers. Therefore it will be necessary to calculate summary statistics, for example the median or other informative percentiles. The

comparison of distributions of LLEs is important because a comparison of different systems' global Lyapunov exponents is not enough to characterize the difference in systems dynamics and predictability. It is possible for systems to have the same global Lyapunov exponent, but very different local behavior (Bailey *et al.*, 1997).

## 5. ANALYSIS AND RESULTS

The unperturbed ecosystem time series (Figure 2a) exhibits large amplitude predator-prey oscillations with a periodicity of  $\sim 100$  days. Between the phytoplankton blooms, nutrient concentrations are high, and phytoplankton biomass is kept low and approximately uniform by zooplankton grazing and, to a lesser degree, natural mortality. Bloom initiation results when the gradual decline in zooplankton biomass and thus grazing pressure reaches a critical threshold where the zooplankton are no longer able to control phytoplankton growth. The maximum bloom phytoplankton concentrations are approximately an order of magnitude larger than the background, and the zooplankton and detrital concentrations show corresponding lagged increases. The bloom termination is governed by nutrient depletion (slowed phytoplankton growth), phytoplankton mortality, and zooplankton grazing. Adding modest levels of dynamical noise to the state variables or model parameters values (e.g. external forcing) alters the magnitude, timing and shape of the simulated phytoplankton blooms but does not change the fundamental character of the ecosystem time series (Figures 2 (b)–(e)).

Figures 3–6 show summary plots for the perturbed (noisy) time series, data cases 1–4, with diagnostics of the nonlinear time series model fits and estimates of the systems' long-term and short-term predictability. For each summary plot, the top panel (a) is the time series of the data generated as described in Section 3. The next two panels (b) and (c) are scatter plots of the actual  $N$  versus predicted  $N$  from the neural network fit to (7) and the corresponding  $N$  residuals versus predicted  $N$ , respectively. Panel (d) is the distribution of local LLEs for 5, 10, 20, 40 and 80 steps ahead. The solid line represents the estimated global Lyapunov exponent. Plot (e) is a plot of the LLE5s over time.

All summary plots indicate that the neural network is fitting the data well even with moderate levels of dynamical noise. For example, for the data case 1 time series, the estimated standard deviation of the residuals is 0.0019 compared to  $\sigma = 0.0020$ . Further examination of the residuals (not shown) indicates the i.i.d. Gaussian random variable assumption is also reasonable for this case. For data case 2, where the noise is proportional to concentration, the  $N$  residuals are considerably larger and scale with the predicted concentration as expected. The residuals for data cases 3 and 4 are intermediate between these two bounds.

The solid line in panel (d) for each of the fits represents the estimated global Lyapunov exponent ( $\hat{\lambda}_{\text{Fit}1} = 0.002$ ,  $\hat{\lambda}_{\text{Fit}2} = 0.004$ ,  $\hat{\lambda}_{\text{Fit}3} = 0.08$ ,  $\hat{\lambda}_{\text{Fit}4} = 0.17$ ). The estimated global exponents for data case 1, 2 and 3 are statistically indistinguishable from zero. The global exponent for data case 4 is statistically different from those from data cases 1 and 2 and zero at the  $\alpha = 0.05$  level. This indicates that correlated noise added to the light parameter significantly affects the long term predictability of the system. The positive exponent and corresponding 95% CI that does not include zero indicates a system that exhibits global sensitivity to initial conditions or chaos.

The statistical distributions and behavior of the local Lyapunov exponents LLEs also tend to differ among the four data cases. An examination of the distribution of the LLE5s or 5-step-ahead LLEs indicates that the LLE5s are similar in range ( $-0.1$ – $0.3$ ) except for data case 4, where the range is somewhat elevated ( $0.1$ – $0.45$ ). However, the median local Lyapunov exponent tends to decrease as the step-ahead time  $n$  is increased for data cases 1 and 2 but remains approximately flat for data cases 3

