AIP Chaos
An Interdisciplinary Journal of Nonlinear Science

Degeneracy of time series models: The best model is not always the correct model

Kevin Judd and Tomomichi Nakamura

Citation: Chaos **16**, 033105 (2006); doi: 10.1063/1.2213957
View online: http://dx.doi.org/10.1063/1.2213957
View Table of Contents: http://chaos.aip.org/resource/1/CHAOEH/v16/i3
Published by the American Institute of Physics.

Degeneracy of time series models: The best model is not always the correct model

Kevin Judd
School of Mathematics and Statistics, The University of Western Australia, Nedlands, Perth, Western Australia, 6009 Australia

Tomomichi Nakamura
Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Received 3 March 2006; accepted 24 May 2006; published online 31 July 2006)

There are a number of good techniques for finding, in some sense, the best model of a deterministic system given a time series of observations. We examine a problem called *model degeneracy*, which has the consequence that even when a perfect model of a system exists, one does not find it using the best techniques currently available. The problem is illustrated using global polynomial models and the theory of Gröbner bases. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2213957]

Suppose one has a deterministic dynamical system from which a time series of state measurements has been collected. The measurements will have some measurement error. The goal is to find the *correct* model of the system, which is, in a sense, the goal of much of physics. Whether a correct model actually exists for any *real* physical system is open to question,[1,2] but let us entertain the idea for the present. We assume then that there is a model *identical* to reality; this is the *correct* model. There have been developed many useful techniques for finding the *best* model of a time series of a dynamical system, where *best* is measured in terms of both accuracy and simplicity given the available information. One purpose of this paper is to illustrate a simple fact that the *best* model of a time series of a dynamical system is not necessarily the *correct* model, and consequently the current best practice for finding the best model will not find the correct model. Some readers may not find this a surprising or profoundly new observation, although it appears not to have been clearly explicated in the literature. The second purpose of this paper, which we believe is truly new, is an investigation of whether there is any relationship between the best model and the correct model. We show that for polynomial difference equations there is a mathematically very beautiful relationship between the best model and the correct model, which we refer to as *model degeneracy*. Our analysis opens the way to finding, or at least further limiting, what the correct model can be.

I. TIME SERIES MODELS

To illustrate our point we consider the chaotic dynamical system described by Henon[3] that, when formulated as a second order difference equation, is given by

$$x(t) = \alpha + \beta x(t-1)^2 + \gamma x(t-2), \tag{1}$$

with $\alpha=1.0$, $\beta=-1.4$, and $\gamma=0.3$. This is our *reality* and, by our definition, the unique *correct* model. (We will see in a moment that mathematically this equation is, in a dynamical sense, not unique.)

Given a time series of observations of $x(t)$, which includes observational errors, the goal is to discover that Eq. (1) is the correct model.

A well-studied problem is the estimation of the parameters, α, β, and γ, given a time series of observations corrupted by noise.[4–14] Most of this work assumes the structure of the difference equation (1) is known, and the only problem is to determine the correct values of the parameters. Here we are concerned with the broader problem of determining the correct structure of the difference equation as well.[13,15]

Our problem to be solved is to start with a class of models and to find the correct model within it. Suppose for the system (1) we restrict attention to the class of the *polynomial difference equations*, where $x(t)$ is expressed as a linear combination of terms of the form $\Pi_i x(t-i)^{n_i}$, where the n_i are non-negative integers. Usually the class is restricted further by having an upper bound on the order $i \leqslant L$, and an upper bound on the degree $\Sigma_i n_i \leqslant D$ for each term. We consider polynomial difference equations because they allow an elegant mathematical analysis of the problem we describe, but the problem is fundamental, and not peculiar to this model class.

The current best practice, in general terms, recommends obtaining the best model from time series data using least squares or total least squares[9,10] to fit parameters and using some information or statistical criteria like the minimum description length,[13,15–18] cross validation, bootstrapping, or a Bayesian approach, to select the best model structure. The statistical literature on cross validation, bootstrapping, and Bayesian methods is immense, but generally not of relevance here because it concerns itself with stochastic systems, and here we are interested in deterministic systems, although also see Sec. III. It can be shown, for example, that Bayesian methods are not practically applicable to systems that exhibit deterministic chaos.[19] We will adopt an information criteria because the concept of entropy is often more familiar to physicists, and in practice performs as well as the statistical methods. Certainly the problem we identify in the next paragraph is fundamental and applies to whatever methodology is used to obtain the model structure.

