

LINKING ECOLOGICAL PATTERNS TO ENVIRONMENTAL FORCING VIA NONLINEAR TIME SERIES MODELS

MERCEDES PASCUAL^{1,3} AND STEPHEN P. ELLNER²

¹*Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 East Pratt Street,
Baltimore, Maryland 21202 USA*

²*Biomathematics Graduate Program, Department of Statistics, North Carolina State University,
Raleigh, North Carolina 27695-8203 USA*

Abstract. The identification of key environmental forcings responsible for population patterns is a pervasive ecological problem and an important application of time series analysis. A common approach, implemented with methods such as cross-correlation and cross-spectral analysis, relies on matching scales of variability. This approach concludes that a population pattern is caused by a physical factor if their variances share a dominant period. In a nonlinear system, however, forcing at one temporal period can produce a response with variability at one or more different periods. Thus, scale-matching methods will be most successful at establishing cause–effect relationships in linear systems, or close to equilibria, where nonlinear systems are well approximated by linear ones. Here, we propose an alternative approach that does not assume linearity and relies on time series models that are both nonlinear and nonparametric. We specifically apply these models to determine the correct but unknown frequency of a periodic forcing. The time series are generated by simulation of a predator–prey model. Under periodic forcing, this type of model is known to be capable of different dynamic regimes, including chaos and quasiperiodicity, in which the power spectra of population numbers exhibit variance at frequencies other than that of the forcing. We show that nonlinear time series models, built with feedforward neural networks, are able to distinguish the correct forcing period in the predator–prey simulations. These results hold under two common limitations of ecological data: the presence of dynamical and measurement noise, and the availability of time series data for only one variable. We discuss future applications of the approach to more general environmental forcings, other than periodic.

Key words: *chaotic dynamics; dynamical noise; environmental forcing; feedforward neural networks; measurement noise; nonlinear ecological response; nonlinear time series analysis; nonparametric time series models; periodic forcing; predator–prey model; quasiperiodic dynamics; temporal scales.*

INTRODUCTION

The problem of linking biological patterns to underlying environmental forcings pervades the study of ecological systems. It is present, for example, in the early dispute between explanations of population patterns based on intrinsic biological processes (Nicholson 1958, Smith 1961) and those based on extrinsic environmental factors (Andrewartha and Birch 1954). Beyond these two alternatives, studies today encompass the interplay of environmental fluctuations *and* biological processes (e.g., Higgins et al. 1997, Ellner et al. 1998, Hofman and Powell 1998), with many pressing questions arising on the effects of environmental variability, particularly in climate.

Key environmental forcings are often identified from time series patterns by matching scales of variability. This approach concludes that an ecological pattern is caused by a physical factor if their patterns of variation

share a dominant temporal period or spatial wave-number (Denman and Powell 1984). Well-known methods that implement this approach are cross-correlation and cross-spectral analysis. When the association between ecological and physical variables obtained with such methods is low, ecologists often invoke the high-dimensional nature of ecological systems—the idea that a large number of factors must influence the dynamics of any ecological variable. An alternative but seldom considered explanation for low associations is nonlinearity. In a nonlinear system, forcing at one temporal period can produce a response with variability at one or more different periods. Thus, scale-matching methods will be most successful at uncovering cause–effect relationships and at forecasting effects in linear systems, or close to equilibria, where nonlinear systems are well approximated by linear ones.

The analysis of a variety of ecological time series has provided evidence for nonlinearity in population growth and ecological interactions (recent examples include Royama 1992, Turchin and Taylor 1992, Ellner and Turchin 1995, Higgins et al. 1997, Leirs et al. 1997,

Manuscript received 3 December 1998; revised 30 August 1999; accepted 6 October 1999.

³ E-mail: mercedes@pampero.umbi.umd.edu

Saitoh et al. 1997, Björnstad et al. 1998, Grenfell et al. 1998, Stenseth et al. 1998). At the ecosystem level, Dwyer et al. (1978) provided compelling evidence for the transfer of variability across scales. A 15-year time series of phytoplankton abundance displayed variability at even-numbered harmonics of the seasonal forcing frequency (Dwyer et al. 1978). A series of microcosm experiments, designed to further examine this response, exhibited significant plankton variability at a number of frequencies, none of which coincided with that of the sinusoidal forcing, including a subharmonic (Dwyer and Perez 1983). Denman and Powell (1984) in a review of plankton patterns and physical processes, pointed out that as often as not, ecological responses could not be linked to a particular physical scale; they invoked nonlinearity as one possible explanation. These marine examples are significant because the physical environment is considered a main determinant of plankton patterns in the ocean (e.g., Mackas et al. 1985, Steele and Henderson 1994, Denman and Gargett 1995).

A transfer of variability across scales is found in a variety of ecological models, particularly those for consumer-resource and host-pathogen interactions (e.g., Schwartz and Smith 1983, Schaffer et al. 1990, Kot et al. 1992, Schwartz 1992, Pascual and Caswell 1997*a, b*). These interactions are well known for their unstable nature leading to oscillatory behavior in the form of persistent cycles or transient fluctuations with slow damping. Such cycles introduce an intrinsic temporal frequency capable of interacting with environmental fluctuations. One well-known example is given by predator-prey models under periodic forcing, where one parameter varies seasonally. These models display a rich array of possible dynamics, including frequency-locking, quasiperiodicity, and chaos (Inoue and Kamifukumoto 1984, Schaffer 1988, Kot et al. 1992, Rinaldi et al. 1993). In these dynamic regimes, predator and prey can display variability at frequencies other than that of the periodic forcing.

These models raise an important empirical question on the analysis of ecological time series: how can we identify environmental forcings related to specific ecological patterns without assuming a priori that systems are linear? When variability transfers across scales, conventional methods might fail to detect relationships between ecological and environmental variables simply because their underlying assumption of linearity is violated.

In this paper, we propose an approach that does not assume linearity and relies on time series models that are both nonlinear and nonparametric. This type of model has been previously applied in ecology to questions on the qualitative dynamics of a system, on sensitivity to initial conditions, and on the degree of predictability vs. noise (e.g., Ellner and Turchin 1995, Ellner et al. 1998, Pascual and Levin 1999). We propose that the same general type of model can be used to

identify a physical variable driving the dynamics of an ecological system, by selecting this variable from among a suite of candidate forcings.

To test and illustrate the proposed approach, we consider the particular application of determining the correct but unknown frequency of a periodic forcing. We generate data for the analysis by simulation of a predator-prey model in which the growth rate of the prey is periodic. This model provides a challenging test: it is known to be capable of different dynamic regimes, including chaos and quasiperiodicity, in which the power spectra of population numbers exhibit variance at frequencies other than that of the periodic forcing. Results show that nonlinear time series models built with feedforward neural networks are able to distinguish the correct forcing period from the predator and prey data in these aperiodic regimes. The approach is robust to two common limitations of ecological data: the presence of dynamic and measurement noise, and the availability of a single time series for only one variable. We end with a discussion of future applications to more general environmental forcings. In actual applications, the forcing variables will not be periodic. Our test cases mimic the situation where there are several candidate forcing variables, and the goal of the data analysis is to determine which of these is best able to account for the observed dynamics of the system. It is not possible, at least by our methods, to reconstruct an unmeasured forcing variable from the observed system dynamics.

THE PROBLEM

Consider an ecological system of interest whose variables $N_i(t)$ ($i = 1, \dots, n$), such as the density of interacting species or the biomass of interacting trophic levels, are influenced by an environmental variable $E(t)$ which fluctuates in time. Further consider that the dynamics of the system are given by

$$\begin{aligned} \frac{dN_1}{dt} &= f_1(N_1, N_2, \dots, N_n, E(t)) \\ \frac{dN_2}{dt} &= f_2(N_1, N_2, \dots, N_n, E(t)) \\ &\vdots \\ \frac{dN_n}{dt} &= f_n(N_1, N_2, \dots, N_n, E(t)) \end{aligned} \quad (1)$$

where the functions f_i are nonlinear and specify the respective rates of change of the variables N_i . The variable $E(t)$, which depends explicitly on time and influences the dynamics without being altered itself by the state of the system, is known as a "forcing," a "driver," or an "exogenous variable." Examples include the seasonal changes of temperature and light, the interannual fluctuations of climate parameters, such as those related to the El Niño Southern Oscillation,

the variation of mixed layer depth in the ocean, and the episodic upwelling events in coastal environments.

Ecologists are often interested in determining the nature of the physical forcing(s) driving the dynamics of a particular variable, say population numbers $N_1(t)$. Of particular interest is the value of the dominant frequency of $E(t)$. Often, however, information about the rest of the system is incomplete: knowledge is lacking on the functional form of the functions f_i and on the other variables of the system N_2, \dots, N_n . We address here whether we can identify the correct but unknown frequency of a periodic forcing from data on a single variable. Existing approaches to this problem assume that the underlying functions f_i are linear. We specifically allow for nonlinearity. The problem of $E(t)$ periodic provides a benchmark for the more general problem of selecting among candidate physical variables, not necessarily periodic, for which time series have been measured.

