

Complexity in Time Series Modeling

Kevin Dooley
Arizona State University
Kevin.dooley@asu.edu

Society for Chaos Theory in Psychology and the Life Sciences, 2(3), 1994

Key to any empirical study is the task of modeling. Empirical models can be used to predict, validate "first principle" theory, enumerate phenomena, suggest new theory, and point to limits of knowledge. Good data modeling is both skillful and artistic.

Much effort has been placed of late in developing means by which we can characterize, and in some cases predict, behavior from nonlinear dynamical systems. Numerous approaches exist: calculating the fractal dimension of an attractor, modeling the time series as fractal brownian motion, searching for nonlinear patterns using neural networks, attempting to discover the underlying topology via attractor reconstruction, prediction via piece-wise splining techniques, modeling the dynamics as one of the elementary catastrophes, extending ARIMA time series models to take into account nonlinearities, etc. (Weigend & Gershenfeld; Johnson & Dooley).

The modeler, faced with the burden of learning and mastering yet another set of techniques, is challenged. As computers open the door to more "complex" modeling approaches, our interests follow. A new set of hypotheses must be tested. New code must be generated. New methodologies must be developed. Just what is the relationship between the complexity of the data we model and the techniques we use to model them with?

The most basic (and perhaps crude) representation of data complexity is the algorithmic information content, or AIC. AIC measures the compressibility of a data string. For example, the binary string "110110110110" could be compressed to "repeat '110' four times". Highly orderly strings will have low AIC, whereas completely random strings will have high AIC. While the AIC for most given data strings is uncomputable, general comments can be made (Gell-Mann).

A dynamical system with a point attractor has minimal AIC--a complete description of system behavior can be compressed into a single number. A periodic system following a limit cycle has an AIC defined by the set of points along its deterministic path--not necessarily a short description, but a finite one nevertheless. An aperiodic (chaotic) system has yet a larger AIC, related to the magnitude of the fractal dimension of its strange attractor. As random noise is added to the system behavior, the data becomes less compressible. Data strings

modeled by linear and nonlinear stochastic time series models typically have some order to them, but also large components of randomness, hence high AIC. These data strings contain a mixture of limit cycle (or strange) and random attractors. At maximum AIC lies random noise with a corresponding random attractor; the shortest description of a random sequence is the sequence itself.

Thus, a spectrum of AIC values, from low to high, underlie a corresponding spectrum of time series model-types that might be used for characterization (Note: the following is not meant to be exhaustive, but rather illustrative of typical issues facing the modeler). Consider a univariate series $y(t)$. Here are different types of models that could be fit, depending on AIC:

AIC	model	example
low	linear deterministic line	
	linear differential	exponential, sine/cosine
medium	nonlinear differential	Lorenz weather equations
	nonlinear difference	exponential time series
	linear difference	ARIMA linear time series
high	random noise	Normal distribution

When AIC is low, modeling is likely to occur with the time/case index "t" being an independent factor. For example, one could fit linear [$y=a+bt$], polynomial [$y=a+bt+ct^2$], and transcendental functions [$y=a*\exp(bt)$, $y=a*\sin(bt)$] to the data. Models forms are likely to be continuous. As AIC increases, modeling shifts to a self-referential (autoregressive) mode, where (e.g.) $y(t-k)$ is used to predict $y(t)$. Model forms become discrete. For example, an AR model would have the form [$y(t)=a+b*y(t-1)$]. For very high AIC, only probabilistic modeling is possible, in the form that "y" follows a probability distribution function.

The nature of prediction also changes across AIC. At low AIC, the goal is a point prediction in time. With random noise, only distributional patterns can be predicted. At intermediate AIC values, prediction is mixed--sometimes the goal is a single point prediction, and sometimes the goal is a boundary prediction (e.g., $P(a<x<b)$).

Information theorists and data compressors are interested in AIC because it measures the gap between the *exact* data string and a compressed representation. Modelers, however, do not expect to recreate or compress the data string exactly. What is important is that non-random features of the data string are extracted and characterized. The model (or schema) of the data separates order from disorder.

"Effective complexity" measures the length of the model used to describe the complexity of the data. One cannot compute effective complexity in an exact manner, because model forms differ so vastly. But it would seem fair to propose

that the effective complexity of a model is related to the number of parameters which define it. In that respect, there seems to be little relationship between AIC and effective modeling complexity. For example, a line has two parameters, the logistics equation has one, and a normal distribution has two, etc. Within a particular model domain, effective complexity will increase as higher order models are used to increase model fit. The concept of parsimony is equivalent to striving for minimal effective complexity.

What is more interesting is to compare the amount of "labor" that the modeler must go through in order to develop a model of a given data string--the complexity of the modeling task itself is referred to as crypticity. It has a maximum value at intermediate AIC values.

Consider the three typical tasks the modeler starts out with:

1. Create a time series plot of the data.
2. Create a histogram of the data.
3. Calculate sample statistics of the data, such as mean, standard deviation, range, etc.

If the data appears to have low AIC, a few simple, additional steps take place: a one or two-parameter line is fit, using linear regression. Residuals may be calculated and examined to determine if a polynomial model is needed, or if the raw data needs a transformation. The same situation exists if high AIC data is hypothesized: distributional parameters are estimated (for example, via moment matching), and a probability plot is generated to test for goodness of fit.