The essential problem we wish to illustrate and investigate is that the *best* model of the time series is not necessarily the *correct* model. To demonstrate this we have examined time series from the Henon system (1) of 1000 and 10 000 observations with Gaussian observational noise at signal to noise ratios of 60, 40, and 20 dB, and multiple realizations of the noise in each case. We obtained the best polynomial model with maximum order $L=3$ and maximum degree $D=3$. To avoid an involved discussion of various model selection methods and introduce difficulties of whether one method might be better than another, we used, for the purposes of this study, an exhaustive search to obtain the best model, that is, for every model structure we obtained a least squares fit of the coefficients and took the best model to be the one that has the minimum description length. [It would have been preferable to use total least squares, but this is prohibitively expensive in an exhaustive search, however, using total least squares does not avoid the problem we are about to describe. For minimum description length we state results for the formula of Judd and Mees (Ref. 20). Other information criteria, such as Schwarz Information Criteria (Ref. 18), Normalized Maximum Likelihood (Ref. 21), give either the same, or very similar, results. Although in this particular example Predictive Description Length did not (Ref. 17).] The best model, for example, using 1000 observations with 40 dB noise, was

$$\begin{aligned} x(t) = 0.786 - 0.419x(t-1) + 0.301x(t-2) \\ - 0.064x(t-3) - 0.766x(t-1)^2 + 0.296x(t-2)^2 \\ - 0.189x(t-1)x(t-3) + 0.885x(t-1)x(t-2)^2. \end{aligned} \quad (2)$$

In all other cases the same model structure was obtained. Furthermore, the coefficients varied by no more than 0.02 from those of (2). There was one exception: when using 10 000 observations with 20 dB noise, and only for some noise realizations, the best model had an additional one or two terms with coefficients no larger than 0.05, and the coefficients of the terms in (2) differing by no more than 0.05.

II. MODEL DEGENERACY

The experiments demonstrate two things: the best model is not the correct model, and the best model is the same for a considerable range of data quantity and quality. Why is the model (2) the best model? Why does it have this particular structure and how is the structure related to the correct model? It should be noted that the structure of this best model (2) is not obviously related to the correct model (1). We claim that the model (2) is one of a large subclass of models that are *equivalent* to the correct model. To see this requires a little formal algebra.

Define $x_i = x(t-i)$, $i \geq 0$, and consider the sequence of multivariate polynomials m_i of the form

$$m_i: x_i - \alpha - \beta x_{i+1}^2 - \gamma x_{i+2}, \quad i \geq 0, \quad (3)$$

for fixed α, β, γ. Note that the difference Eq. (1) is formally equivalent to $m_0=0$, and, in general, $m_i=0$ are equivalent time-shifted copies of Eq. (1). Now consider a multivariate polynomial $p=\Sigma_{j=0}^{k} q_j m_j$, where the q_j are arbitrary polynomials in the variables x_i, $i \geq 0$. Observe that the equation $p=0$ formally defines a difference equation, and this difference equation is dynamically equivalent to (1), in the sense that any time series of (1) (without noise) also satisfies the difference equation defined by $p=0$. In mathematical terminology the set of all polynomials is a *Ring* and the subset of these defined by $\Sigma_{j=0}^{k} q_j m_j$, for arbitrary polynomials q_j, forms an *Ideal*.[22,23] (A Ring is a set of elements on which addition, multiplication, and scalar multiplication is defined. An Ideal is a subset of a Ring that is closed under scalar multiplication and addition, and multiplication by any element of the Ring.) This particular ideal will be referred to as the ideal generated by the polynomials m_i, $i \geq 0$ and will be denoted $\langle m_* \rangle$. The ideal $\langle m_* \rangle$ can be formally identified as the equivalence class of polynomial difference equations equivalent to the difference equation $m_0=0$; the models in this class may be referred to as *degenerate* models. [Mathematically any polynomial in $\langle m_* \rangle$ provides a *perfect* model, which has dynamics identical (conjugate) to the correct model (1).]