Although we have sketched the problem in continuous time (system 1), we could as well have used a discrete dynamic system. The method proposed below applies to a time series of observed values, with observations made at discrete intervals of time and denoted by N_t (where the subscript specifying the population has been dropped). A time series of the forcing at the same time intervals is denoted by E_t .

THE GENERAL APPROACH

Nonlinear time series models

The general approach relies on nonlinear time series models and on concepts from dynamical systems familiar to ecologists from the literature on chaos and prediction (e.g., Kot et al. 1988, Sugihara and May 1990, Tong 1990, Ellner and Turchin 1995). It consists of modeling the dynamics of a variable of interest, N_t , with a nonlinear time series model of the form

$$\frac{dN}{dt} = f(N_t, N_{t-\tau}, N_{t-2\tau}, \dots, N_{t-(d-1)\tau}, E_t) \quad (2)$$

where f is a nonlinear function, τ is a chosen lag, d is the number of time-delay coordinates, and E_t represents an environmental forcing affecting the dynamics of N_t . Or in discrete time,

$$N_{t+T_p} = f(N_t, N_{t-\tau}, N_{t-2\tau}, \dots, N_{t-(d-1)\tau}, E_t) \quad (3)$$

where T_p is a prediction time. The basic model in these equations has three key features. First, the function f is not specified in a rigid form. Instead, the functional form of the model is determined from the data (technically, the model is nonparametric) (Häerdle 1990, Wahba 1990, Green and Silverman 1994, Ellner and Turchin 1995). This is appealing because we often lack the information to specify exact functional forms.

Second, the model uses time-delay coordinates. This is rooted in a fundamental result from dynamical systems theory, known as “attractor reconstruction” (Tak-

ens 1981, Sauer et al. 1991), which tackles the problem of not knowing (and therefore, not having measured) all the interacting variables of a system. Takens’s Theorem essentially tells us that we can use lagged values of a single variable as surrogates for the unobserved variables of a system. Specifically, if the attractor of the system lies in an n -dimensional space, but one only samples the dynamics of a single variable $x(t)$; then, for almost every time lag τ and for large enough d , the attractor of the d -dimensional time series

$$X_t = [x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(d-1)\tau}] \quad (4)$$

is qualitatively similar to the unknown attractor of the n -dimensional system (Takens 1981; for ecological discussion see Kot et al. 1988). The “embedding dimension” d needs to be sufficiently high but not larger than $2n + 1$. Because the original dimension n is generally unknown, the choice of d becomes a problem of model selection.

Third, the model incorporates the effect of an exogenous variable E_t , which influences the dynamics without being altered itself by the state of the system. The inclusion of an exogenous variable E_t in Eqs. 2 or 3 is based on an extension by Casdagli (1992) of Takens’s Theorem to input–output systems. In the model, the effect of any measured environmental covariates, or of periodic “clock” variables that represent seasonality, is allowed to be nonlinear. To determine the unknown frequency of a periodic forcing, E_t takes the form of a periodic clock consisting of a sine and cosine, but not necessarily seasonal. Eq. 2 becomes

$$\frac{dN}{dt} = f\left(N_t, N_{t-\tau}, N_{t-2\tau}, \dots, N_{t-(d-1)\tau}, \sin\frac{2\pi}{T}t, \cos\frac{2\pi}{T}t\right) \quad (5)$$

where T denotes the unknown forcing period. The problem becomes one of selecting among different models with different values of T .

Model selection and computational methods

Analysis of the data can be based either on Eq. 2 or on Eq. 3, and one or the other may be advantageous depending on the application. In order to use Eq. 2, there are two requirements: the system must be such that a continuous-time model is appropriate (e.g., overlapping generations with continuous breeding), and sampling must be frequent enough that the data can be used to estimate the instantaneous rates of change of the measured state variables. In such cases Eq. 2 would be the natural choice because it is closer to the biological reality. If the data are too coarsely sampled in time, or if there are discrete nonoverlapping generations, then a model of the form of Eq. 3 would be the better option. However the use of Eq. 3 is dependent on the tacit assumption that the dynamics between times t and $t + T_p$ can be predicted from the value of the forcing variable E at the start of the time interval.

In this paper we report results for models of the form of Eq. 2, but we have partially repeated the analyses using Eq. 3 and results are essentially the same (for data on a single species free of noise and for data on a single species with both measurement and dynamical noise).

The dependent variable dN/dt in Eq. 2 is estimated from the time series of population numbers by local polynomial regression of $N(t_i)$ on t_i (Fan and Gijbels 1996). In local polynomial regression, the fitted curve is constructed point by point. The fitted value at a point t_i is obtained by fitting a low-order polynomial by weighted least squares, giving greatest weight to values of the time series with t_j near t_i and omitting values with $|t_i - t_j|$ above some cutoff. The derivative of the fitted curve at t_i was used as the estimate of $dN/dt(t_i)$. We used cubic polynomials, with cutoff equal to three sampling times.

To implement Eqs. 2 or 3, we use a class of time series models known as feedforward neural networks (FNN). FNN models were originally inspired by the neural architecture of the brain, but they are used here strictly as a statistical model. Previous studies of applications to nonlinear dynamics make FNN one of the methods of choice for time series in ecology whose typical length rarely exceeds 500 data points (McCaffrey et al. 1992, Ellner and Turchin 1995). In FNN, f is given by

$$f(x_1, x_2, \dots, x_m) = \beta_0 + \sum_{i=1}^k \beta_i G\left(\sum_{j=1}^m \gamma_{ij} x_j + \mu_i\right) \quad (6)$$

where G is a sigmoid function $G(y) = e^y/(1 + e^y)$, β_i , γ_{ij} , and μ_i are model parameters, m is the number of independent variables, and k is known as the “number of neurons.” As in previous studies, the value of k is limited a priori because FNN can otherwise approximate with arbitrary accuracy any smooth function on a bounded region in m -dimensional space (Barron 1991a, b). Given k and the set of independent variables (x_1, x_2, \dots, x_m) , the model parameters $(\beta_i, \gamma_{ij}, \mu_i)$ are estimated by ordinary least squares. (The FNN models are fitted using FUNFITS, a suite of S/Fortran functions that run in S-Plus [Nychka et al. 1998]).⁴

The independent variables (i.e., the arguments of the function f) are in our models, the lagged values of N_t and the exogenous variable or periodic clock, as well as other species numbers when observed. Comparison of models with different sets of x 's (including different T 's) and different values of k is done using a Generalized Cross Validation (GCV) model-selection criterion. GCV is one of several similar criteria in which a fitted model's mean-square residual error is “penalized” by a quantity that increases with the number of fitted parameters. The penalty term compensates for the fact that residuals from a fitted model underestimate

the true prediction error on data that were not used to fit the model (Linhart and Zucchini 1986, Efron and Tibshirani 1993). The GCV criterion function is

$$V_c = \left(\frac{\text{RMS}}{1 - p\frac{c}{n}}\right)^2 \quad (7)$$

where RMS is the root-mean-square residual error of the fitted model, p is the number of fitted parameters, and n is the sample size. T is not counted as a fitted parameter because we are comparing among a few candidate values rather than allowing T to vary arbitrarily. V_1 is the standard GCV criterion (see Wahba 1990). We used $c = 2$ based on Nychka et al. (1992). This slight over-penalization of model complexity creates a small bias towards simpler models, but greatly reduces the chances of spuriously selecting an overly complex model. In our analyses the value of c is often irrelevant, as we are mostly comparing models that differ only in the choice of exogenous forcing function and therefore have the same values of n and p .

Finally, to guarantee that N_t and $N_{t-\tau}$ are sufficiently uncorrelated, the value of τ is chosen as the time lag for which the autocorrelation function first crosses 0.5 (for details see Ellner and Turchin 1995: Appendix 2).

THE SIMULATED PREDATOR-PREY DATA

Predator-prey models provide ideal candidates to generate the data to evaluate the proposed approach. Under periodic forcing, they are known to be capable of different dynamic regimes, including chaos and quasi-periodicity, in which variability occurs at periods other than that of the forcing (e.g., Schaffer 1988, Kot et al. 1992). We use here a standard predator-prey system with logistic growth of the prey and a saturating functional response of the predator, in which the intrinsic growth rate of the prey is periodic in time.

Let $P(t)$ and $H(t)$, respectively, denote the prey and predator numbers at time t . The dynamics of the predator-prey system are given by

$$\begin{aligned} \frac{dP}{dt} &= R \left[1 + \varepsilon \sin \frac{2\pi}{T_f} t \right] P \left(1 - \frac{P}{K} \right) - \frac{AC_1 P}{C_2 + P} H \\ \frac{dH}{dt} &= \frac{C_1 P}{C_2 + P} H - MH. \end{aligned} \quad (8)$$

The parameters R , ε , and T_f denote the mean, amplitude, and period of the intrinsic growth rate of the prey. K , M , and $1/A$ denote the carrying capacity of the prey, the death rate of the predator, and the yield coefficient of prey to predator, respectively. The constants C_1 and C_2 parameterize the saturating Type II functional response.

