Ecological Complexity 32 (2017) 134–143

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

Original Research Article

Circumventing structural uncertainty: A Bayesian perspective on nonlinear forecasting for ecology

Stephan B. Munch, Dr.[a,b,*], Valerie Poynor[c], Juan Lopez Arriaza[d]

[a] Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 110 Shaffer Rd., Santa Cruz, CA 95060, United States
[b] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, United States
[c] Department of Mathematics, California State University, Fullerton, CA 92831, United States
[d] Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064, United States

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form 24 August 2016
Accepted 30 August 2016
Available online 7 September 2016

Keywords:
Structural uncertainty
Nonlinear forecasting
Time-delay embedding
Gaussian process
Hierarchical model
Nonstationarity

ABSTRACT

As a consequence of the complexity of ecosystems and context-dependence of species interactions, structural uncertainty is pervasive in ecological modeling. This is particularly problematic when ecological models are used to make conservation and management plans whose outcomes may depend strongly on model formulation. Nonlinear time series approaches allow us to circumvent this issue by using the observed dynamics of the system to guide policy development. However, these methods typically require long time series from stationary systems, which are rarely available in ecological settings. Here we present a Bayesian approach to nonlinear forecasting based on Gaussian processes that readily integrates information from several short time series and allows for nonstationary dynamics. We demonstrate the utility of our modeling methods on simulated from a wide range of ecological scenarios. We expect that these models will extend the range of ecological systems to which nonlinear forecasting methods can be usefully applied.

Published by Elsevier B.V.

1. Introduction

Ecosystems involve a large and often unknown number of organisms and environmental factors. These components interact within and across species, age groups, genotypes, and environmental factors through time leading to systems that can be extremely complex. While it is possible to disentangle these sources of complexity for a handful of experimentally tractable, well-studied systems, identifying models for less-studied or intractable systems is a daunting task. Due to the system complexity, seemingly slight changes in model structure can lead to qualitatively different predictions (Wood and Thomas, 1999; Walters et al., 2016). This is particularly relevant when models are needed to inform conservation and management decisions (Lee et al., 1999; Wood and Thomas, 1999).

Alternatively, nonparametric time series methods allow us to study the dynamics of a system without having to specify a model. These methods originated in the 1980's and 90's based on Takens' (1981) theorem of time-delay embedding. Although initially restricted to single time series from an autonomous, deterministic system, these methods have since been generalized to multiple time series (Deyle and Sugihara, 2011) from non-autonomous systems with deterministic (Stark, 1999) and stochastic forcing (Stark et al., 2003). These methods have been of great use in physics (Buzug and Pfister, 1992), neurobiology (Kannathal et al., 2005), and econometrics (Mayfield and Mizrach, 1992) where long time series that are relatively free of observation noise are fairly common. Although a correctly specified parametric model is able to extract more information about the system (e.g. estimates of relevant parameters, reduced uncertainty), the insights gained from nonparametric methods tend to be robust to model misspecification.

Ecological applications of nonlinear forecasting were popular in the 1980's and 90's, including outstanding work by Sugihara (1994), Schaffer (1985), and Ellner and Turchin (1993); see Hastings et al. (1993) for a review. In the current literature, these methods seem to have been supplanted by more 'mechanistic' state-space models (see Patterson et al., 2008; and references therein) or linear models with time varying coefficients (e.g. Ives and Dakos, 2012).

* Corresponding author at: Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, United States.
E-mail address: steve.munch@noaa.gov (S.B. Munch).

http://dx.doi.org/10.1016/j.ecocom.2016.08.006
1476-945X/Published by Elsevier B.V.

The primary objections to using time-delay embedding in ecology seem to be that ecological time series are noisy, too short to define an attractor, and not stationary (Sugihara et al., 1990; Grenfell et al., 2001). Modern statistical methods, developed outside the time-delay embedding literature, may mitigate these objections. Hierarchical approaches allow information to be shared across data sets without assuming that they are identical (e.g., Shi et al., 2005; Bjornstad and Grenfell, 2001; Royle and Dorazio, 2008; Halstead et al., 2012). Nonstationary dynamics, in which the system drivers change through time, can be accommodated by allowing parameters to change as well (West and Harrison, 1997; Wikle, 2003; Ives and Dakos, 2012).

Here, we develop a Bayesian nonparametric framework for time-delay embedding that makes use of these modern statistical ideas. To do so, we use Gaussian process models to infer dynamics in delay coordinates. The chief advantages of the Gaussian process (GP) are its simple parameterization and ability to estimate with precision complicated nonlinear functions (O'Hagan, 1978). We then extend the GP framework to incorporate hierarchical inference from multiple time series and allow for nonstarionarity.

2. Methods and results

To begin, we briefly describe time delay embedding. We then introduce Gaussian processes as a tool for time-delay embedding and present a model specification that allows us to identify the relevant lags in the series. We extend this model to a hierarchical form that accommodates information from multiple related times series. Finally, we demonstrate how nonstationarity can be incorporated into the Gaussian process time-delay embedding model. These methods are then applied to a sequence of simulated data sets.

