DETECTING CHAOS REQUIRES CAREFUL ANALYSIS OF PERIODIC DATA

Torsten Lindström
University of Kalmar, Kalmar, Sweden

EXTENDED ABSTRACT

This paper determines the properties of mechanistic and data-based models in population biology side by side. The mechanistic modeling approach starts from mechanisms assumed to take part in the process under study. For instance, a predator-prey model in continuous time may be written as follows,

$$\dot{x} = rx(1-x) - axy/(1+abx)$$
$$\dot{y} = -\delta y + amxy/(1+abx)$$

The first term in the first equation describes growth, natural death, and competition, the second term describes predation. In the second equation, the second term describes the conversion of prey biomass into predator biomass, whereas the first term describes natural death of the predator population. Each term in the equations can thus be interpreted as a model of each mechanism that is taken into account. The different terms have been justified with various successes, and the second term in the first equation has that far the most compelling mechanistic explanation [1]. The mechanistic justification of the first term of the first equation has, on the other hand been called into question [2].

Differential equation models has an advantage over discrete time models in population dynamics [3, 4] in that on an infinite-decimal time-scale, the different processes tend to operate independently, so that we can take into account more processes and build up larger models by adding more terms to the models [5]. Consider, for instance, the discrete predator-prey model [6].

$$X' = \frac{\beta X \exp(-Y)}{1 + X \dfrac{1 - \exp(-Y)}{Y}}$$
$$Y' = MX(1 - \exp(-Y))$$

Here the different processes do not operate independently, because the predation process (in the first equation described by the factor $\exp(-Y)$) removes individuals from the prey population, which relaxes the competition process (described by the factor $1/(1+X)$). Therefore, the denominator in the first equation becomes complicated and this puts limits on how far we can proceed when formulating population models in discrete time [6, 7].

We shall restrict this study to one dimensional discrete population models, since in this case there exist a theory of their generic dynamical properties [8]. In many cases, resources are used by individuals who do not obtain enough resources for reproduction, which causes a

decrease of population size [9] and therefore a large class of one-dimensional population models possess humped right hand sides.

The mechanistic model under study in this paper is the Sheperd model, which describes the growth of a single species population under non-linear intra-specific competition corresponding to a humped right hand side [4]. As predicted by May and Oster [10] and Devaney [8] such models undergo the period-doubling route to chaos in many cases. It is no difference if we select another model to describe the nonlinear competition mechanisms, in most cases the same bifurcations will happen if the growth rate parameter is increased [8]. The reason for restricting this first study to single species models is that there exists sufficiently general theory for how many periodic attractors such maps may have [11]. In fact, it turns out that the considered cases do not possess more than at most one periodic attractor.

When ecological data are collected, the map describing the nonlinear mechanisms is hardly available, and we may call into question whether such a correct model even exists [12]. In order to select an appropriate model, different approaches have been developed and most attempts use least squares selection or likelihood methods combined with some procedure that penalizes the number of parameters used in the model. One well-known procedure is cross-validation (CV) that essentially trains the model to predict data left out [13]. Each data point in the material is left out from the data material after each other and the remaining data are used for selecting the model that predicts the data point that was left out. Generalized cross-validation (GCV) were used for many decades instead of cross-validation in order to correct for outliers a save computation efforts [14]. Related statistical methods are the Akaike information criterion AIC [15] and its modifications [16]. These methods possess cross-validation properties, too, but the idea with those methods is to find a model that is as close as possible to the correct model if such a model exists.

Recently, Lindström and Thunberg [17] stressed that the generic dynamical properties of typical maps fitted to data differ from those of maps usually accepted for mechanistic modelling. In fact, the simplest nonlinear models used in statistical data analysis (continuous TAR (1) models) do not possess period-doublings at all, so their build-up of complicated dynamics is entirely different from the generic bifurcation routes in the mechanistic modelling approach [8, 17]. Yet, in one-dimension they still possess at most one periodic attractor for each parameter value and this property agrees with the corresponding property for many mechanistic single species models. The objective of this paper is to elucidate the relation between the dynamics of mechanistic models and the dynamics of commonly used statistical models fitted to their data. We explore this relationship in the noise-free limit simply because the relationship tends to be most easy to understand in this case. This is most evident in the case when the dynamics of the mechanistic model corresponds to a fixed point. If a line is fitted to this data-point it describes the fixed point equally well regardless of which slope is selected. If a small amount of noise is added, the slope of the line agrees reasonably well with the slope of the mechanistic map at the fixed point. We show that the dynamics of continuous TAR (1)-models can describe the dynamics of the mechanistic Sheperd map close to its attractor as long as the mechanistic map possesses either a fixed point or a periodic attractor of period two and that AIC can select perfectly between fixed point cases and two-periodic cases in the noise free limit. Outside this regime AIC cannot select a TAR model describing the deterministic dynamics correctly in the noise-free limit. If three- or four-periodic attractors are to be described, there are many cases described by corresponding three- or four-periodic TAR-repellers, respectively, and chaotic TAR-attractors.