The model is simplified by introducing the dimensionless variables $p = P/K$ and $h = AH/K$. Time and the forcing period are both multiplied by the mean

⁴ A Unix version of FUNFITS, as well as a user manual, can be found at (<http://goldhill.cgd.ucar.edu/stats/Funfits>)

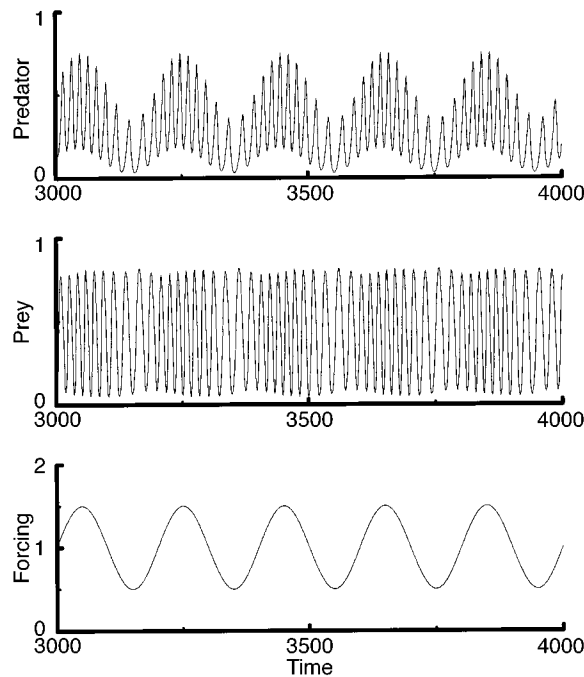


FIG. 1. Quasiperiodic dynamics of predator and prey numbers.

growth rate. Thus, the equations for the predator-prey dynamics become

$$\begin{aligned} \frac{dp}{dt} &= \left[1 + \varepsilon \sin \frac{2\pi}{T_f} t \right] p(1 - p) - \frac{ap}{1 + bp} h \\ \frac{dh}{dt} &= \frac{ap}{1 + bp} h - mh \end{aligned} \quad (9)$$

with the new parameters

$$a = \frac{C_1 K}{C_2 R} \quad b = \frac{K}{C_2} \quad m = \frac{M}{R}. \quad (10)$$

The model is integrated numerically with a fourth-order Runge-Kutta scheme with a fixed time step. The parameters in all simulations are $\varepsilon = 0.5$, $a = 5$, $b = 5$, and $m = 0.6$. With these parameters and no forcing, the predator-prey system exhibits a limit cycle. With periodic forcing, complex dynamics become possible. We choose two different values for the forcing period T_f , which generate respectively two different types of dynamic regimes and therefore, power spectra. For a forcing period $T_f = 200$, much larger than the natural period of oscillation, the dynamics are quasiperiodic (Figs. 1 and 2). The power spectrum displays multiple peaks that differ from the forcing frequency, but also a peak at the forcing frequency $f = 1/200$ (Fig. 2). For a forcing period $T_f = 10$ closer to the natural period of oscillation, the dynamics are chaotic (Figs. 3 and 4). The chaotic nature of the dynamics is demonstrated by a positive dominant Lyapunov exponent ($\lambda = 0.036$ bits per unit time). The power spectrum for chaotic

dynamics shows variability at all frequencies, but displays here dominant peaks at the forcing frequency $f = 1/10$ and at harmonics and subharmonics of this frequency (Fig. 4). Our goal is to identify the value of the forcing period T_f from time series of the predator and/or prey, with and without noise present. We consider both errors in the measurements and errors that affect the dynamics directly.

To simulate the effect of measurement errors, we first simulated the models without noise, and then added normally distributed errors to the log-transformed population densities for one of the species. This produces measurement errors whose coefficient of variation is constant over the full range of population densities. The noise level is indicated by the parameter CV, which is the standard deviation of the normally distributed errors and is approximately the coefficient of variation

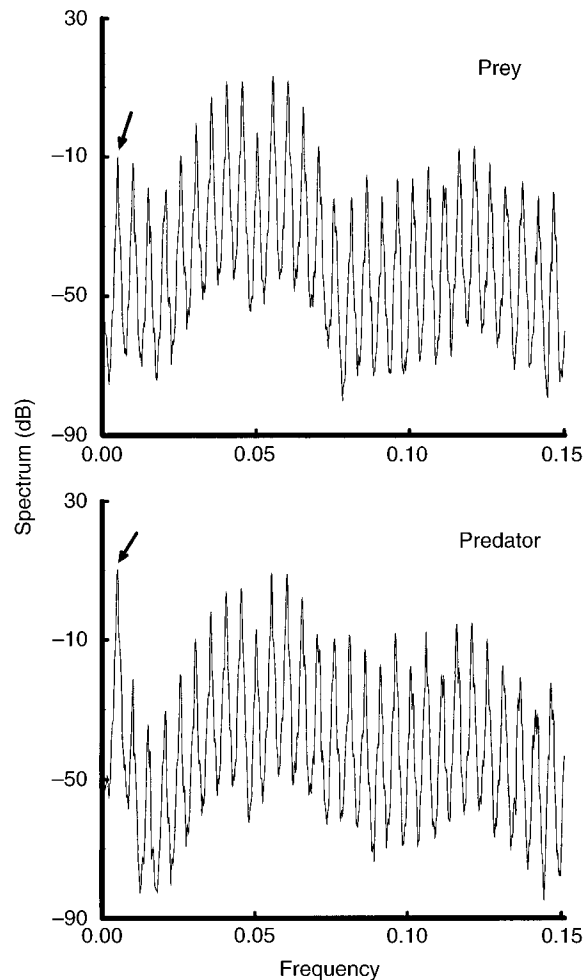


FIG. 2. The power spectrum of predator and prey numbers in the quasiperiodic regime. The arrow indicates the peak in the spectrum corresponding to the forcing frequency. (The power spectrum in this and other figures is estimated as a smoothed periodogram and expressed in decibels. In all the figures, we plot the part of the power spectrum containing the dominant peaks.)

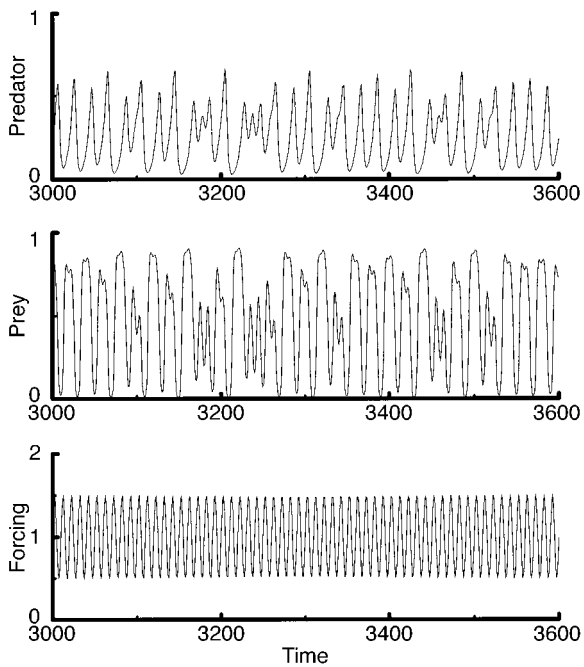


FIG. 3. Chaotic dynamics of predator and prey numbers.

of the errors on the untransformed scale (see *Appendix: Measurement noise* for further details). To implement dynamical noise, the errors must enter in the simulation itself. We assume here that the effect of any unmeasured variables affecting the dynamics can be modeled as random multiplicative perturbations to the finite rate of increase between times t and $t + 1$ (see *Appendix: Dynamical noise* for details on the simulations).

Several cases are considered in the analysis below. We start with the ideal case in which data for both species are available, the dimensionality of the system is known, and there is no error in the measurement of population numbers. We quickly relax this unlikely scenario: all following cases consider that a single time series is available, first free of noise and then with the different types of errors. Measurement and dynamical noise are considered first separately and then jointly; in all cases with errors, replicates are considered.

RESULTS

The ideal case: data on both species

When data for both species are available, the time delay coordinates in Eq. 5 are not needed and the time series model can be written as

$$\frac{dP}{dt} = f\left(P, H, \sin\frac{2\pi}{T}t, \cos\frac{2\pi}{T}t\right) \quad (11)$$

where T denotes the unknown period of the clock. To infer the value of the original forcing period, models with different T values are compared. Candidate values for T can be selected based on the power spectrum and/

or the population time series. The respective models are then fitted to the data and compared via the cross-validation criterion V_c (Eq. 7). All results are initially obtained for time series that are 250 data points long (sampled at $\Delta t = 1$). Shorter time series are then considered.

Quasiperiodic dynamics ($T_f = 200$).—One candidate for the forcing period is $T = 200$. This value is apparent in the long period modulation of the time series (Fig. 1), as well as in the power spectrum (Fig. 2), provided the time series is sufficiently long. A shorter oscillation with $T \approx 20$ is also apparent in both the predator and prey time series.

