NOTE

Quantitative argument for long-term ecological monitoring

Alfredo Giron-Nava¹, Chase C. James¹, Andrew F. Johnson¹, David Dannecker², Bethany Kolody¹, Adrienne Lee², Maitreyi Nagarkar¹, Gerald M. Pao³, Hao Ye¹, David G. Johns⁴, George Sugihara¹,*

¹Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92037, USA  
²Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92037, USA  
³Salk Institute for Biological Studies, 10010 N Torrey Pines Road, La Jolla, CA 92037, USA  
⁴Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

ABSTRACT: Although it seems obvious that with more data, the predictive capacity of ecological models should improve, a way to demonstrate this fundamental result has not been so obvious. In particular, when the standard models themselves are inadequate (von Bertalanffy, extended Ricker etc.) no additional data will improve performance. By using time series from the Sir Alister Hardy Foundation for Ocean Science Continuous Plankton Recorder, we demonstrate that long-term observations reveal both the prevalence of nonlinear processes in species abundances and an improvement in out-of-sample predictability as the number of observations increase. The empirical results presented here quantitatively demonstrate the importance of long-term temporal data collection programs for improving ecosystem models and forecasts, and to better support environmental management actions.

KEY WORDS: Long-term monitoring · Predictability · Nonlinearity · Time series · Population dynamics · Ecological data

INTRODUCTION

Although there is a strong consensus for the benefits of long-term ecological monitoring and the creation of long-term datasets, there are significant fiscal and political challenges facing the continuation of these efforts. Given typical funding horizons and the inherent costs associated with long-term research programs (Callahan 1984, Magurran et al. 2010), it is no surprise that the majority of ecological research occurs within short time frames (Magnuson 1990).

Ecosystem dynamics, however, may not be decipherable from observations on short timescales. Indeed, there is a growing view that nonlinear and non-stationary dynamics are common and may require longer time series for study (Sugihara & May 1990, Hsieh et al. 2005, Sugihara et al. 2012, Glaser et al. 2014). While most ecological models can be fit to short time series, they may subsequently fail when used to make predictions into the future or to explain changes that occur over the long term (Pilkey & Pilkey-Jarvis 2007, Evans et al. 2012). The challenge of models to predict well has led some to suggest that scientific efforts, and particularly those in fisheries science, should be shifted away from forecasting as an attainable goal (Schindler & Hilborn 2015). This view contradicts our natural expectation that ecosystems behave in nonrandom ways and that with more data (i.e. longer time series), models should generally improve. Although we commonly acknowledge out-
of-sample prediction to be the ultimate test for understanding, many current parametric ecosystem models fail to pass this test (Magurran et al. 2010). A possible reason is that these models are not well suited for describing nonlinear and nonstationary dynamics. Insofar as nonlinearity is to be expected in ecosystems, it is almost guaranteed that improperly formulated models will fail to achieve successful predictions, no matter how much additional data are provided.

We use 2 distinct measures to quantify the amount of data in each time series: time series length and data availability. Time series length refers to the number of data points in a time series, whereas data availability is defined as the number of nonzero values within a time series. We use the term data availability because we recognize that zeros could represent an absence of an organism or a lack of detection, and we are unable to distinguish between the two. Our analyses that examine the effect of time series length use subsampled segments from the time series with the fewest nonzero values or greatest data availability ‘Effect of time-series length’.

Effect of data availability

We investigate the degree of nonlinearity and predictability using 2 methods of time series analysis: simplex projection (Sugihara & May 1990) and S-maps (Sugihara 1994). Simplex projection is used to assess the one-step-ahead predictive skill for each time series and is measured by the Pearson correlation coefficient ($\rho$) between predictions and observations. Following Glaser et al. (2014), to minimize overfitting, we use leave-one-out cross validation, where the point being forecast is excluded from the data used to construct the forecast. Different values for the embedding dimension ($E$) were tested (ranging from $E = 1$ to $E = 10$), and the optimal embedding dimension was selected as that which produced the highest $\rho$. We note that the optimal $E$ is a property of the data (how noisy and how long the time series are) as well as the underlying system (how complex it is in terms of numbers of variables) and indicates the number of lags that best resolves one-step-ahead forecasts. While $E$ can be informative about the complexity of the system or the number of drivers (Liu et al. 2012, Glaser et al. 2014), we caution against over-interpretation.