It can be easily verified that the formal polynomial of (2) is in the ideal $\langle m_* \rangle$ of (1) with the stated values of α, β, γ, because it has a factorization

$$m_0 - (0.21 + 0.63x_1)m_1, \quad (4)$$

to within a numerical accuracy of 0.01. Factorization over an ideal requires using a Gröbner basis for the ideal.[12,13] It is easily shown that defining the ideal $\langle m_* \rangle$ as we have results in the m_i being a Gröbner basis with the variable ordering $x_0 > x_1 > \ldots > x_n$. It can then be easily shown (by writing the polynomial with symbolic coefficients, performing the multivariate polynomial division by the Gröbner basis, then equating the coefficients of the remainder polynomial to zero), that any difference equation with the structure of (2) is equivalent to (1) if the coefficients have the form

$$\begin{aligned} &1:\ \alpha + \alpha A - \alpha^2 B + \alpha^2 \beta, \quad x_1^2:\ B, \\ &x_1:\ A, \quad x_2^2:\ \beta A - \alpha\beta B + \alpha\beta^2, \\ &x_2:\ \gamma, \quad x_1 x_3:\ \gamma B - \beta\gamma, \\ &x_3:\ \gamma A - \alpha\gamma B + \alpha\beta\gamma, \quad x_1 x_2^2:\ \beta B - \beta^2, \end{aligned}$$

where A and B are free parameters. The corresponding factorization is

$$m_0 + [A - B\alpha + \alpha\beta + (B-\beta)x_1]m_1.$$

One usually thinks of the minimum description length, and other methods mentioned previously, as a modern formulation of Ockham's Razor: the simplest model is best. But we have seen that the application of this principle obtains (2), which is not the simplest model, and even though (2) is correct in the mathematical sense, it is not obvious that there is a simpler model (1).

III. IMPERFECT MODEL SCENARIO

Thus far we have restricted attention to the perfect model scenario, where our model class includes the correct (or perfect) model. The perfect model scenario is, of course, fiction, because even the most well-designed physical experiments have thermal noise or other apparently random perturbations of the system state. One might question whether our assumption of a deterministic model is ever valid, and hence question of whether our results are relevant or useful. A discussion of the appropriateness of deterministic models is beyond us, but we can easily demonstrate that the phenomenon of degeneracy we describe is robust, in the sense that the phenomenon occurs in the imperfect model scenario (when using a deterministic model for a stochastic system), and the phenomenon occurs with stochastic models too.

Consider now the situation where our system has a correct model,

$$x(t) = \alpha + \beta x(t-1)^2 + \gamma x(t-2) + \delta(t), \qquad (5)$$

with parameters as in (1) and the $\delta(t)$ are independent, identically distributed Gaussian random variables with mean zero and standard deviation σ. Suppose as modelers we treat the system as though it were a deterministic system, use the deterministic polynomial model class (which is now imperfect), and apply an exhaustive search for the minimum description length model in this model class as before.

We performed experiments with time series having lengths and observational noise as previously described and σ=0.0006 and 0.006. (For comparison, the 40 dB observation noise has a standard deviation of 0.007 21.) We also performed experiments with zero observational noise. In all cases the best model obtained by an exhaustive search of the minimum description length obtains either the model (1) or a degenerate model of this, usually (2), and occasionally a variant of (2) with the extra terms having much smaller coefficients.

This experiment demonstrates that degeneracy occurs in the imperfect deterministic model class, and the perfect stochastic model class. We conclude that degeneracy is a robust phenomenon and not restricted to deterministic models.

IV. DISCUSSION

We have demonstrated that even for a simple example of the perfect model class using an exhaustive search and current best practices, one does not obtain the correct model from data, but rather a larger equivalent (degenerate) model. We believe this occurs because the best predictive model of noisy data is not the correct model. Observe, for example, that the best model is larger than the correct model, and includes terms that have larger degree and longer lags than the correct model. A slightly simplistic interpretation of this phenomenon is the best model uses the information from further in the past to better estimate the current state.

