

In the format provided by the authors and unedited.

Nonlinear population dynamics are ubiquitous in animals

T. J. Clark * and Angela D. Luis

Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT, USA. *e-mail: tyler3.clark@umontana.edu

Nonlinear population dynamics are ubiquitous in animals

Tyler James Clark & Angela D. Luis

Supplementary Information

1. Supplementary Methods

Time-series data selection procedure. Time-series data were obtained from a variety of sources, including global databases, past peer-reviewed literature which collated databases, and gray literature. This included:

1) NERC Centre for Population Biology, Imperial College (2010). *The Global Population Dynamics Database Version 2*. <http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html>. This database includes ~ 5,000 time-series of 1400 animal and plant species throughout the world. The database was accessed using the R package “rgpdd”⁵⁸.

2) Other Databases, including:

a) Living Planet Index (2016) *Living Planet Index database*.

<http://www.livingplanetindex.org>. This database is managed by the Zoological Society of London and WWF International, including 14,152 time-series of more than 4,000 animal species, which were filtered so that none overlapped with information already collected from the Global Population Dynamics Database (GPDD).

b) ICES Database of Stock Assessments (2018). *International Council for the Exploration of the Sea*. ICES, Copenhagen, Denmark. <http://www.ices.dk>.

3) Past peer-reviewed literature which had previously collated data from gray literature, websites, and books to augment the GPDD^{37,50}.

4) Literature searches. Primary literature was found by searching for key terms (e.g. “time-series”) or known, well-studied species.

Filtering procedure. Past research has identified that the GPDD includes inconsistencies and ambiguities which may lead to improper analysis if uncorrected³⁷. To resolve this, we strictly filtered datasets according to the following criteria:

- 1) Time-series were limited to those which were ≥ 30 population counts per time series, the minimum number suggested for our nonlinear forecasting method³².
- 2) Other animal datasets, including bivalves, sharks, gastropods, and crustaceans were removed because there were not enough time-series of each taxonomic class for statistical analysis.
- 3) The GPDD was assessed with a subjective reliability score from 1 (poor) to 5 (excellent)³⁷. Data were removed if ≤ 2 . An example of a dataset ranked as a 1 is historical fur trapping or export records from North America.
- 4) All log-transformed ($\log_{10}$) datasets were back-transformed to produce values directly proportional to the population size.
- 5) Time-series were removed which were exclusively made up of zeroes, or which had negative abundances.
- 6) Time-series were required to vary over at least five different unique values.
- 7) It was common for time-series to include a long sequence of zeros, which could represent population extirpation, emigration, or sampling error. Due to uncertainty about the meaning of these values, we performed separate supplementary analyses under three separate filtering criteria: 1) Exclude time-series with > 20 repeating zeroes; 2) Exclude time-series with > 5

repeating zeroes; 3) Exclude time-series with > 1 repeating zeroes. These analyses do not change our findings or conclusions (Extended Data Fig. 6), so we included all time-series regardless of repeating zeroes for our primary nonlinear forecasting analysis.

8) Due to overrepresentation of a small number of well-studied species (e.g., multi-species monitoring of aphids and moths at single field stations)³⁶, we conducted statistical analysis only on a single, random time-series from each species. Subsequent random sampling did not change our results.

2. Supplementary Tables

Table 1. Model selection results for analysis of nonlinearity.

Model	Parameters	ΔAIC
TC + PC1 + E + ρ + CV	6	0
TC + PC1 + E + ρ + N + CV	7	1.80*
TC + PC1 + PC2 + E + ρ + N + CV	8	3.80
TC + PC1 + PC2 + E + ρ + N + CV + G	9	5.10
TC + PC1 + PC2 + E + ρ + N + CV + G + Data Type	10	11.50
Intercept-only	1	12.55
TC + PC1 + PC2 + E + ρ + N + CV + G + Data Type + Data Type*CV	11	23.70

BIC and AIC model results were the same, so we only show AIC results for simplicity. *TC* is

taxonomic class, *PC1* and *PC2* are principal components of life-history traits, *E* is embedding dimension, ρ is forecast skill, *N* is sample size, *CV* is coefficient of variation, *G* is number of generations monitored in each time-series, *Data Type* represents different sampling methods, and “Intercept-only” is a model without any covariates.