S-maps are used to test for nonlinear state dependence (Sugihara 1994). This involves demonstrating curvature in the attractor and is quantified by the improvement in $\rho$ ($\Delta \rho$) obtained with a nonlinear versus a linear model. Statistical significance for $\Delta \rho$ was determined using a randomization test. For each time series, we generated 100 surrogate time series and calculated a $\Delta \rho$ for each surrogate using the previously determined $E$. This creates a null distribution that can be compared to the original (nonshuffled) $\Delta \rho$. Nonlinearity is indicated when the $\Delta \rho$ statistic for the original time series is greater than the 0.95 quantile of the null distribution.
To examine the effect of time series length as a property independent of taxonomic identity, we focused on the 23 time series with the greatest data availability (≥461 nonzero data points) and which were identified as significantly nonlinear. For each time series, we investigated the effect of varying the data length on predictability using randomly selected contiguous segments of length: 25, 50, 100, 200, 400. The subsampling was repeated 25 times at each length for each species (so that 575 [23 × 25] subsamples were used to compute forecast performance at each length). Predictions were made using a form of leave-one-out cross validation where the predicted point is excluded to minimize overfitting. We then tested for nonlinear state dependence using S-maps as described in the previous subsection.

**Significance of nonlinearity as a function of data completeness**

To test whether data availability or time series length are significant predictors of nonlinearity, we treated data availability and time series length by binning them into discrete categories (Fig. 1) and recorded the detection of nonlinearity as a binary response. Then, we fit a logistic regression using the R glm function (binomial family).

**RESULTS AND DISCUSSION**

The S-map analysis shows that nonlinear dynamics are more readily identified as data availability increases (Fig. 1a; p < 0.01; logistic regression, df = 207). For the 90 taxa with the lowest data availability (≤37 nonzero data points out of 672), only 11% (10 taxa) showed significant nonlinear dynamics. In contrast, of the 90 taxa with 38 to 460 nonzero values, 57% (51 taxa) showed significant nonlinear dynamics. Finally, among the 28 taxa with the highest data availability (≥461 nonzero values), 82% (23 taxa) showed significant nonlinear dynamics. In other words, the time series that are most complete also show stronger evidence for nonlinear dynamics. To test whether this effect could be driven by the specific taxa that happen to appear most often in the data, we also analyzed subsampled time series (artificially shortened from the 23 nonlinear taxa with the most data availability), finding a similar pattern (Fig. 1b; p < 0.01; logistic regression, df = 23). It has been recognized that aggregating taxon levels can obscure nonlinear dynamics detection (Liu et al. 2014); however, we have shown that nonlinearity detection increases with time series length regardless of the taxonomic resolution. The subsampling procedure here obviates any taxon-specific effect related to aggregation that could reduce nonlinearity in more aggregated (linearly summed) groupings (Sugihara et. al. 1999).

A similar advantage with increased data holds for prediction; in general, greater data availability also corresponds to higher forecast skill (Fig. 2a). Using the same subsamples as in the previous paragraph, Fig. 2b shows that predictability also increases with time series length. Thus, we expect that as more data are collected, populations will be more readily identified as nonlinear and that the longer time series will enable better predictions. However, we note that even at the longest time series lengths (≥600 nonzero values in Fig. 2a or 400 time points in Fig. 2b), the level of forecast skill can vary substantially. These differences in predictability could partially reflect differences in the natural response times (e.g. generation times), leading to differences in the density of points on the attractor (Table 1), with shorter generation times producing denser attractors;
however, they could also reflect exogenous effects, such as environmental drivers, that are not captured in the abundance time series and may therefore need to be included explicitly in forecast models (see Dixon et al. 1999, Deyle et al. 2013, and Harford et al. 2017 for examples).

Finally, we note that in addition to leave-one-out cross validation, for the 23 taxa with the highest data availability, nearly identical out-of-sample prediction results are obtained with 2-fold cross validation. Here each half of the data series is used to predict the other half, and the predictions for the entire series are combined to calculate performance (Fig. 3). This demonstrates the robustness of the cross validation results and shows that the dynamics are stationary or essentially the same in both halves (Sugihara & May 1990).