When the model for $T = 200$ is compared to that for $T = 20$, the cross-validation criterion selects the correct model. Models with the same number of neurons k are compared and the model with smaller V_c is selected. For all k values within the range considered (1 to 5), the model with $T = 200$ is selected (e.g., $V_c = 3.2 \times 10^{-5}$ vs. $V_c = 1.5 \times 10^{-3}$ for $T = 20$, for $k = 3$). Fig. 5 shows that the correct model fits the data extremely

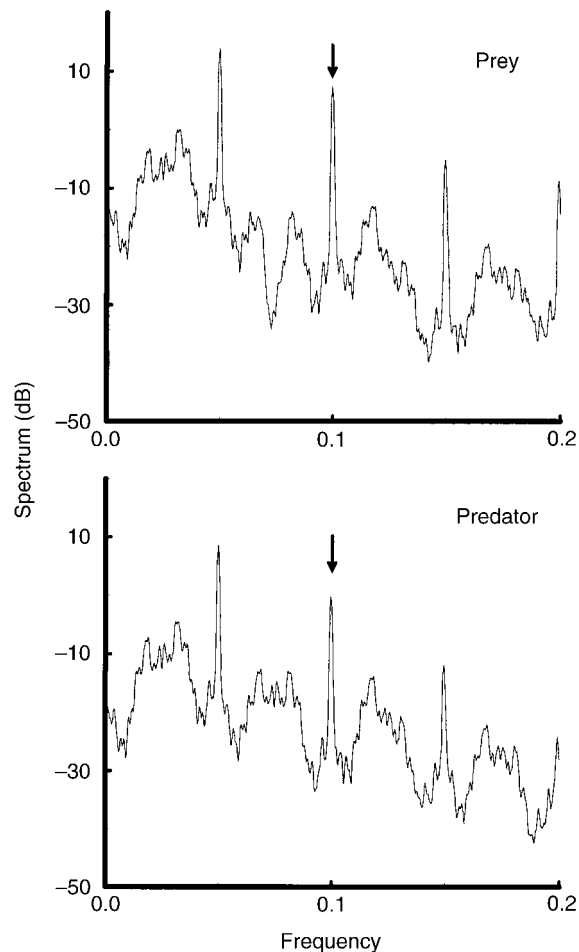


FIG. 4. The power spectrum of predator and prey numbers in the chaotic regime. The arrow indicates the peak in the spectrum corresponding to the forcing frequency.

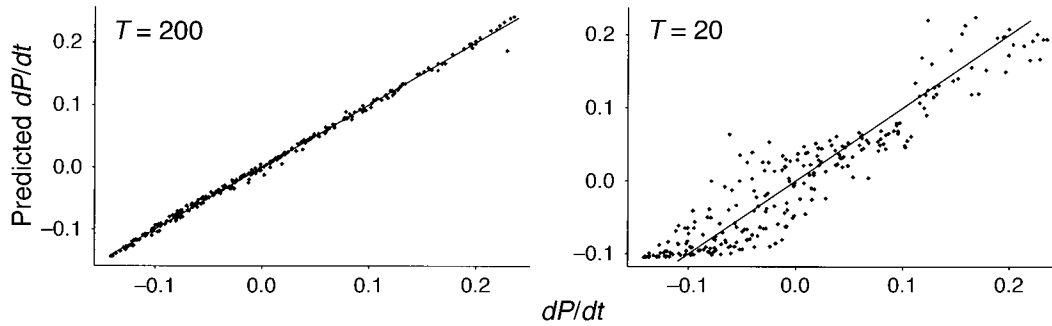


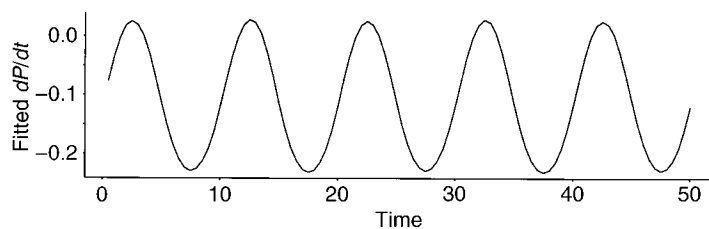
FIG. 5. Predicted vs. observed values for the two time series models with forcing periods 200 and 20, respectively ($k = 3$).

well, accounting for a large fraction of the variance in the time series ($r^2 = 99.74\%$ vs. $r^2 = 88.11\%$ for $T = 20$). We also fitted models with T corresponding to other peaks in the power spectrum near $T = 20$, with very similar results. Thus $T = 200$ is clearly picked out as being more consistent with the data than $T \approx 20$.

Chaotic dynamics ($T_f = 10$).—From the power spectrum of either prey or predator numbers, the obvious candidates for the forcing period are the two main peaks at $f = 0.1$ or $f = 0.05$ (Fig. 4), corresponding to $T = 10$ or 20 . When time series models with these forcing periods are fitted, the model with $T = 10$ has the smaller V_c ($V_c = 4.6 \times 10^{-5}$ and 1.5×10^{-4} , respectively, at $k = 3$). However both models fit the derivatives obtained from the time series data extremely well ($r^2 = 99.74\%$ and 99.13% , respectively), so it would be hard to claim that one is clearly preferred over the other.

With hindsight, this result is not surprising. Provided with the subharmonic $T = 20$ as input, the neural network model is able to create an output at the base period $T = 10$, for example, by approximating the function $f(\sin(x), \cos(x)) = \sin(x)^2$. An additional step is therefore necessary, to determine whether the fitted model with input forcing period $2T$ is internally generating a forcing period of T . To do this, the fitted function f is plotted as a function of time for constant values of P and H . This plot allows us to “see” the forcing produced by FNN from the periodic clock with all other independent variables kept constant. Fig. 6 shows that the model for $2T = 20$ is indeed creating a forcing of period 10. Based on these results, we choose the model with the correct period $T = 10$.

FIG. 6. The fitted derivative of prey numbers as a function of time for constant values of prey and predator numbers ($P = H = 0.5$).



The common case: data on a single species

The above analyses rely on the availability of data for both the predator and prey. In most ecological applications, however, the dimensionality of the system (the number of relevant variables) is not known and data is available for only one species. In this case, lagged variables are needed. Eq. 5 rewritten for the prey data becomes

$$\frac{dP}{dt} = f\left(P_t, P_{t-\tau}, P_{t-2\tau}, \dots, P_{t-(d-1)\tau}, \sin\frac{2\pi}{T}t, \cos\frac{2\pi}{T}t\right). \tag{12}$$

If both dynamical noise and measurement errors are absent, then Takens’ Theorem guarantees a faithful embedding of the attractor with only a few lags, which is topologically and dynamically equivalent to the two-species dynamics. It is therefore no surprise that the correct forcing period can also be identified by fitting models to one of the time series using time-delay coordinates. Analyses exactly parallel to those in *The ideal case: data on both species* pick out the correct forcing in both the quasiperiodic and chaotic cases from the prey data alone, provided the number of lags d is larger or equal to 2. Indeed models with $d \geq 2$ are always selected by the cross-validation criterion (lower V_c values).

Table 1 shows that for quasiperiodicity, the correct period ($T = 200$) is selected by the cross-validation criterion (lower V_c values), and that there is a marked decrease in V_c from $d = 1$ to $d = 2$. For chaos, Table 2 shows that when a forcing period $T = 10$ is specifically compared to its subharmonic $2T$, results based

TABLE 1. Comparison of models with forcing periods 200 and 20, respectively, for quasiperiodic dynamics when only the prey time series is available for analysis.

Model parameters	$T = 200$		$T = 20$	
	V_c	r^2 (%)	V_c	r^2 (%)
$d = 1, k = 5$	6.05×10^{-3}	56.54	8.76×10^{-3}	37.13
$d = 2, k = 5$	3.06×10^{-6}	99.98	1.99×10^{-4}	98.72
$d = 3, k = 5$	2.26×10^{-6}	99.99	1.32×10^{-5}	99.92

Note: In this and other tables, results are shown for the number of neurons k corresponding to the best model for each embedding dimension d (lowest V_c).

uniquely on the cross-validation criterion are not clear-cut. The V_c values for the right period are only slightly smaller than those for the wrong period even for $d \geq 2$. As before, we take one more step and determine if the model with period $2T$ is fitting the data well by generating a forcing of period T . To examine whether the forcing generated from the clock with period 20 has a significant component at period 10, we fix all independent variables $P_t, P_{t-\tau}, \dots, P_{t-(d-1)}$ other than the clock to constant values. Although the shape of the fitted function f vs. time varies with the specific choice of these constants, it typically contains a significant contribution at period 10 (compare Fig. 7A, B). A systematic way to take this variation into account is to choose the constants from the observed values of $P_t, P_{t-\tau}, \dots, P_{t-(d-1)}$ and to loop over all such vectors of observed values. For each one, the corresponding function $f(t, T = 20)$ is computed, as well as its power spectrum. By averaging these spectra for the different frequencies, a mean power spectrum is computed. For the models with $T = 20$, the mean power spectrum shows a peak at period 10 comparable or larger than that at period 20. We conclude that the dynamics are consistent with forcing at period 10.