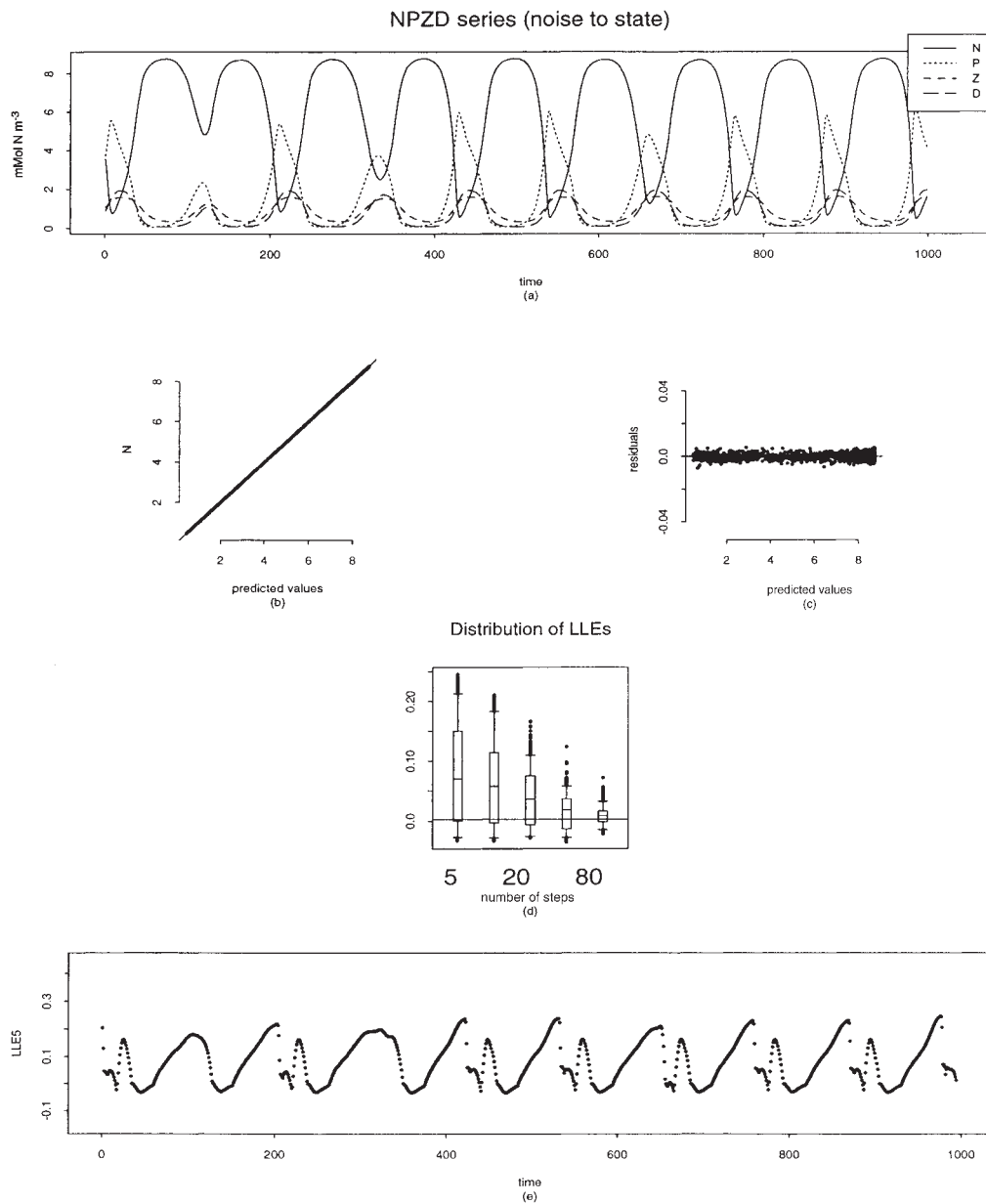


Figure 3. (a) Time series of NPZD generated data with i.i.d. noise added to the state variables. (b) Predicted values versus fitted values of the neural network fit to N. (c) Predicted values versus residuals of the neural network fit to N. (d) Distribution of LLEs for system of (a). (e) Time series of the 5-step-ahead LLEs for the system of (a)

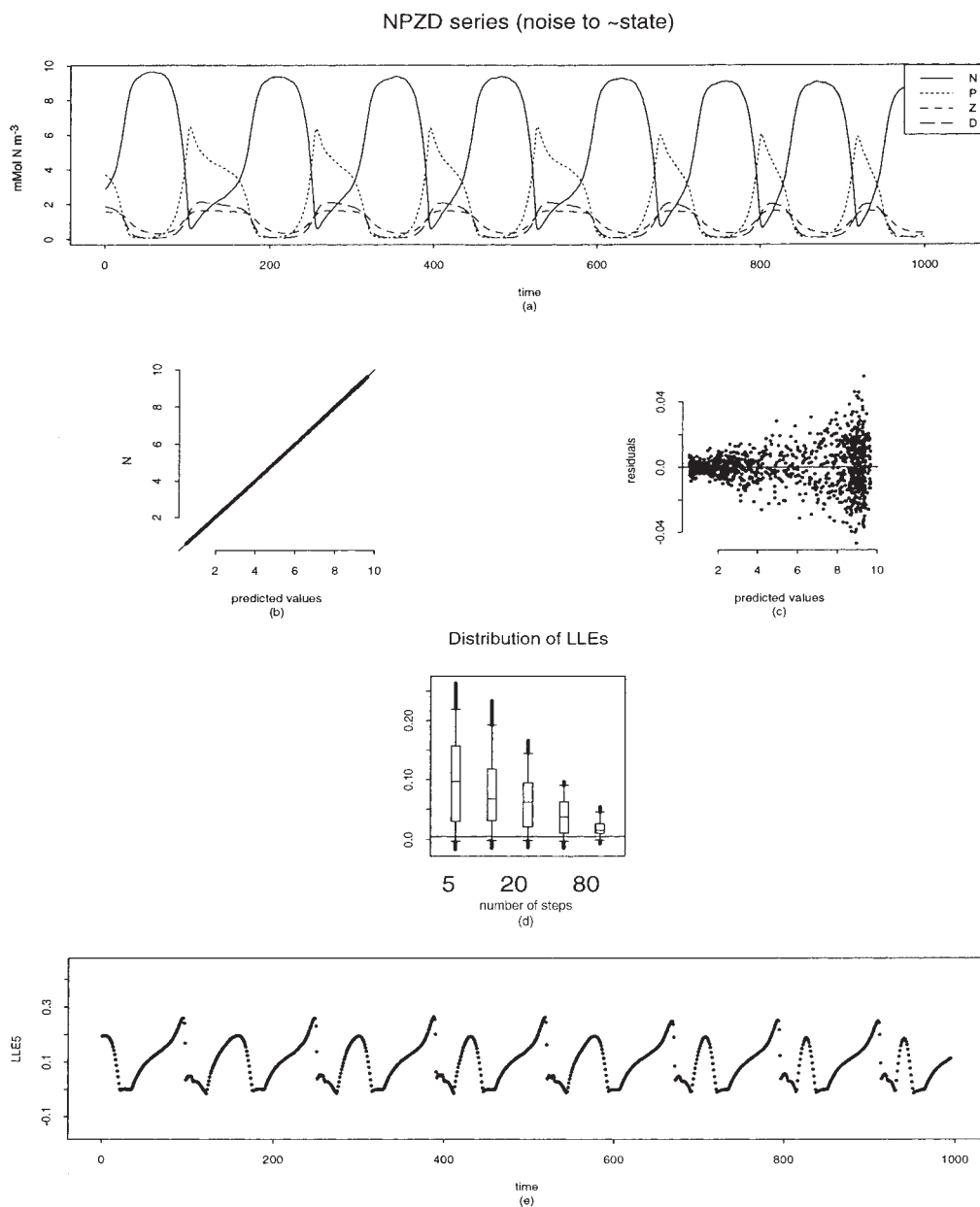


Figure 4. (a) Time series of NPZD generated data with i.i.d. noise added to the state variables, which is proportional to the state variables. (b) Predicted values versus fitted values of the neural network fit to N. (c) Predicted values versus residuals of the neural network fit to N. (d) Distribution of LLEs for system of (a). (e) Time series of the 5-step-ahead LLEs for the system of (a)