There is now a long history of applying Takens' theorem and time delay embedding in the ecological literature, see, e.g., Schaffer (1985), Ellner and Turchin (1993), and Sugihara (1994). However, most of the descriptions of the idea are steeped in the alien vernacular of topology. While there *are* deeper insights to be gained from the topological viewpoint, the practical upshot of Takens' theorem is that we are justified in modeling the dynamics of a single time series y_t (t = 1, . . . ,T) as a function of its lags. That is, $y_t = f(y_{t-1}, \ldots y_{t-L})$ for some unknown function f and 'embedding dimension' L which is at least twice the dimension of the attractor (Takens, 1981). Here, a fixed time step of 1 is assumed in keeping with the majority of ecological time series applications. In settings where the data are continuously sampled through time, an appropriate time lag, Δ, must also be determined and the model is $y_t = f(y_{t-\Delta}, \ldots y_{t-L\Delta})$.

Various approaches to nonlinear forecasting can be thought of as approximating the unknown function f, including polynomials (e.g. Turchin and Ellner, 1995), support vector machines (e.g. Mukherjee et al., 1997), and neural networks (e.g. Bakker et al., 2000). A particularly useful way to approximate f is using locally-weighted multiple linear regression, as in Sugihara's S-Map (Sugihara 1994). Specifically, a locally linear model of the form $y_t = \sum_{i=1}^{L} \beta_{t,i} y_{t-i} + \varepsilon_t$ is fit to the time series by weighted least squares. We have highlighted this method in particular because it was precisely locally linear models that motivated O'Hagan (1978) to introduce Gaussian processes (GP) as priors for flexible regression modeling from a Bayesian point of view. Here, we use the tools of Bayesian GP regression to construct a hierarchical approach to nonlinear forecasting that allows integration of information from multiple time series and explicitly deals with nonstationarity. The main text lays out the model specification and the simulations used to test each model. Further details of prior specification and posterior inference are provided in the Appendices.

2.1. Gaussian process time-delay embedding

Assume we have a scalar time series $y_1, \ldots, y_T$, and the goal is to estimate the unknown function f that maps the history of y into the future. To simplify notation, we'll use $\mathbf{x}_t = \{y_{t-1}, \ldots, y_{t-L}\}$ to represent the 'delay-coordinate vector' so that we are attempting to fit a model of the form $y_t = f(x_t) + \varepsilon_t$ for $t \in \{L+1, \ldots, T\}$. The errors ε_t are explicitly included here to account for approximation errors as well as process noise. For convenience, we assume that ε_t is (at least approximately) normally distributed with mean 0 and variance V_ε.

The shape of the function f is unknown and we would like to estimate it from the available data. In a Bayesian context, we do so by assigning a prior to f and updating the distribution over f given the observed data. Since we are inferring a function, we need a prior on a space of functions and the natural place to look for these is the theory of stochastic process. The Gaussian process is particularly convenient to work with as a prior for uncertain regression functions (O'Hagan, 1978). GP models have been used widely in spatial statistics under the moniker Kriging (Cressie, 1993). In addition, they have been used in population modeling to estimate the form of density dependence (Munch et al., 2005), test for the presence of Allee effects (Sugeno and Munch, 2013), and as a tool to assess model misspecification (Thorson et al., 2014). Rasmussen and Williams (2006) is an excellent source for additional background on modeling with Gaussian processes.

The GP is a continuous generalization of the multivariate normal distribution and as such is completely defined in terms of a mean and covariance. However, because it is a distribution on a function space, the mean and covariance are functions as well, denoted by $\mu(w)$ and $\Sigma(w, w')$, respectively. Here w and w' denote two arbitrary 'inputs'. At a single input, the marginal distribution for $f(w)$ is Gaussian with mean $\mu(w)$ and variance $\Sigma(w, w)$. For any finite collection of input points, $w = \{w_1, \ldots, w_n\}^T$ (superscript T denotes transpose), the marginal distribution is multivariate normal with mean vector $\mu(w) = \{\mu(w_1), \ldots, \mu(w_n)\}^T$ and covariance matrix $\Sigma(w, w^T)$ {i.e. the covariance matrix is constructed by evaluating the covariance function at all pairs of inputs, such that the i,j^{th} element is $\Sigma(w_i, w_j)$}.

In the present application we set the mean function to zero, $\mu = 0$, to indicate that we do not have any *a priori* information on the shape of the function we want to infer. This is particularly the case for time-delay embedding where the 'true' function is bound to be something rather complicated. In other applications, such as modeling density dependence or population productivity, we can use standard parametric models as the prior mean and use the GP to infer model misspecification (see e.g. Thorson et al., 2014; Sugeno and Munch 2013).

Setting the mean function to zero means that the covariance function informs the shape of f by specifying how strongly correlated realizations of f are at different inputs. In general, the slower the correlation decays with increasing separation between inputs, the smoother realizations of f will be. There are many choices for the covariance function (see e.g. Rasmussen and Williams, 2006; Paciorek and Schervish, 2004). The squared exponential correlation function, $R(d) = exp[-d^2]$ where $d = w - w'$, is among the most widely used.

In the present application the 'inputs' are the delay coordinate vectors, $\mathbf{x}_t = \{y_{t-1}, \ldots, y_{t-L}\}$ and we need to specify the covariance between f evaluated at the delay coordinates for two different times, e.g. $f(\mathbf{x}_t)$ and $f(\mathbf{x}_s)$ for times t and s, respectively. We build