With lags and no noise, the models with the wrong period fit the data quite well (high r^2 's in Tables 1 and 2). This accurate fit is possible because the lagged variables, besides the wrong "clock," can be used by the fitted model to make up for the lack of forcing at the right period. As we will show later, the presence of noise interferes with this use of the lagged variables, as they become an imperfect "clock" facilitating the discrimination among models.

Measurement noise.—The time series analyzed so far lack one essential feature of ecological data, namely the presence of measurement errors. In the presence of measurement noise, a faithful reconstruction of the attractor is no longer possible. Fig. 8A illustrates that the noise quickly degrades the smoothness of the reconstructed trajectories in phase-space. Nevertheless, we find that the forcing period can be identified correctly at appreciable noise levels (CV = 0.2). Fig. 8B shows a typical time series for the chaotic regime and measurement noise with CV = 0.2. Comparisons of models (Eq. 12) with similar complexity (equal k and

TABLE 2. Comparison of models with forcing periods 10 and 20, respectively, for chaotic dynamics with only the prey time series available for analysis.

Model parameters	$T = 10$		$T = 20$	
	V_c	r^2 (%)	V_c	r^2 (%)
$d = 1, k = 5$	4.53×10^{-3}	78.31	9.62×10^{-4}	95.39
$d = 2, k = 5$	3.72×10^{-5}	99.84	5.55×10^{-5}	99.76
$d = 3, k = 5$	2.56×10^{-5}	99.90	3.77×10^{-5}	99.86

d) lead to the selection of the correct period (Table 3). The models with $T = 10$ have the smallest value of the cross-validation criterion provided the number of lags d exceeds one. The overall best model corresponds to $d = 3$ and $T = 10$. The models with $T = 10$ also explain a larger fraction of the variance, and the selection of the correct model can be based solely on the cross-

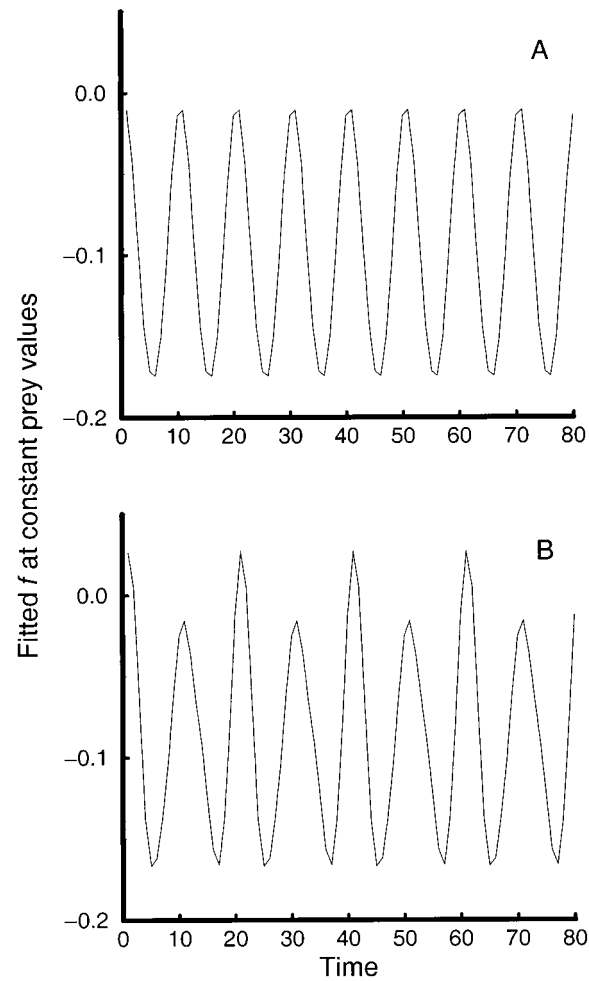


FIG. 7. The fitted function f as a function of time for constant values $P_t = 0.24$ and $P_{t-3} = 0.51$. In (A), the model with the correct period, $T = 10$; in (B), the model with the wrong period, $2T = 20$ ($d = 2, k = 5$). Notice that the latter uses period 20 to generate a forcing similar to the correct one, with variance at period 10.

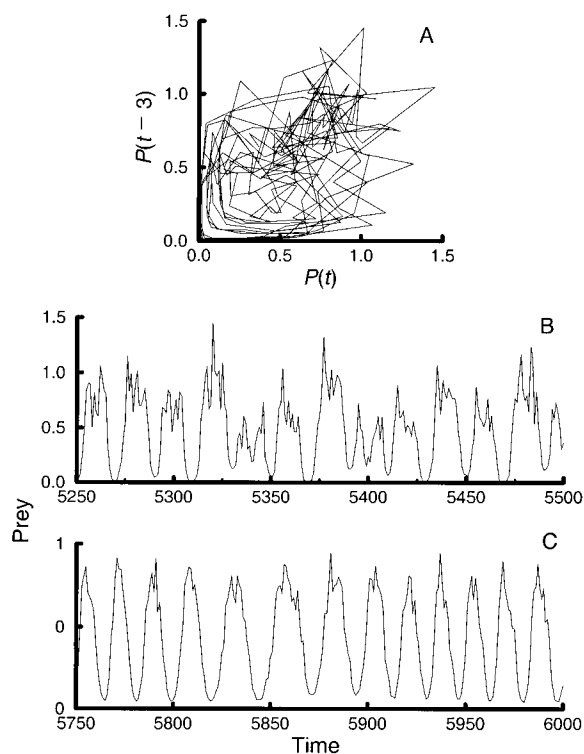


FIG. 8. Prey data with measurement errors. Panels (A) and (B) show the reconstructed trajectory in two dimensions and the prey time series, respectively, for the chaotic simulation ($CV = 0.2$). Panel (C) shows the time series for the quasiperiodic simulation ($CV = 0.1$).

validation criterion. It can be verified, however, that the best model for the wrong period $T = 20$ clearly generates period $T = 10$.

The analysis for the quasiperiodic regime appears somewhat more sensible to noise. For the prey time series, the correct forcing period can only be identified up to $CV = 0.1$ (Table 4). It must be noted, however, that the analysis of prey-only data for the quasiperiodic case is particularly challenging because the prey in the noise-free simulation displays an extremely weak modulation at period 200 (see Fig. 1). Fig. 8C shows that the noise effectively blurs this modulation from the time series for a $CV = 0.1$. This modulation is, on the other hand, more apparent in the predator dynamics, making the analysis of predator-only data more robust to noise. For the predator time series, the correct period is clearly identified up to $CV = 0.2$ ($V_c = 4.22 \times 10^{-3}$ and $r^2 = 73.89\%$ for the best model with $T = 200$ vs. $V_c = 5.12 \times 10^{-3}$ and $r^2 = 68.13\%$ for the wrong period $T = 20$, with $d = 3$ and $k = 5$).

The above results can be replicated with similar outcomes. For both chaos and quasiperiodicity, we repeated the analyses for four different sections of the time series and different sequences of random numbers. For all the replicates, the correct forcing period was selected.

TABLE 3. Comparison of models with forcing periods 10 and 20, respectively, for chaotic dynamics and measurement noise with only the prey time series available for analysis.

Model parameters	$T = 10$		$T = 20$	
	V_c	r^2 (%)	V_c	r^2 (%)
$d = 1, k = 2$	2.36×10^{-2}	30.44	1.94×10^{-2}	42.70
$d = 2, k = 4$	1.83×10^{-2}	59.14	1.98×10^{-2}	55.79
$d = 3, k = 5$	1.80×10^{-2}	68.90	2.25×10^{-2}	60.72
$d = 4, k = 3$	1.97×10^{-2}	56.07	2.06×10^{-2}	53.93

Note: $CV = 0.2$.

These results demonstrate that the method is able to distinguish the correct forcing period for nonnegligible levels of measurement noise. A different type of error is present, however, in ecological data as the result of noise that affects the dynamics directly.

Dynamical noise.—In the presence of dynamical noise, a faithful embedding is no longer guaranteed unless we can also (somehow) observe the noise (Casdagli 1992). Put another way: with dynamical noise present, having only one of the time series rather than both entails the inevitable loss of some information about the dynamics. Nonetheless, we find that as for measurement noise, the forcing period can be identified correctly up to $CV = 0.2$.

The resulting time series for the chaotic simulation is shown in Fig. 9B. With noise added to the simulation, the power spectrum of the time series has changed in an important way with the appearance of a dominant peak not present before at period $T = 30$ (Fig. 9A). An interesting question becomes whether we can now distinguish the correct period $T = 10$, still reflected in the spectrum, from this dominant period. We add this comparison to our analyses.

The results are summarized in Table 5. Once the embedding dimension $d \geq 2$, the models with the correct forcing period are selected when comparing models with similar complexity (equal k and d). This is also the case when comparisons are made across the best (lowest V_c) model for each T . Interestingly, the correct forcing period can be selected based solely on the cross-validation criterion (Table 5). Thus, some dynamical noise appears to facilitate the choice of the right model. One possible explanation for this phe-

TABLE 4. Comparison of models with forcing periods 200 and 20 respectively, for quasiperiodic dynamics and measurement noise when only the prey time series is available for analysis.