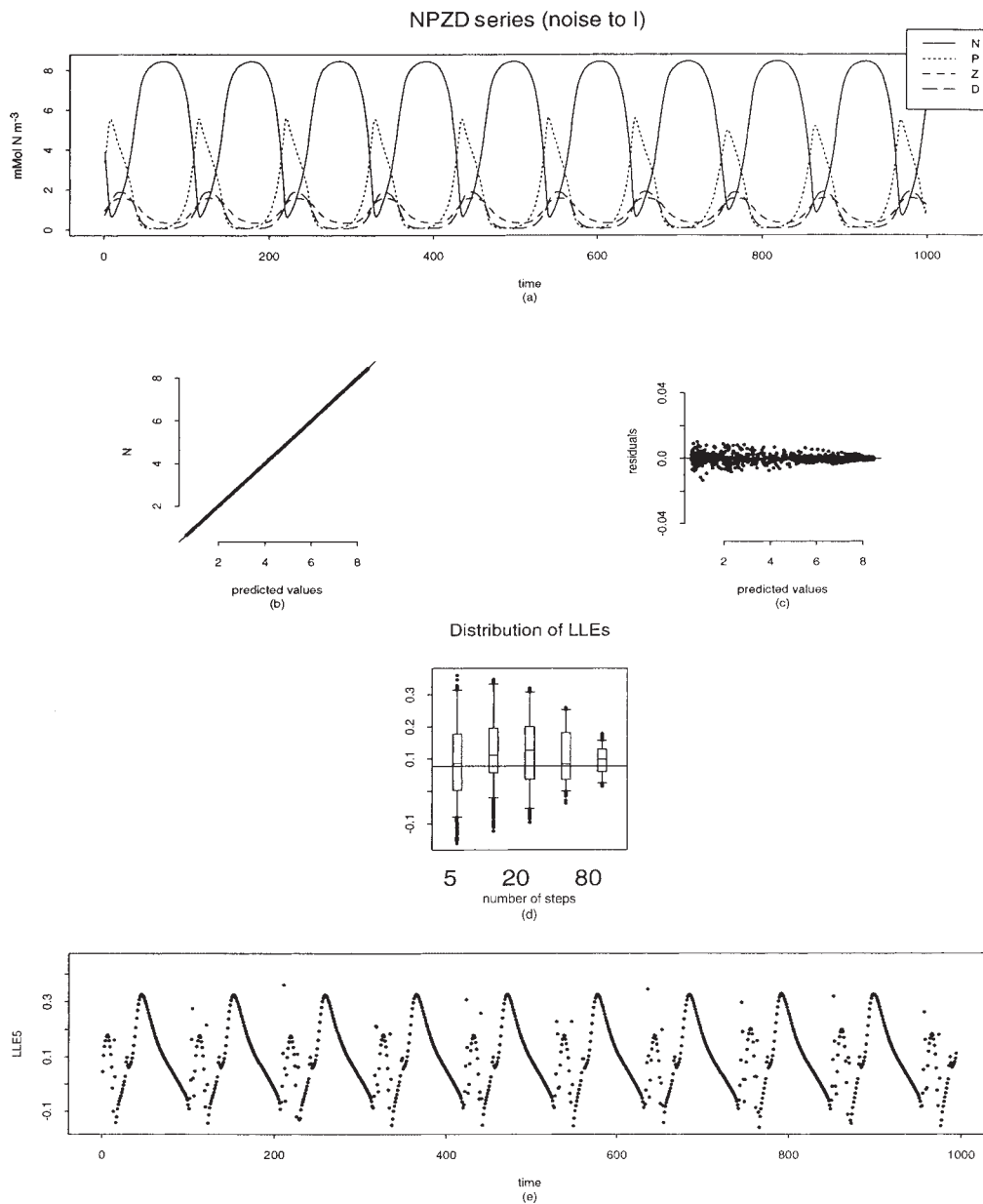


Figure 5. (a) Time series of NPZD generated data with i.i.d. noise added to the light parameter  $I$ . (b) Predicted values versus fitted values of the neural network fit to  $N$ . (c) Predicted values versus residuals of the neural network fit to  $N$ . (d) Distribution of LLEs for system of (a). (e) Time series of the 5-step-ahead LLEs for the system of (a)