Table 1. Taxa whose time series have more than 600 nonzero values and the Pearson correlation coefficient ($\rho$) from nonlinear prediction (from Fig. 2a)

<table>
<thead>
<tr>
<th>Taxon</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oithona sp.</td>
<td>0.71</td>
</tr>
<tr>
<td>Total copepods</td>
<td>0.70</td>
</tr>
<tr>
<td>Para-pseudocalanus sp.</td>
<td>0.67</td>
</tr>
<tr>
<td>Acartia sp.</td>
<td>0.58</td>
</tr>
<tr>
<td>Chaetognaths</td>
<td>0.54</td>
</tr>
<tr>
<td>Chlorophyll index</td>
<td>0.53</td>
</tr>
<tr>
<td>Calanus I–IV</td>
<td>0.48</td>
</tr>
<tr>
<td>Calanus finmarchicus</td>
<td>0.44</td>
</tr>
<tr>
<td>Fish larvae</td>
<td>0.34</td>
</tr>
<tr>
<td>Calanus helgolandicus</td>
<td>0.34</td>
</tr>
<tr>
<td>Temora longicornis</td>
<td>0.31</td>
</tr>
<tr>
<td>Total hyperiids</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Documenting that ecosystems are capable of exhibiting predictability and that their dynamics are nonlinear is of considerable interest to agencies involved in data collection and to resource managers who must use those data. The reasonable expectation is that predictive models will improve with more data. However, if a model is a poor description of a system, no amount of additional data will improve predictability (MacNally 2000, DeAngelis & Yurek 2015). For example, fishery stock prediction has been a difficult challenge despite continual model refinements and new data inputs (Ward et al. 2014, Schindler & Hilborn 2015, Ye et al. 2015). The limitations of current ecosystem models to predict out of sample may arise in part from observational error in the data. However, we believe that a more significant problem is the mismatch between the traditional assumptions of equilibrium and linearity (DeAngelis & Waterhouse 1987) and the increasing observations that many ecological systems are not in equilibrium but are demonstrably nonlinear (Fromentin & Powers
Although short time series can be challenging to the identification of nonlinear dynamics and construction of empirical models, there have been several recent advances on this front (Hsieh et al. 2008, Glaser et al. 2011, Clark et al. 2015, Ye & Sugihara 2016).

Data-driven approaches where causal variables and functional relationships are determined empirically may offer a viable alternative to inductive equation-based approaches. For example, Sugihara et al. (2012) introduce an EDM method for using time series to identify the causal drivers of ecosystem dynamics, and several others (Dixon et al. 1999, Deyle et al. 2013, Harford et al. 2017) provide examples of incorporating these environmental effects into EDMs to forecast future ecosystem states—including apparently random events such as red tides (McGowan et al. 2017). These approaches do not rely on hypothesized equations but instead infer relationships deductively as they appear in the data. With continued monitoring and longer time series, the ability of such techniques to describe nonlinear behavior will improve our understanding of ecological mechanisms, where unraveling the interdependence between environmental factors and endogenous population dynamics is certain to be critical for managing ecosystems in the context of climate change (Persing et al. 2015). Thus, as predictive data-driven approaches continue to gain traction, investments in long-term data collection will surely yield long-term payoffs.

CONCLUSIONS

The CPR at SAHFOS is a prime example of how continuous long-term projects help to create new perspectives in ecology. By providing robust, long-term datasets, this program captures variability in species abundance that occurs on temporal scales ranging from interannual cycles to multidecadal shifts (Barton et al. 2003). Furthermore, SAHFOS, in collaboration with partners from around the globe, is now involved in conducting plankton counts in nearly every major ocean basin on earth. In addition to providing a long-term description of population dynamics, these spatially broad environmental monitoring programs can provide regional insights into global problems. Beyond the common sense value of collecting these data, our analyses provide a quantitative justification for continued support of these programs. These data are critical for predictability and understanding, which are particularly important given increasing threats to global ecosystems, such as human exploitation pressures and climate change (Halpern et al. 2008).


Acknowledgements. Thanks to Martin Edwards, Darren Stevens, and SAHFOS for providing the authors with CPR data. This analysis was supported by US Department of Defense–Strategic Environmental Research and Development Program 15 RC-2509; Lenfest Ocean Program 0002 8335; the Sugihara Family Trust; the Deutsche Bank–Jameson Complexity Studies Fund; and the McQuown Fund and McQuown Chair in Natural Sciences, University of California San Diego. A.G.N. was funded by CONACYT (CVU 579904) and Fulbright Garcia-Robles (LASPAU ID 20140963) doctoral program fellowships. A.F.J. was supported by NSF grant DEB-1632648 (2016). B.K. and M.N. were funded by NSF grant DGE-1144086. The authors declare no conflicts of interest.

LITERATURE CITED