Model parameters	$T = 200$		$T = 20$	
	V_c	r^2 (%)	V_c	r^2 (%)
$d = 1, k = 4$	10.09×10^{-3}	39.79	13.46×10^{-3}	19.64
$d = 2, k = 5$	3.92×10^{-3}	81.11	4.36×10^{-3}	79.03
$d = 3, k = 3$	3.77×10^{-3}	77.95	4.42×10^{-3}	75.75
$d = 4, k = 3$	3.80×10^{-3}	79.13	4.25×10^{-3}	76.64

Note: $CV = 0.1$.

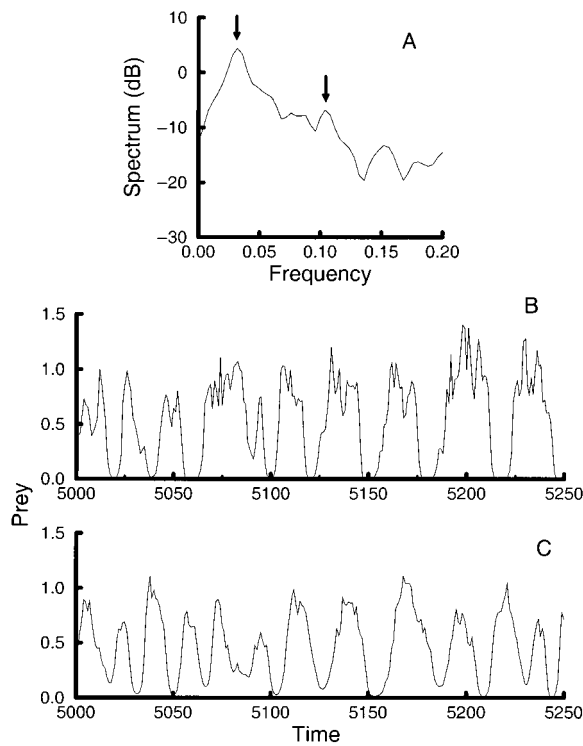


FIG. 9. Prey data with dynamical errors. (A) The power spectrum of the prey chaotic time series with arrows indicating the forcing frequency (1/10) and the dominant frequency (1/30). (B) The prey time series for the chaotic simulation (CV = 0.2). (C) The prey time series for the quasiperiodic simulation (CV = 0.1).

nomenon is that dynamical noise causes the system to visit a larger region of state space than it would if noise were absent. This “reveals” more of the underlying system map f , which often can more than compensate for the loss of information due to noise (Schaffer et al. 1986). The situation is analogous to that in linear regression of Y on X , where a higher variance in X produces smaller mean-square errors for estimates of the regression coefficients, all else being equal. It can also be verified that the models with the wrong period $2T = 20$ clearly generate a forcing of period 10, which further supports the correct choice of model.

In the presence of dynamical noise, another comparison becomes possible, that of autonomous to periodically forced models. This comparison allows us to

TABLE 6. Comparison of models with forcing periods 20 and 200 respectively, for the quasiperiodic simulation with dynamical noise when only the prey time series is available for analysis.

Model parameters	$T = 20$		$T = 200$	
	V_c	r^2 (%)	V_c	r^2 (%)
$d = 1, k = 2$	10.94×10^{-3}	25.01	12.24×10^{-3}	13.58
$d = 2, k = 4$	4.56×10^{-3}	75.3	4.42×10^{-3}	76.04
$d = 3, k = 5$	4.96×10^{-3}	79.04	4.43×10^{-3}	81.26

Note: CV = 0.1.

specifically address whether the dynamics could have been generated by the feedbacks between endogenous variables without any external forcing (Ellner and Turchin 1995). We have found that in the absence of dynamical noise, the method is not able to clearly distinguish between an autonomous FNN model of the form

$$\frac{dP}{dt} = f(P_t, P_{t-\tau}, P_{t-2\tau}, \dots, P_{t-(d-1)\tau}) \quad (13)$$

and its periodic counterpart in Eq. 12. The distinction is not clear because the autonomous model can fit the data well by using time delay coordinates in place of the forcing. Dynamical noise sabotages this replacement, by forcing the system away from a regular trajectory that can serve as a surrogate “clock.” It is then possible to identify that the dynamics are actually non-autonomous: models of similar complexity have lower V_c when periodically forced. For example, the best forced model with one lag ($d = 2, k = 4$) has a $V_c = 1.63 \times 10^{-2}$ and an $r^2 = 57.45\%$, while the corresponding autonomous model with three lags ($d = 4, k = 4$) has a $V_c = 1.86 \times 10^{-2}$ and an $r^2 = 51.29\%$.

For the quasiperiodic case, the analysis appears somewhat more sensitive to dynamical noise. For a CV = 0.1 the distinction of the correct model is barely possible: the cross-validation criterion V_c is only slightly smaller for the correct frequency (embedding dimensions $d \geq 2$, Table 6). As for measurement noise, however, dynamical noise with a CV = 0.1 has effectively blurred the modulation at period 200, making the analysis based on prey data more sensitive to noise than that based on predator data (Fig. 9C). For example, for a CV = 0.1, the analysis is able to clearly identify the correct forcing period from the predator time series ($V_c = 0.99 \times 10^{-3}$ vs. $V_c = 1.54 \times 10^{-3}$ for $T = 200$

TABLE 5. Comparison of models with forcing periods 10 and 20 respectively, for the chaotic simulation with dynamical noise when only the prey time series is available for analysis.

Model parameters	$T = 10$		$T = 20$		$T = 30$	
	V_c	r^2 (%)	V_c	r^2 (%)	V_c	r^2 (%)
$d = 1, k = 2$	2.53×10^{-2}	12.87	2.39×10^{-2}	17.56	2.22×10^{-2}	23.39
$d = 2, k = 4$	1.63×10^{-2}	57.45	1.72×10^{-2}	55.09	1.87×10^{-2}	51.14
$d = 3, k = 4$	1.65×10^{-2}	60.48	1.79×10^{-2}	57.09	1.94×10^{-2}	53.56

Note: Coefficient of variation CV = 0.2.

and $T = 20$ respectively, with $d = 2$ and $k = 5$). It is also able to select the periodic model ($T = 200$) over its autonomous counterpart. For a $CV = 0.2$, the selection of the correct forcing period is still possible ($V_c = 6.18 \times 10^{-3}$ vs. $V_c = 6.86 \times 10^{-3}$ for $T = 200$ and $T = 20$ respectively, with $d = 2$ and $k = 5$).

For the chaotic regime, we have replicated the comparison of periods $T = 10$ and $2T = 20$ by repeated simulations of the model with dynamical noise. For four such simulations and different sections of the time series, the correct period is always identified. We further replicated the results by considering shorter time series.

Time series length in dynamical systems is generally evaluated relative to the time taken by a trajectory to travel around the attractor. For our specific purpose, we must also consider time series length relative to the forcing period. For the quasiperiodic simulation, the length considered so far (250) is already close to the forcing period. For the chaotic simulation, it is possible to challenge the analysis further by limiting time series length. For 100 data points, we have found that the analysis correctly identifies the forcing period $T_f = 10$, when data is available for only one species and dynamical noise is present ($CV = 0.2$). It also correctly identifies that the system is nonautonomous. Selection of the correct periodic model is possible based solely on the cross-validation criterion ($V_c = 1.46 \times 10^{-2}$ and $r^2 = 65.88\%$ vs. $V_c = 1.74 \times 10^{-2}$ and $r^2 = 59.36\%$ for $T = 10$ and $T = 20$, respectively, with $d = 2$ and $k = 2$).

For a lower noise level ($CV = 0.1$), it is possible to push the limit of time series length even further to only 50 data points. Because such short time series contain only a few cycles, results can vary with the particular time segment chosen for the analysis. Thus, we have repeated the analysis for different time segments. In the analysis of 10 such segments, we find that for most of them (8/10) we can select the correct forcing period. In five of these, the correct period is selected solely from the cross-validation criterion, while in the remaining three, the actual period generated by the forcing must be examined. Only for 2 of the 10 segments, the method fails to select the correct forcing period. For higher noise levels ($CV = 0.2$), results become more dependent on the specific time segment analyzed. Only in 6 of the 10 segments, were we able to select the correct forcing period.

Dynamical and measurement noise.—The results are also robust to the presence of *both* types of errors. We illustrate this for a chaotic time series with $CV = 0.1$ for dynamical noise and $CV = 0.2$ for measurement noise. Fig. 10 (A, C) shows the resulting time series and its power spectrum. The presence of noise has altered the spectrum, which now exhibits variance not only at periods $T = 10$ and $2T = 20$ but also at a period ≈ 27 . The noise has also noticeably degraded the smoothness of the reconstructed trajectories (Fig. 10B).