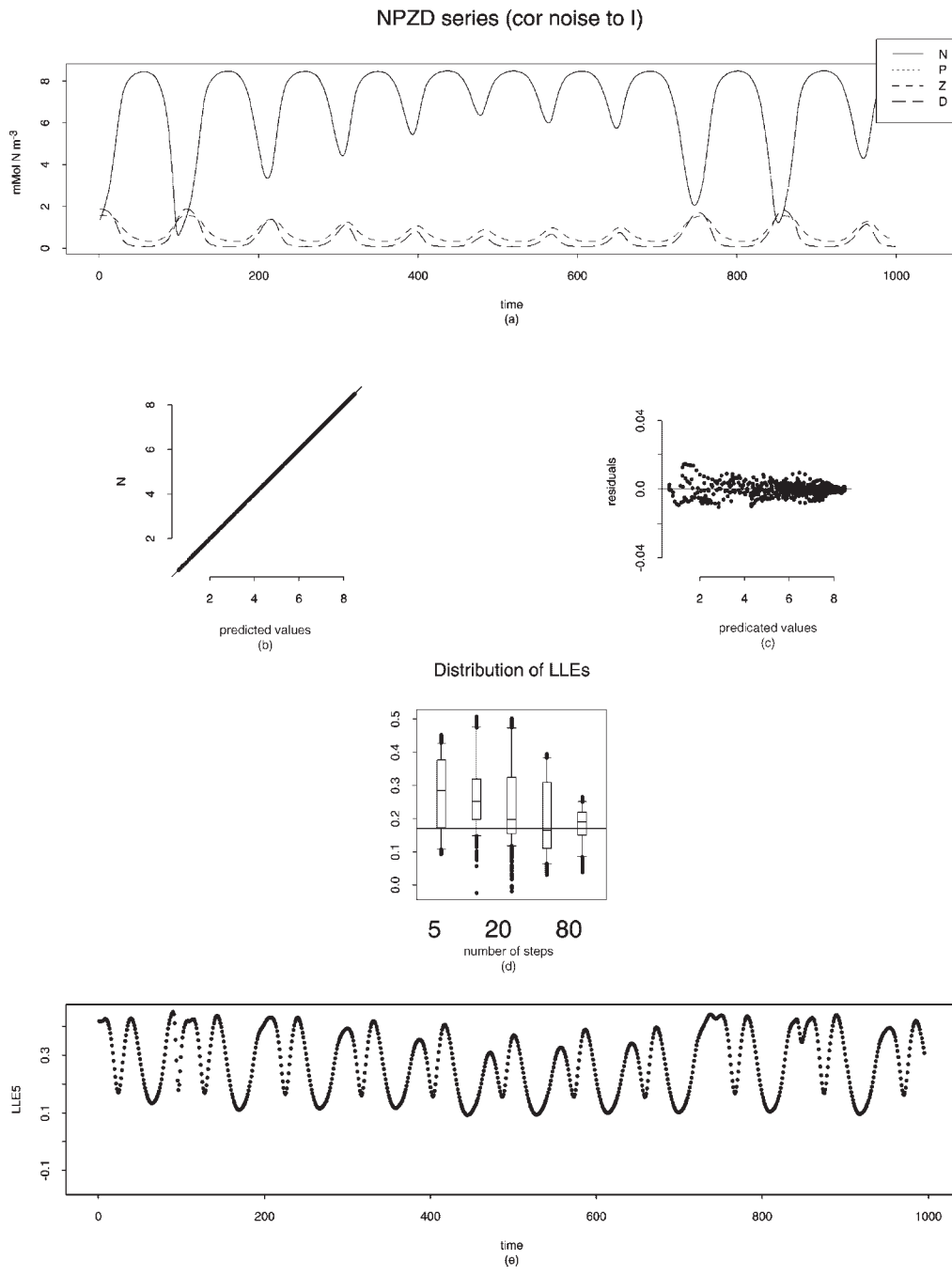


Figure 6. (a) Time series of NPZD generated data with correlated noise added to the light parameter  $I$ . (b) Predicted values versus fitted values of the neural network fit to  $N$ . (c) Predicted values versus residuals of the neural network fit to  $N$ . (d) Distribution of LLEs for system of (a). (e) Time series of the 5-step-ahead LLEs for the system of (a)

and 4. The plots of the LLE5s time series (Figures 3–6e) also show differences in the temporal patterns of predictability for the separate cases. Therefore, adding different types of noise to the system can qualitatively change the short-term predictability of the system. It should be noted that as the number of steps-ahead becomes large the distribution of local exponents will converge to a normal distribution (Bailey *et al.*, 1997). By 100-steps ahead, the distribution of LLEs is approximately normal, so that the comparison for LLEs is only valid for relatively small  $n$ .

The estimated LLEs for the fit to data case 1 are chosen to illustrate in more detail the patterns of predictability for a single system. The relationship of LLE5 to the simulated limit cycle is shown more clearly in Figures 7(b, c), scatter plots of LLE5 versus  $P$  and LLE5 versus  $dP/dt$ , respectively. Periods when the LLE5s are in the top 10% of the LLE5 distribution are identified (coded by 10s) on the time series of  $P$ , Figure 7(a) and the scatter plots.

The LLE time series (Figure 3e) consists of small, negative values (weak convergence, moderate predictability) during the termination of the bloom as the system stabilizes to a steady-state, low phytoplankton background. The LLE5 values then begin to increase and transition to rapidly rising, positive values at and following bloom initiation. The top 10% of the LLE5s are found during the middle of the bloom when  $dP/dt$  is near its maximum ( $0.1\text{--}0.4\text{ mMol N m}^{-3}\text{ d}^{-1}$ ) and are positive at the 95% confidence level. These are times, therefore, when there is short-term trajectory divergence and the largest unpredictability. In most but not all of the blooms, the LLE5 values decrease again towards the peak of the bloom and actually dip slightly negative as the bloom starts to decline. There is then another period of relative high LLE5 before the circuit begins again.

This rather complex pattern of local predictability has a relatively simple explanation. Near the termination and peak of the bloom, the system's trajectory is strongly constrained despite any noise by strong grazing pressure ( $Z \gg P$ ) or overall nutrient depletion (low relative  $P$  growth). In the two cycles when the bloom did not fully draw down the ambient nutrients (time  $\sim 100$  and 300 days) the minima in LLE5 are not observed near the bloom peak. Contrastingly, during the growth and decline phases of the bloom, predictability is low, and small amounts of noise can greatly alter the path of the system, for example changing the size of the peak or the timing of the termination. The same general patterns are found for data case 2, where the model noise is added proportional to the state variables.