We continue by analyzing the results obtained when fitting a Ricker [18] model to the Sheperd data. The Ricker model belongs to the class of mechanistic models that have negative Schwartzian derivative, possessing a unique attractor, and undergo the period doubling route to chaos. There is therefore some hope that fitting this false nonlinear model to the Sheperd data could recapitulate the dynamic patterns of the true nonlinear Sheperd model. But in fact, it turns out that this model is able to follow the dynamic behavior of the Sheperd model only up to its second period doubling bifurcation. According to information theoretic criteria, a slightly false nonlinear model is far behind the AR-TAR approach. Therefore, we cannot expect that we are able to get any other result than that AR-TAR models describe data best unless we have included the correct nonlinear model in the model set, which cannot be assumed.

We close this study by fitting broken line models with a possibly large number breakpoints to the Sheperd data. These models are Multivariate Adaptive Regression Spline (MARS) models [19] and they are ranked in this study by the GCV criterion. These maps should have enough flexibility to be able to estimate the derivatives and location at an arbitrary periodic solution, but since a limit amount of data is assumed, this will not be the case. The qualitative dynamical properties of these models are not fully known and we cannot, for instance, guarantee these models to have at most one periodic attractor, as were the case with the AR and TAR models. Our study indicates that the switching behavior reported for the AR-TAR approach is still present and remain a good explanation for what happens when attempting to estimate data from mechanistic models with models highly ranked through criteria like AIC or GCV. On the other hand it is impossible to recover the underlying mechanisms by use of linear (AR) models since they possess unrealistic ecological properties like unbounded per capita population growth rates [20] and these problems are not removed by use of TAR models. We can therefore, in general, never expect that models fitted to data describes dynamical properties or the underlying mechanisms of governed by the underlying true mechanistic model adequately. However, attempting to fit a slightly false nonlinear model that possesses the same bifurcation sequence as the mechanistic model to data does even worse.

REFERENCES

[1] C. S. Holling. Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, 91(7):385--398, 1959.
[2] B. W. Kooi, M. P. Boer, and S. A. L. M. Kooijman. On the use of the logistic equation in models of food chain. *Bulletin of Mathematical Biology*, 60:231--246, 1998.
[3] M. Gyllenberg, I. Hanski, and T. Lindström. A predator-prey model with optimal suppression of reproduction in the prey. *Mathematical Biosciences* 134:119--152, 1996.
[4] M. Gyllenberg, I. Hanski, and T. Lindström. Continuous versus discrete single species population models with adjustable reproductive strategies. *Bulletin of Mathematical Biology*, 59(4):679--705, 1997.
[5] J. A. J. Metz and O. Diekmann. *The Dynamics of Physiologically Structured Populations*. Springer-Verlag, 1986.
[6] T. Lindström. On the dynamics of discrete food-chains: Low- and high-frequency behavior and chaos. *Journal of Mathematical Biology*, 45:396--418, 2002.
[7] T. Lindström. Dependencies between competition and predation - and their consequences for initial value sensitivity. *SIAM Journal of Applied Mathematics*, 59(4):1468--1486, 1999.

[8] R. Devaney. *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley Publishing Company, Inc, 1989.

[9] A. Lomnicki. *Population Ecology of Individuals*, volume 25 of *Monographs in Population Biology*. Princeton University Press, Princeton, New Jersey, 1988.

[10] R. M. May and G. F. Oster. Bifurcations and dynamic complexity in simple ecological models. *The American Naturalist*, 110(974):573--599, 1976.

[11] J. Guckenheimer, G. Oster, and A. Ipaktchi. The dynamics of density dependent population models. *Journal of Mathematical Biology*, 4:101--147, 1977.

[12] K. P. Burnham and D. R. Andersson. *Model selection and multimodel inference*. Springer, 2002.

[13] T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Chapman & Hall, London, 1990.

[14] P. J. Green and B. W. Silverman. *Nonparametric regression and Generalized Linear Models*, Volume 58 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1994.

[15] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *2nd International Symposium on Information Theory*, pages 267--281. Akademia Kiado, Budapest, 1973.

[16] C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297--307, 1989.

[17] T. Lindström and H. Thunberg. An elementary approach to dynamics and bifurcations of skew tent maps. Preprint, 2007.

[18] W. E. Ricker. Stock and recruitment. *Journal of the Fisheries Research Board of Canada*, 11(5):559--623, 1954.

[19] J. H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics* 19(1):1--141, 1991.

[20] M. Lima. The dynamics of natural populations: feedback structures in fluctuating environments. *Revista chliena de historia natural*, 74(2):317--329, 2001.