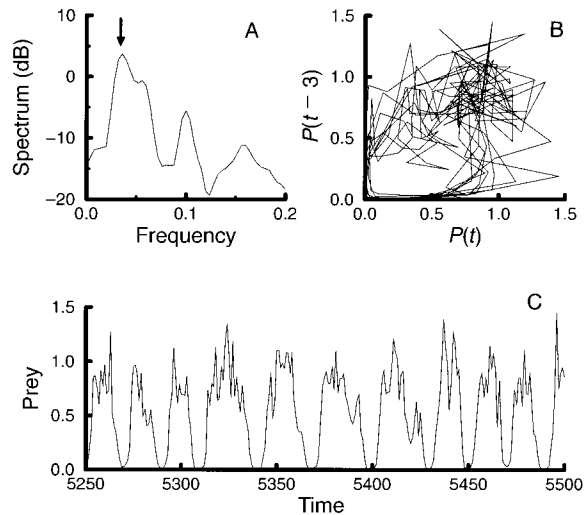


FIG. 10. Prey data in the chaotic regime with both measurement and dynamical errors ($CV = 0.2$ and $CV = 0.1$, respectively). Panel (A) shows the power spectrum of the time series with a dominant peak at frequency $1/27$. Panels (B) and (C) show the reconstructed trajectory in two dimensions and the prey time series, respectively.

The comparison of T and $2T$, however, correctly identifies the forcing period: the selected model for $T = 10$ has a smaller value of V_c ($V_c = 23.28 \times 10^{-3}$ and $r^2 = 59.09\%$ vs. $V_c = 26.98 \times 10^{-3}$ and $r^2 = 52.60\%$, for $d = 3$ and $k = 4$), and the model with the wrong period $2T = 20$ but similar complexity (equal k and d) clearly generates period 10. These results remain unchanged for four replicates whose time series differ in the dynamical and measurement errors. The analysis further selects period 10 over the wrong but dominant period 27 (with $V_c = 27.57 \times 10^{-3}$ and $r^2 = 51.55\%$, for $d = 3$ and $k = 4$).

DISCUSSION

We have proposed an approach to complement traditional statistical methods for relating ecological responses to underlying environmental forcings. In particular, the proposed method does not assume that the system responds linearly to the forcing, an assumption likely to be violated in many ecological systems. We have shown that nonlinear time series models correctly identify the unknown frequency of the periodic forcing, even in the presence of measurement and/or dynamical noise and with data on only one of the two interacting species. The presence of noise appears to facilitate the discrimination among models, up to the high noise levels for which the approach breaks down.

This work has considered only one among several possible applications of the proposed approach. The method should apply more broadly to identify environmental forcing(s) by distinguishing among candidate variables for which time series are available. We emphasize, however, that the test problems considered

here are especially hard ones, involving in one case a periodic forcing and its subharmonic, instead of aperiodic signals, and in the other, a forcing variable whose period is close to that of the dominant oscillations in the data.

The proposed approach should also prove useful in short-term prediction of ecological dynamics as a function of environmental variability. An example is found in the recent modeling of measles epidemics that incorporates forcing by a seasonal clock (Ellner et al. 1998). Future work should also consider forcing by specific environmental variables whose fluctuations are measured in time (e.g., Dixon et al. 1999).

Two areas of ecological research—fisheries and epidemiology—appear as potential candidates for the application of the proposed approach. Although seemingly unrelated, they share some important prerequisites: the availability of data, the presence of nonlinearity in standard dynamic models, and the current interest in environmental effects on their dynamics. The role of climatic and environmental variables has become an important current issue in the dynamics of many diseases (e.g., Colwell 1996, Patz et al. 1996, Linthicum et al. 1997). Better knowledge on the effects of climate variability has potential application to short-term prediction and can contribute to the implementation of timely preventive measures, particularly in regions of the globe where resources for prevention are either limited or not in place.

In fisheries, nonlinearity is present in standard dynamic models known as stock-recruitment maps (Cushing 1983). Prediction with these models has remained a difficult if not elusive goal (Rothschild 1986). The International Council for the Exploration of the Sea recently held a symposium on physical-biological interactions in the recruitment dynamics of exploited marine populations, in which a major section was devoted to climate variability and recruitment dynamics (Fogarty et al. 1997). The incorporation of environmental effects is recognized as an important challenge in the management of marine fisheries (Hofmann and Powell 1998). Nonparametric and nonlinear time series models built not only from data on stock and recruitment, but also from time-delay coordinates and environmental variables, could provide a useful framework for prediction.

Our choices with regard to methods—model family, model selection criterion, and so on—were based mainly on their success in related applications. We could not defend a claim that our choices are the best possible, and we therefore make no such claim. We have shown what can be done with one sensible set of choices, and if others can find improvements, so much the better for the general approach that we are proposing here.

As with linear methods, the nonlinear dynamic models described here do not necessarily imply the existence of direct causal connections among the variables.

However, fitting alternate dynamic models can be an effective way to evaluate and compare the degree of consistency between alternative mechanistic hypotheses and the available data (Hilborn and Mangel 1997, Kendall et al., *in press*). Those hypotheses flagged as most consistent with the data can then be the target for more mechanistic investigations or experimental study. For purposes of prediction, the models provide a framework for incorporating known ecological structure into data-driven, statistical forecasts. Here we have shown that incorporating data on specific environmental forcing variables is a potentially useful, yet largely unexplored avenue for unearthing causal relationships, which can then be exploited for making more accurate forecasts.

ACKNOWLEDGMENTS

M. Pascual is pleased to acknowledge the support of the James S. McDonnell Foundation through a Centennial Fellowship, the Knut and Alice Wallenbergs Foundation through funding to the University of Maryland, and the Office of Naval Research through the URIP award to the Woods Hole Oceanographic Institution, grant number ONR-URIP N00014-92-J-1527. S. P. Ellner is pleased to acknowledge a grant from the Andrew W. Mellon Foundation to S. P. Ellner and Nelson G. Hairston, Jr.

LITERATURE CITED

- Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University of Chicago Press, Chicago, Illinois, USA.
- Barron, A. R. 1991a. Complexity regularization with application to artificial neural nets. Pages 561–576 in G. Rouskas, editor. Nonparametric function estimation and related topics. Kluwer Academic, Dordrecht, The Netherlands.
- Barron, A. R. 1991b. Approximation and estimation bounds for artificial neural networks. Pages 243–249 in Proceedings of the Fourth Annual Workshop on Computational Learning Theory, University of California at Santa Cruz, 5–7 August 1991. Morgan Kaufmann, San Mateo California, USA.
- Bjørnstad, O. N., M. Begon, N. C. Stenseth, W. Falck, S. M. Sait, and D. J. Thompson. 1998. Population dynamics of the Indian meal moth: demographic stochasticity and delayed regulatory mechanisms. *Journal of Animal Ecology* **67**:(1)110–126.
- Casdagli, M. 1992. A dynamic systems approach to modeling input-output systems. Pages 265–282 in M. Casdagli and S. Eubank, editors. Nonlinear modeling and forecasting. Addison-Wesley, New York, New York, USA.
- Colwell, R. R. 1996. Global climate and infectious disease: the cholera paradigm. *Science* **274**:2025–2031.
- Cushing, D. H., editor. 1983. Key papers on fish populations. IRL Press, Oxford, UK.
- Denman, K. L., and A. E. Gargett. 1995. Biological-physical interactions in the upper ocean: the role of vertical and small scale transport processes. *Annual Review of Fluid Mechanics* **27**:225–255.
- Denman, K. L., and T. M. Powell. 1984. Effects of physical processes on planktonic ecosystems in the coastal ocean. *Oceanography and Marine Biology Annual Review* **22**: 125–168.
- Dixon, P. A., M. J. Milicich, and G. Sugihara. 1999. Episodic fluctuations in larval supply. *Science* **283**:1528–1529.
- Dwyer, R. L., S. W. Nixon, C. A. Oviatt, K. T. Perez, and T. J. Smayda. 1978. Frequency response of a marine ecosystem subjected to time-varying inputs. Pages 19–38 in J. H.