The estimated LLE5s for the fits to data cases 1–4 are chosen to illustrate the qualitatively different short-term predictability of the systems observed by adding different types of noise to the system. Figure 8 is a series of scatter plots for estimated LLE5s versus  $P$ , LLE5s versus  $dP/dt$ , LLE5s versus  $1/P dP/dt$ , and  $P$  versus  $1/P dP/dt$  for fit to data case 1 (a)–(d), fit to data case 2 (e)–(h), fit to data case 3 (i)–(l), and fit to data case 4 (m)–(p). The quantity  $1/P dP/dt$  is the relative or fractional growth rate. The shapes of the phytoplankton limit cycle outlined by the  $P$  versus  $1/P dP/dt$  plots are quite similar across all four data cases. The local predictability patterns are also similar in that the LLE5 values are typically large (divergent) during the growth phase of the bloom, when  $1/P dP/dt$  is near a maximum. Striking differences exist in other aspects of the patterns, however, with distinct behavior among data case 1 + 2 (similar), data case 3, and data case 4. In particular, the presence of a secondary period of divergence during the collapse of the bloom is case dependent, as is the magnitude and, in some cases, strong variability of the LLE5s during the low plankton biomass period between blooms.

Model diagnostics and checking if the model is appropriate is an important part of the analysis, since the estimation and measures of uncertainty of the dynamical systems quantities is based on the model assumptions. To investigate the magnitude of noise that can be added to the system and still give reasonable neural network fits, a large amount of noise,  $\sigma_e = 0.01$  is added to each state variable to generate data case 5. This amount of dynamical noise forces the system in and out of the limit cycle regime by altering  $N_0$  as shown in Figure 9(a) with time series plots of  $(N, P, Z, D)$ . A larger sample

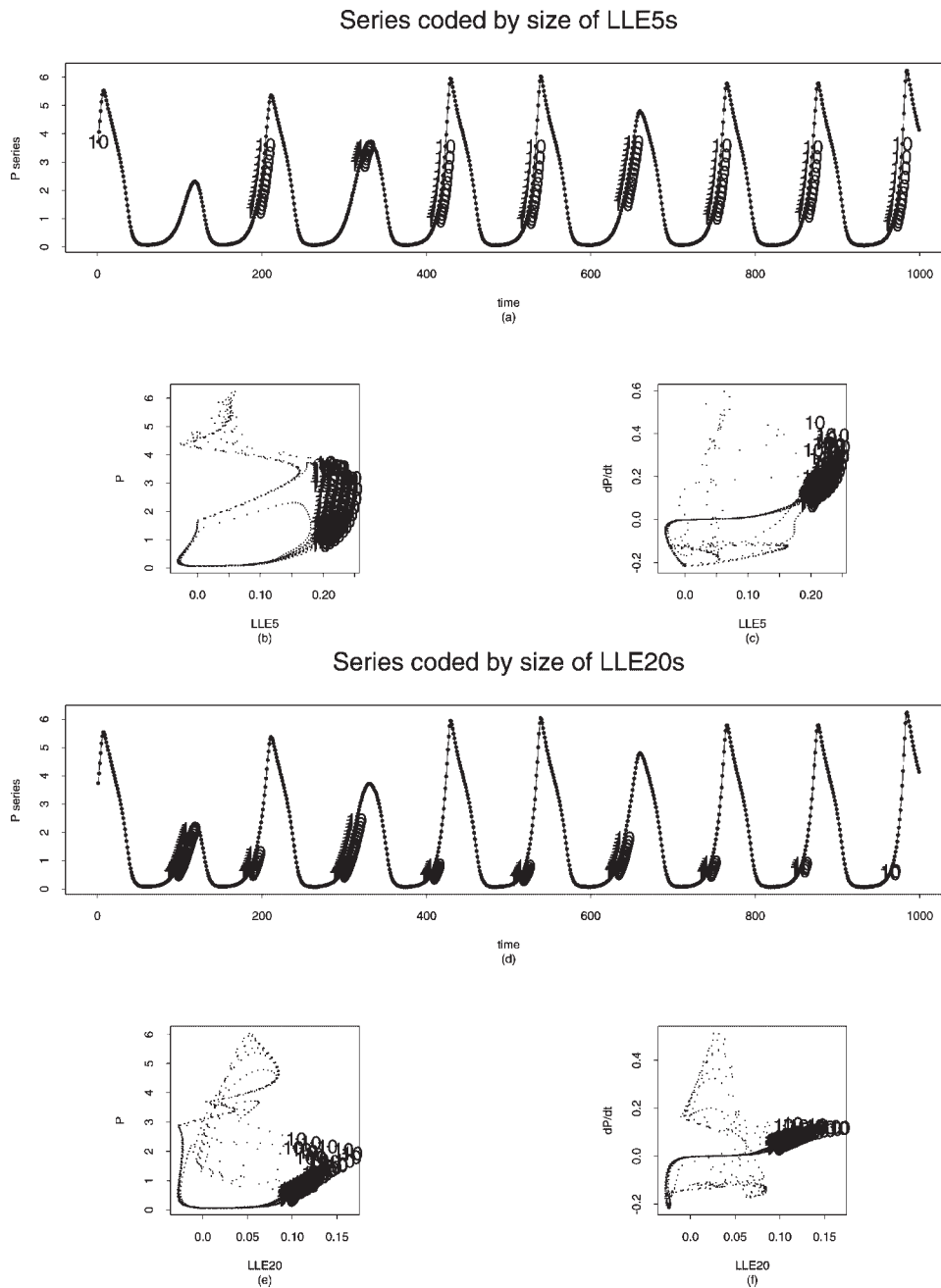


Figure 7. (a) Time series of  $P$  coded to identify the times at which the LLE5 is in the top 10% of the distribution of LLE5s. (b) LLE5 versus  $P$  coded to identify the LLE5s in the top 10% of the distribution of LLE5s. (c) LLE5 versus  $dP/dt$  coded to identify the LLE5s in the top 10% of the distribution of LLE5s. (d) Time series of  $P$  coded to identify the times at which the LLE20 is in the top 10% of the distribution of LLE20s. (e) LLE20 versus  $P$  coded to identify the LLE20s in the top 10% of the distribution of LLE20s. (f) LLE20 versus  $dP/dt$  coded to identify the LLE20s in the top 10% of the distribution of LLE20s