Global polynomial models have useful algebraic properties that enabled us to reveal the underlying structure and nature of the problem. This is not necessarily of any immediate practical value, because the best model may be in the equivalence class of many other smaller models and it may be impossible to determine which, if any, is the correct (or simpler) model. On the other hand, we observe in the exhaustive search that models in the equivalence class of the correct model were noticeable "local" minima of the description length, and certain model selection algorithms[20,24] also reveal these local minima. This could enable the correct model to be identified as being one of the local minimum for which many of the other local minima are degenerate models.

The problem of model degeneracy is not special to polynomial models and occurs in other models classes, such as radial basis models and neural nets. The difficulty with these classes is they do not have the algebraic structure (rings, ideals, and Gröbner bases) that allow the easy dismantling of degeneracies.

Further investigating these ideas and techniques may reveal new general methods to identify correct models avoiding the problem of model degeneracy in arbitrary model classes and without excessive computation.

[1] M. Cartwright, *How the Laws of Physics Lie* (Oxford University Press, Oxford, 1983).

[2] A. Chalmers, *Science and its Fabrication* (Open University Press, Buckingham, 1990).

[3] M. Henon, "A two-dimensional mapping with a strange attractor," Commun. Math. Phys. **50**, 69 (1976).

[4] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1996).

[5] H. D. I. Abarbanel, T. A. Carroll, L. M. Pecora, J. J. Sidorowich, and L. S. Tsimring, "Predicting physical variables in time-delay embedding," Phys. Rev. E **49**, 1840 (1994).

[6] Edited by G. Gouesbet, S. Meunier-Guttin-Cluzel, and O. Menard, *Chaos and its Reconstruction* (NOVA, New York, 2003).

[7] Peter Grassberger, Thomas Schreiber, and Carsten Schaffrath, "Nonlinear time sequence analysis," Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 521 (1991).

[8] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Number 7 in Cambridge Nonlinear Science Series (Cambridge University Press, Cambridge, 1997).

[9] E. J. Kostelich, "Problems in estimating dynamics from data," Physica D **58**, 138 (1992).

[10] P. McSharry and L. A. Smith, "Better nonlinear models from noisy data: Attractors with maximum likelihood," Phys. Rev. Lett. **83**, 4285 (1999).

[11] A. I. Mees, *Nonlinear Dynamics and Statistics* (Birkhauser, Boston, 2000).

[12] D. A. Smirnov, B. P. Bezruchko, and Y. P. Seleznev, "Choice of dynamical variables for global reconstruction of model equations from time series," Phys. Rev. E **65**, 026205 (2002).

[13] H. Tong, *Non-linear Time Series: a Dynamical Systems Approach* (Oxford University Press, Oxford, 1990).

[14] H. U. Voss, J. Timmer, and J. Kurths, "Nonlinear dynamical system identification from uncertain and indirect measurements," Int. J. Bifurcation Chaos Appl. Sci. Eng. **14**, 1905 (2004).

[15] A. I. Mees, "Parsimonious dynamical reconstruction," Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 669 (1993).

[16] H. Akaike, "A new look at the statistical identification model," IEEE Trans. Autom. Control **19**, 716 (1974).

[17] J. Rissanen, *Stochastic Complexity in Statistical Inquiry* (World Scientific, Singapore, 1989), Vol. 15.

[18] G. Schwarz, "Estimating the dimension of a model," Ann. Stat. **6**, 461 (1978).

[19] K. Judd, "Chaotic-time-series reconstruction by the Bayesian paradigm: Right results by wrong methods," Phys. Rev. E **67**, 026212 (2003).

[20] K. Judd and A. I. Mees, "On selecting models for nonlinear time series," Physica D **82**, 426 (1995).

[21] J. Rissanen, "MDL Denoising," IEEE Trans. Inf. Theory **46**, 2537 (2000).

[22] W. W. Adams and P. Loustaunau, *Introduction to Gröbner Bases*, Number 3 in Graduate Studies in Mathematics (American Mathematical Society, Providence, 1991).

[23] R. Fröberg, *Introduction to Gröbner Bases* (Wiley, New York, 1997).

[24] T. Nakamura, K. Judd, and A. I. Mees, "Refinements to model selection for nonlinear time series," Int. J. Bifurcation Chaos Appl. Sci. Eng. **13**, 1263 (2003).