- Thorp and J. W. Gibbons, editors. Energy and environmental stress in aquatic ecosystems. NTIS No. CONF-771114. U.S. DOE Symposium Series Number 48.
- Dwyer, R. L., and K. T. Perez. 1983. An experimental examination of ecosystem linearization. *American Naturalist* **121**:(3)305–323.
- Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, New York, New York, USA.
- Ellner, S. P., B. A. Bailey, G. V. Bobashev, A. R. Gallant, B. T. Grenfell, and D. W. Nychka. 1998. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *American Naturalist* **151**:425–440.
- Ellner, S., and P. Turchin. 1995. Chaos in a 'noisy' world: new methods and evidence from time series analysis. *American Naturalist* **145**:343–375.
- Fan, J., and I. Gijbels. 1996. Local polynomial modeling and its applications. Chapman & Hall, New York, New York, USA.
- Fogarty, M. J., H. Loeng, T. R. Osborn, and J. G. Sheperd, co-conveners. 1997. Recruitment dynamics of exploited marine populations: physical–biological interactions. International Council of Environmental Studies ICES, Copenhagen, Denmark. International Symposium, 22–24 September, Baltimore, USA.
- Green, P. J., and B. W. Silverman. 1994. Nonparametric regression and generalized linear models. Chapman & Hall, New York, New York, USA.
- Grenfell, B. T., K. Wilson, B. F. Finkenstadt, T. N. Coulson, S. Murray, S. D. Albon, J. M. Pemberton, T. H. Clutton-Brock, and M. J. Crawley. 1998. Noise and determinism in synchronized sheep dynamics. *Nature* **394**:(6694)674–677.
- Häerdle, W. 1990. Applied nonparametric regression. Cambridge University Press, Cambridge, UK.
- Higgins, K., A. Hastings, J. N. Sarvela, and L. W. Botsford. 1997. Stochastic dynamics and deterministic skeletons: population dynamics of Dungeness crab. *Science* **276**:1431–1435.
- Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.
- Hofmann, E. E., and T. M. Powell. 1998. Environmental variability effects on marine fisheries: four case histories. *Ecological Applications* **8**:S23–S32.
- Inoue, M., and H. Kamifukumoto. 1984. Scenarios leading to chaos in a forced Lotka–Volterra model. *Progress in Theoretical Physics* **71**:(5)931–937.
- Kendall, B. E., C. J. Briggs, W. W. Murdoch, P. Turchin, S. P. Ellner, E. McCauley, R. Nisbet, and S. N. Wood. 1999. Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80**:1789–1805.
- Kot, M., G. S. Saylor, and T. W. Schultz. 1992. Complex dynamics in a model microbial system. *Bulletin of Mathematical Biology* **54**:(4)619–648.
- Kot, M., W. M. Schaffer, G. L. Truty, D. J. Graser, and L. F. Olsen. 1988. Changing criteria for imposing order. *Ecological Modelling* **43**:75–110.
- Leirs, H., N. C. Stenseth, J. D. Nichols, J. E. Hines, R. Verhagen, and W. Verheyen. 1997. Stochastic seasonality and nonlinear density-dependent factors regulate population size in an African rodent. *Nature* **389**:(6647)176–180.
- Linhart, H., and W. Zucchini. 1986. Model selection. Wiley, New York, New York, USA.
- Linthicum, K. J., A. Anyamba, C. J. Tucker, P. W. Kelley, M. F. Myers, and C. J. Peters. 1999. Climate and satellite indicators to forecast rift valley fever epidemics in Kenya. *Science* **285**:397–399.
- Mackas, D. L., K. L. Denman, and M. R. Abbott. 1985. Plankton patchiness: biology in the physical vernacular. *Bulletin of Marine Science* **37**:652–674.
- McCaffrey, D., S. Ellner, D. W. Nychka, and A. R. Gallant. 1992. Estimating the Lyapunov exponent of a chaotic system with nonlinear regression. *Journal of the American Statistical Association* **87**:682–695.
- Nicholson, A. J. 1958. Dynamics of insect populations. *Annual Review of Entomology* **3**:107–136.
- Nychka, D. W., S. Ellner, D. McCaffrey, and A. R. Gallant. 1992. Finding chaos in noisy systems. *Journal of the Royal Statistical Society Series* **54**:399–426.
- Nychka, D., P. D. Haaland, M. A. O'Connell, and S. P. Ellner. 1998. FUNFITS: Data analysis and statistical tools for estimating functions. Pages 159–179 in D. Nychka, W. W. Piegorsch, and L. H. Cox, editors. Case studies in environmental statistics. Lecture Notes in Statistics Volume 132. Springer-Verlag, New York, New York, USA.
- Pascual, M., and H. Caswell. 1997a. From the cell-cycle to population cycles in phytoplankton–nutrient interactions. *Ecology* **78**:897–912.
- Pascual, M., and H. Caswell. 1997b. Environmental heterogeneity and biological pattern in a chaotic predator–prey system. *Journal of Theoretical Biology* **185**:1–13.
- Pascual, M., and S. A. Levin. 1999. From individuals to population densities: searching for the intermediate scale of nontrivial determinism. *Ecology* **80**:2225–2236.
- Patz, J. A., P. R. Epstein, T. A. Burke, and J. M. Balbus. 1996. Global climate change and emerging infectious diseases. *Journal of the American Medical Association* **275**:(3)217–223.
- Rinaldi, S., S. Muratori, and Y. Kuznetsov. 1993. Multiple attractors, catastrophes, and chaos in seasonally perturbed predator–prey communities. *Bulletin of Mathematical Biology* **55**:15–36.
- Rothschild, B. J. 1986. Dynamics of marine fish populations. Harvard University Press, Cambridge, Massachusetts, USA.
- Royama, T. 1992. Analytical population dynamics. Chapman & Hall, New York, New York, USA.
- Saitoh, T., N. C. Stenseth, and O. N. Björnstad. 1997. Density dependence in fluctuating grey-sided vole populations. *Journal of Animal Ecology* **66**:(1)14–24.
- Sauer, T., J. A. Yorke, and M. Casdagli. 1991. Embedology. *Journal of Statistical Physics* **65**:579–616.
- Schaffer, W. M. 1988. Perceiving order in the chaos of nature. Pages 331–350 in M. S. Boyce, editor. Evolution of life histories in mammals, theory and pattern. Yale University Press, New Haven, Connecticut, USA.
- Schaffer, W. M., S. Ellner, and M. Kot. 1986. Effects of noise on some dynamic models in ecology. *Journal of Mathematical Biology* **24**:479–523.
- Schaffer, W. M., L. F. Olsen, G. L. Truty, and S. L. Fulmer. 1990. The case for chaos in childhood epidemics. Pages 138–166 in S. Krasner, editor. The ubiquity of chaos. AAAS, Washington D.C., USA.
- Schwartz, I. B. 1992. Small amplitude, long period outbreaks in seasonally driven epidemics. *Journal of Mathematical Biology* **30**:473–491.
- Schwartz, I. B., and H. L. Smith. 1983. Infinite subharmonic bifurcations in an SEIR model. *Journal of Mathematical Biology* **18**:233–253.
- Smith, F. E. 1961. Density dependence in the Australian thrips. *Ecology* **42**:403–407.
- Steele, J. H., and E. W. Henderson. 1994. Coupling between physical and biological scales. *Philosophical Transactions of the Royal Society of London Series B* **343**:5–9.
- Stenseth, N. C., K.-S. Chan, E. Framstad, and H. Tong. 1998. Phase- and density-dependent dynamics in Norwegian lemmings: interaction between deterministic and stochastic

- processes. Proceedings of the Royal Society of London Series B **265**:1957–1968.
- Sugihara, G., and R. M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**:734–741.
- Takens, F. 1981. Detecting strange attractors in turbulence. Pages 366–381 in D. Rand and L. S. Young, editors. *Dynamical systems and turbulence*, Warwick 1980. Lecture Notes in Mathematics Volume 898. Springer-Verlag, New York, New York, USA.
- Tong, H. 1990. *Non-linear time series, a dynamic system approach*. Oxford University Press, Oxford, UK.
- Turchin, P., and A. D. Taylor. 1992. Complex dynamics in ecological time series. *Ecology* **73**:289–305.
- Wahba, G. 1990. *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

APPENDIX

Measurement noise

Error-free time series for P and H were obtained by numerical solution of the differential equations (Eq. 7). To these values we added lognormal errors, by adding normal errors on log-transformed scale:

$$\log \hat{P}(t_i) = \log P(t_i) + CV \times e_i$$

$$\log \hat{H}(t_i) = \log H(t_i) + CV \times e_i$$

where \hat{H} , \hat{P} are the measured values of H , P , and e_i denotes an independently and identically distributed random variable drawn from a normal distribution with mean 0, and variance 1. Back-transforming to arithmetic scale and doing a Taylor-series expansion in CV gives

$$\hat{P}(t_i) \doteq P(t_i)(1 + CV \times e_i)$$

$$\hat{H}(t_i) \doteq H(t_i)(1 + CV \times e_i)$$

so that the noise parameter CV is (approximately) the coef-

ficient of variation of the measured values, as stated in the text.

Dynamical noise

Dynamical noise was also lognormal, with constant variance on the log-transformed scale, and was applied as follows. Given values of P and H at time $t_i = ih$, the differential equations were solved to give noise-free values \tilde{P} and \tilde{H} at time $t_{i+1} = (i + 1)h$. Dynamical noise was then applied by setting the following:

$$\log P(t_{i+1}) = \log \tilde{P}(t_{i+1}) + CV \times e_i$$

$$\log H(t_{i+1}) = \log \tilde{H}(t_{i+1}) + CV \times e_i$$

where e_i denotes as before an independently and identically distributed random variable from $\mathcal{N}(0, 1)$. Thus conditional on $P(t_i)$, $\log P(t_{i+1})$ is normally distributed with mean $\log \tilde{P}(t_{i+1})$ and variance CV^2 . As with measurement noise, these normal perturbations on a log-transformed scale imply that the departures from the deterministic one-step-ahead solution have coefficient of variation approximately equal to CV.