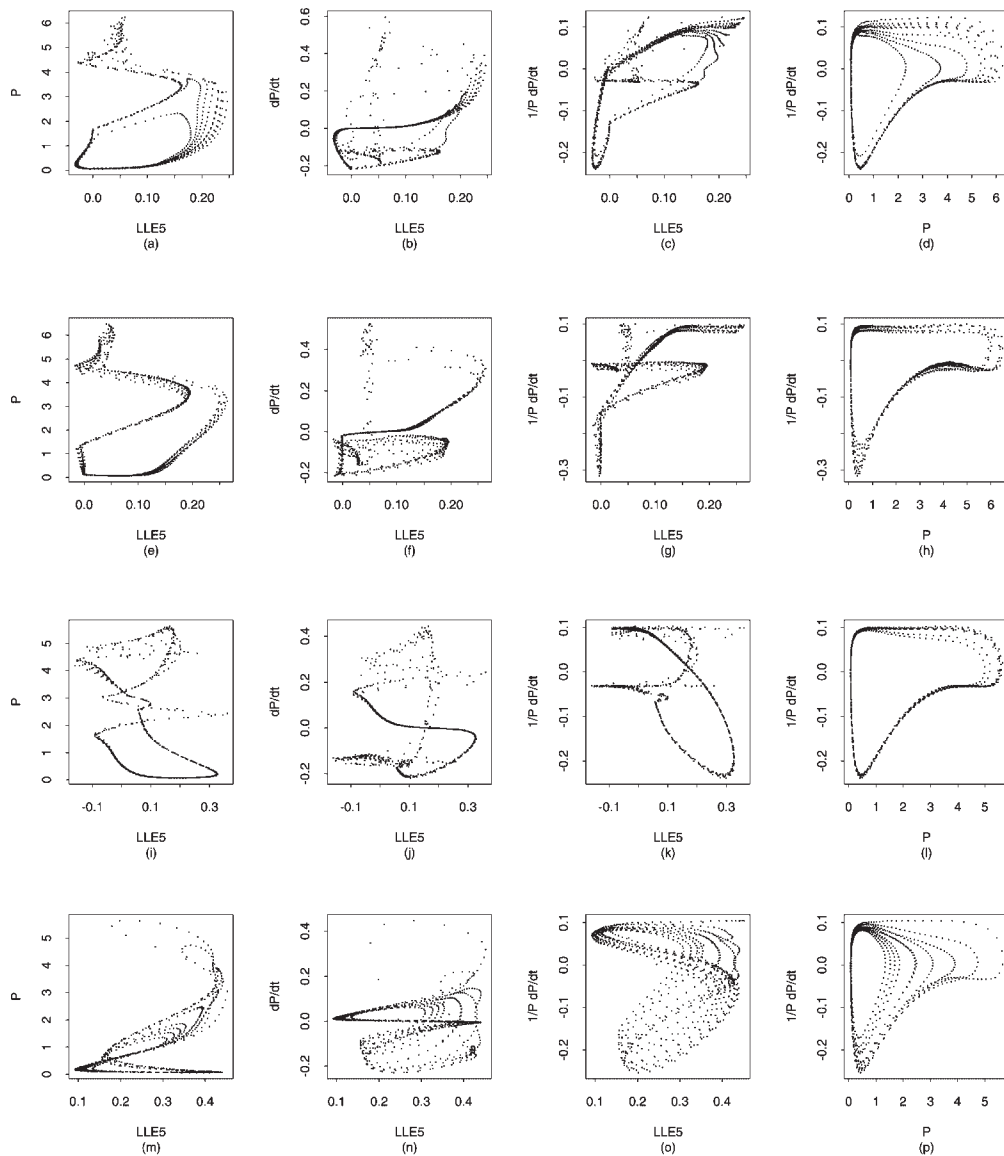


Figure 8. Fit 1: (a) LLE5 versus  $P$ ; (b) LLE5 versus  $dP/dt$ ; (c) LLE5 versus  $1/P dP/dt$ ; (d)  $P$  versus  $1/P dP/dt$ . Fit 2: (e) LLE5 versus  $P$ ; (f) LLE5 versus  $dP/dt$ ; (g) LLE5 versus  $1/P dP/dt$ ; (h)  $P$  versus  $1/P dP/dt$ . Fit 3: (i) LLE5 versus  $P$ ; (j) LLE5 versus  $dP/dt$ ; (k) LLE5 versus  $1/P dP/dt$ ; (l)  $P$  versus  $1/P dP/dt$ . Fit 4: (m) LLE5 versus  $P$ ; (n) LLE5 versus  $dP/dt$ ; (o) LLE5 versus  $1/P dP/dt$ ; (p)  $P$  versus  $1/P dP/dt$

size of 2000 is chosen so that more types of different dynamics can be analyzed. Note that the initial parameters for  $I$  and  $N_0$  were chosen to fall close to the boundary of the limit cycle regime. Figures 9(b)–(e) present the corresponding residuals of the fits to  $N$ ,  $P$ ,  $Z$ ,  $D$ , respectively, for data case 5. An examination of the residuals over time shows no significant patterns, indicating that the neural network models are robust to changes in dynamic behavior and to a significant level of dynamical noise.