*Because ΔAIC was < 2 between our top model and this one⁵⁹, there may be evidence that *N* (representing time-series length) is an important coefficient in a possible model structure for

nonlinearity. However, the coefficient's estimate in log odds = -0.247, SE = 0.239, $p = 0.300$ for this model, so we expect that N has a negligible role in explaining the variation in nonlinearity for our data.

Table 2. Model selection results for statistical analysis of dimensionality, E .

Model	Parameters	ΔAIC
PC1 + N	3	0
PC1 + PC2 + N	4	1.22*
PC1 + PC2 + N + CV	5	3.17
PC1 + PC2 + N + CV + G	6	5.15
Intercept-only	1	5.97
TC + PC1 + PC2 + N + CV + G	7	9.5
TC + PC1 + PC2 + Trophic + N + CV + G	8	15.22
TC + PC1 + PC2 + Trophic + N + CV + G + Data Type	9	20.1
TC + PC1 + PC2 + Trophic + N + CV + G + Data Type + Data Type*CV	10	23.67

TC is taxonomic classification, *PC1* and *PC2* are principal components of life-history traits,

Trophic is the trophic level (e.g., carnivore), N is sample size, *CV* is coefficient of variation, G is number of generations monitored in each time-series, *Data Type* represents different sampling methods, and “Intercept-only” is a model without any covariates.

*Because ΔAIC was <2 between our top model and this one⁵⁹, there may be evidence that *PC2* (representing fecundity in the principal components analysis) is an important coefficient in a possible model structure for dimensionality. However, the coefficient's estimate in log odds = -0.032, $p = 0.513$ for this model, so we expect that *PC2* has a negligible role in explaining the variation in dimensionality for our data.

Table 3. Model selection results for statistical analysis of forecast skill, ρ .

Model	Parameters	ΔAIC
TC + PC1 + Predict Category + CV	5	0
TC + PC1 + Predict Category + CV + G	6	0.64*
TC + PC1 + PC2 + Predict Category + CV + G	7	1.56*
TC + PC1 + PC2 + Predict Category + N + CV + G	8	2.62
TC + PC1 + PC2 + Predict Category + N + CV + G + Data Type	9	4.32
TC + PC1 + PC2 + Predict Category + N + CV + G + Data Type + Data Type*SD	10	6.75
Intercept only	1	307.68

TC is taxonomic classification, *PC1* and *PC2* are principal components of life-history traits,

Predict Category is linear, nonlinear, or non-predictable, *N* is sample size, *CV* is coefficient of variation, *G* is number of generations monitored in each time-series, *Data Type* represents different sampling methods, and “Intercept-only” is a model without any covariates

*Because ΔAIC was <2 between our top model and these ones⁵⁹, there may be evidence that *PC2* (representing fecundity in the principal components analysis) and *G* (number of generations monitored) are important coefficients in a possible model structure for dimensionality. However, *PC2* coefficient’s estimate in log odds = -0.016, SE = 0.059, $p = 0.784$ and *G* coefficient’s estimate in log odds = -0.060, SE = 0.0607, $p = 0.323$, so we expect that both have a negligible role in explaining the variation in dimensionality for our data.

3. Supplementary Results

There were control variables that were retained in our most parsimonious models for *E*, ρ , and nonlinearity. Time-series which were longer (*N*) were more likely to have higher dimensionality (*E*) (Extended Data Fig. 4). This is consistent with past analysis using nonlinear forecasting to

predict time-series^{7,32}. Additionally, time-series with a higher CV (coefficient of variation) were found to have a higher nonlinearity and forecast skill (ρ) (Extended Data Fig. 2; 5), suggesting that this modeling technique may be better at forecasting for time-series with high variability⁷.

Supplementary References.

58. Boettiger, C., Harte, T., Chamberlain, S. & Karthik, R. rgpdd: R Interface to the Global Population Dynamics Database. R package version 0.1 <https://github.com/ropensci/rgpdd> (2017).
59. Burnham, K. P. & Anderson, D. R. Model Selection and Inference: A Practical Information-Theoretic Approach (Springer, 2013).