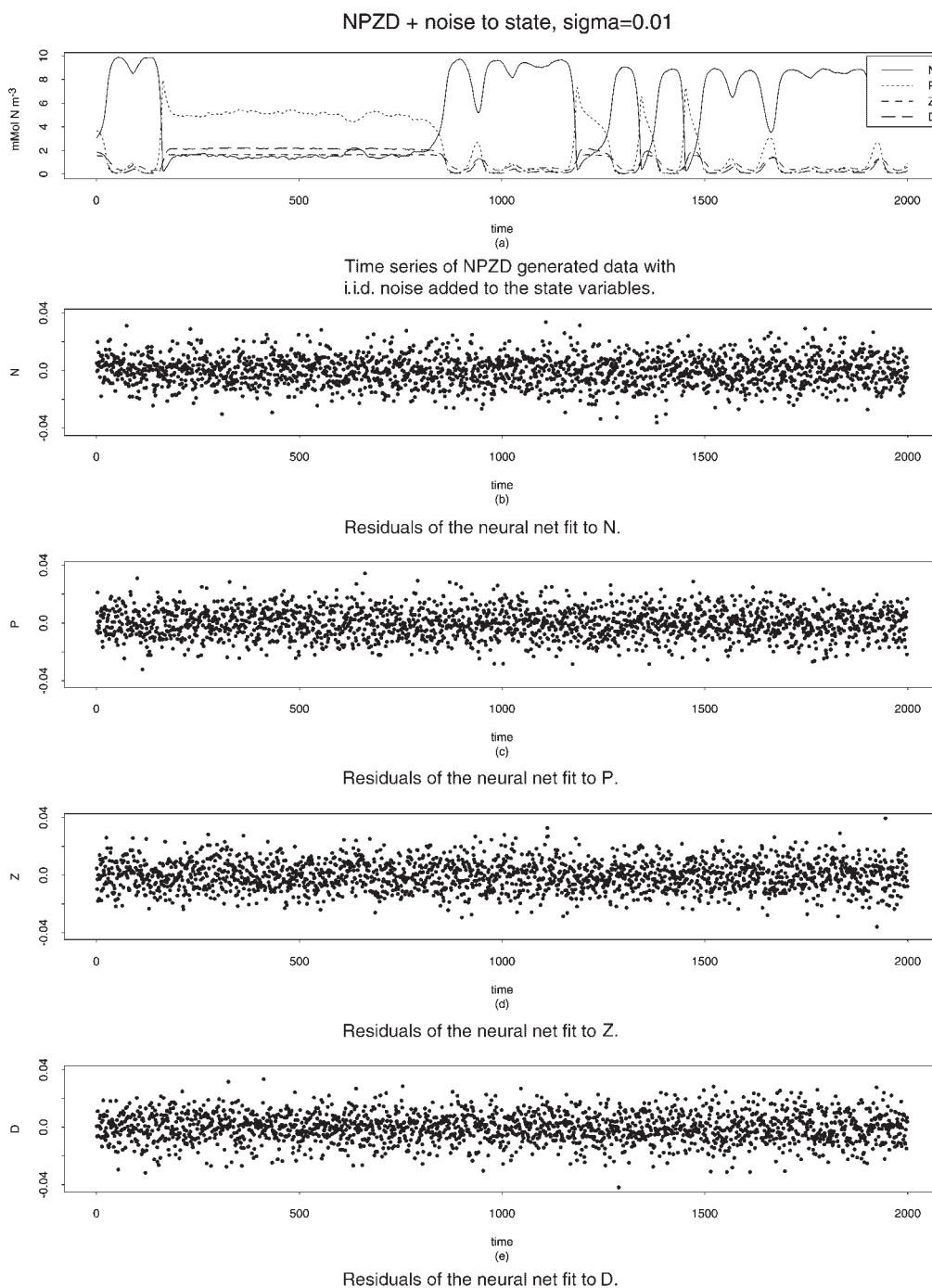


Figure 9.

## 6. DISCUSSION

The estimated global Lyapunov exponent, summary statistics of LLEs, and the LLE time series are useful diagnostics generated from our analysis. The estimated global Lyapunov exponent is a single number which describes the long-term predictability of the system of interest, and the LLEs quantify the system's local or short-term predictability. Our experiments suggest that these measures, particularly the LLE time series, can vary significantly depending on the nature of the noise applied to the system—this, despite the fact that the limit cycles exhibited in the time series and phase diagrams of the model variables, phytoplankton, zooplankton, etc., are qualitatively similar across cases. An improved understanding of the role of stochastic dynamics will be of significant value in the interpretation of field data, relative impact of intrinsic versus extrinsic variability, and development of better numerical simulations.

The degree to which real ecosystems exhibit the oscillatory and other classical nonlinear behavior found in simple ecosystem models is not fully resolved, though specific examples from natural time series can certainly be found (Turchin and Taylor, 1992; Grenfell *et al.*, 2001). At present, there is only a limited amount of oceanic field data to compare with the results of this study, but the methodology presented in this article can be used to compare the predictability of generated noisy systems with data as it becomes available in the future. The nonlinear time series models and estimation techniques described in this article can be extended to data in which not all of the state variables in the system are measured. For example, if only zooplankton data is available, then it is possible to assume that the data are a univariate time series  $\{x_t\}$ . The difference is that  $X_t$  in (8) will be a vector of time-lags of the variable  $x_t$ ,  $X_t = (x_t, x_{t-1}, \dots, x_{t-d+1})$ .

Our finding that local predictability varies in a well defined manner with the oscillatory cycle has implications for parameter optimization studies (Fasham, 1995) and forecasting of, for example, harmful algal blooms (Truscott and Brindley, 1994). To date, most attempts to fit ecosystem models to field data have emphasized observational error (Carpenter *et al.*, 1994), not model or dynamical error. Characterizing the effect of noise on simple ecosystem models is also a first step toward stochastic marine ecosystem models similar to those under development for terrestrial systems (Moorcroft *et al.*, 2001). In particular, validated high resolution sub-gridscale models are need to parameterize unresolved variability for regional and global simulations (Doney, 1999).

## REFERENCES

- Ascioti FA, Beltrami E, Carroll TO, Wirick C. 1993. Is there chaos in plankton dynamics? *Journal of Plankton Research* **15**(6): 603–717.
- Bailey BA. 1996. *Asymptotics and applications of local Lyapunov exponents*. Ph.D. Thesis, Statistics Department, North Carolina State University, Raleigh, NC 27695–8203.
- Bailey BA, Ellner S, Nychka DW. 1997. Chaos with confidence: asymptotics and applications of local Lyapunov exponents. *Proceedings of the Fields/CRM Workshop on Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences* **11**: 115–133.
- Carpenter S, Cottingham K, Stow C. 1994. Fitting predator-prey models to time series with observation errors. *Ecology* **75**: 1254–1264.
- Case T. 2000. *An Illustrated Guide to Theoretical Ecology*. Oxford University Press: UK.
- Cullen JJ. 1990. On models of growth and photosynthesis in phytoplankton. *Deep Sea Research* **37**: 667–683.
- Doney SC. 1999. Major challenges confronting marine biogeochemical modeling. *Global Biogeochemical Cycles* **13**(3): 705–714.
- Doney S, Glover D, Najjar R. 1996. A new coupled, one-dimensional biological-physical model for the upper ocean: applications to the JGOFS Bermuda Atlantic time series (BATS) site. *Deep-Sea Research II* **43**: 591–624.

- Eckmann J-P, Ruelle D. 1985. Ergodic theory of chaos and strange attractors. *Review of Modern Physics* **57**: 617–656.
- Ellner S, Nychka D, Gallant A. 1992. *LENNS, a Program to Estimate the Dominant Lyapunov Exponent of Noisy Nonlinear Systems from Time Series Data*. Institute of Statistics Mimeo Series 2235, North Carolina State University, Statistics Department, North Carolina State University, Raleigh, NC 27695–8203.
- Eppley RW. 1980. Estimating phytoplankton growth rates in the central oligotrophic oceans. In *Primary Productivity in the Sea*, Falcowski PG (ed.). Plenum Press: New York; 231–240.
- Fasham MJR. 1995. Variations in the seasonal cycle of biological production in subarctic oceans: a model sensitivity analysis. *Deep-Sea Research* **42**(7): 1111–1149.
- Fasham MJR, Ducklow HW, McKelvie SM. 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research* **48**(3): 591–639.
- Geider RJ, MacIntyre HL, Kana TM. 1997. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll *a*:carbon ratio to light, nutrient limitation and temperature. *Marine Ecology Progress Series* **148**: 187–200.
- Grenfell B, Wilson K, Finkenstadt B, Coulson T, Murray S, Albon S, Pemberton J, Clutton-Brock TH, Crawley M. 2001. Noise and determinism in synchronized sheep dynamics. *Nature* **394**: 674–677.
- Hansen PJ, Bjornsen PK, Hansen BW. 1997. Zooplankton grazing and growth: scaling within the 2–2000  $\mu\text{m}$  body size range. *Limnology and Oceanography* **42**(4): 687–704.
- Harrison WG. 1992. Regeneration of nutrients. In *Primary Productivity and Biogeochemical Cycles in the Sea, Vol. 43 of Environmental Science Research*, Falcowski PG, Woodhead AD (eds). Plenum Press: New York; 385–407.
- Hastings A, Powell T. 1991. Chaos in a three-species food chain. *Ecology* **72**: 896–903.
- Lima ID, Olson DB, Doney SC. 2002. Intrinsic dynamics and stability properties of size-structured pelagic ecosystem models. *Journal of Plankton Research*. in press.
- May R. 1973. *Stability and Complexity in Model Ecosystems. Monographs in Population Biology*. Princeton University Press.
- Moorcroft P, Hurtt G, Pacala S. 2001. A method for scaling vegetation dynamics: the ecosystem demography model (ed). *Ecological Monographs*: Princeton **71**: 557–586.
- Nychka D, Ellner S, Gallant AR, McCaffery D. 1992. Finding chaos in noisy systems. *Journal of the Royal Statistical Society, Series B, Methodological* **54**: 399–426.
- Platt T, Sathyendranath S, Ulloa O, Harrison WG, Hoepffner N, Goes J. 1992. Nutrient control of phytoplankton photosynthesis in the Western North Atlantic. *Nature* **352**: 220–231.
- Popova E, Fasham M, Osipov A, Ryabchenko V. 1997. Chaotic behaviour of an ocean ecosystem model under seasonal external forcing. *Journal of Plankton Research* **19**: 1495–1515.
- Press W, Flannery B, Teukolsky SA, Vetterling W. 1992. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press: New York.
- Ryabchenko VA, Fasham MJR, Kagan BA, Popova EE. 1997. What causes short-term oscillations in ecosystem models of the ocean mixed layer? *Journal of Marine Systems* **13**: 33–50.
- Steele J. 1974. *The Structure of Marine Ecosystems*. Harvard University Press: Cambridge, MA.
- Truscott J, Brindley J. 1994. Ocean plankton populations as excitable media. *Bulletin Mathematical Biology* **56**: 981–998.
- Turchin P, Taylor A. 1992. Complex dynamics in ecological time-series. *Ecology* **73**: 289–305.