These modeling tasks are of Order-1, or O1 complexity. I will define O1, O2, and O3 as three prototypical levels of crypticity (modeling complexity), defined by the following characteristics:

- computation time--how long does it take the computer to estimate parameters, perform statistical tests, generate plots, etc.?
- tool availability--how commonly available are the tools; are the techniques needed available in common main-frame or PC software packages, or must new code be written?
- investigative certainty--as modeling progresses, is the modeling methodology well defined, or is the "decision tree" quite complicated, often calling for backtracking and modeler intuition; do the techniques have parameters themselves which must be adjusted?
- noise sensitivity--how robust are the techniques to random noise?
- theoretical basis--how complex is the underlying theory; how much background knowledge does it take for the modeler to carry through the modeling tasks, make appropriate decisions, and understand the implications of the model?

At low AIC values, fitting a straight line to data takes little computational time; techniques are readily available, even on hand calculators; there is little uncertainty about "what" to do; parameter estimates are wonderfully insensitive to noise (and many other abnormalities); and the theory behind it is simple--part of any introductory data analysis course at a college level, and covered in many cases in high school.

Likewise, at the other end of the AIC spectrum, fitting data to a statistical distribution is of O1 complexity. It takes little computation time to compute standard statistics and draw a histogram; these tools exist in any data processing program--even probability plots are part of popular spreadsheet programs now; there can be some "art" to fitting distributions, (e.g.) if one is allowed to entertain a full range of hypotheses, but there are some good rules of thumb available; probability plots and histograms are good at detecting outliers and multiple modes, and thus are robust to "noise"; and the simpler plotting techniques and statistical calculations are part of any introductory data analysis course.

At low-to-medium AIC, where linear differential and difference equations are fit, modeling is of O2 complexity. Consider as examples the Fourier representation of a signal (a decomposition of the data into a set of sinusoidal functions) and an ARIMA time series model (a set of self-referential difference equations). Computation time (on a PC) is on the order of fractions of minutes, or minutes, as compared to seconds. Fourier and ARIMA methods are available only on "high end" PC packages, or in numerical libraries on mainframe computers. In Fourier modeling, one must be careful of aliasing effects, and if spectrum analysis is used, window parameters must be chosen; ARIMA modeling is notoriously intuitive, and typically many models must be tried before a "best" model is found. Both techniques are robust to noise. The theoretical basis for these techniques is not simple, and is typically not part of a student's study until graduate work.

If we are to buy into the notion that much "interesting" phenomena in social systems and human behavior exist at the "edge of chaos", or in a low dimensional state of chaos, then we are in for a challenge, because intermediate values of AIC require modeling of O3 complexity. Consider the estimation of attractor fractal dimension and the first Lyapunov exponent--necessary first steps to claim chaotic behavior (Peitgen, Jurgens, and Saupe).

Computation time is on the order of minutes and hours depending on the algorithm chosen; some parametric modeling techniques developed for chaotic data can take several hours. Tools are not readily available. A few specialized programs have been written, but the algorithms contained in these commercial programs (e.g. original Grassberger and Procaccia correlation dimension algorithm) tend to be outdated by the time they are brought to market. State-of-

the art algorithms are traded on the Internet, where one must come prepared with some programming skills in order to take full advantage of the offerings.

There are no well-accepted, time-tried methodologies for testing chaos. Investigative paths are still in large part determined by the extent of knowledge of the modeler, and the techniques they have access to. Almost all the algorithms themselves have parameters which must be defined, and only rules of thumb exist for parameter settings (e.g. Peters). Many of the algorithms are incredibly sensitive to noise, even at levels of a noise-to-signal ratio of one percent (Johnson and Dooley). The theoretical basis for these techniques is daunting--one has to be prepared to read "Physica D"-type journals in order to keep up with the state-of-the art. Thorough understanding of deterministic chaos and dynamical systems--a complex task in itself--is required before one can delve into the empirical modeling of such systems.

One immediate danger that is apparent is that researchers in some cases have approached the empirical modeling of complex systems with an assumption that one can get away with an "O1" or "O2" level of effort--e.g. a single method applied a single way, leading to grandiose conclusions, with no sense of testing linear or null hypotheses, or checking for alpha error, or adjusting algorithmic parameters for better fit. It is interesting to note that in almost *all* the publications in this area across the different domains, the investigators discover low dimensional chaos. Could it be that chaos is that prevalent? Are studies which do *not* find chaos being withheld? Or could it be that the algorithms used to test for chaos have very large false alarm rates (Theiler et al.)?

Finally, it is worth noting that modeling complexity (crypticity) is more related to the *hypothesized* AIC rather than the actual. For instance, even if a data series is completely random and thus could be easily modeling with a probability density function, the modeler may expend an O3-effort looking for deterministic chaos in the system.

To summarize, systems which yield data which have a high degree of order (linear determinism) or disorder (randomness) are simple to model. Systems which have a moderate degree of order or disorder can be modeled by linear differential or difference forms. Modeling efforts here are more involved, although expertise in such can still be readily gained. Systems with intermediate values of order and disorder--at the so-called "edge of chaos", or with low dimensional chaos--are the most difficult to model. Researchers should carefully consider the value of such modeling versus the effort extended. Algorithmic and methodological advances must be made before modeling of such systems can become more commonplace.

I wish to thank Honeywell Solid State Electronics Center and 3M Engineering Systems and Technology Laboratory for support of this research.

References:

- Gell-Mann, M. (1994). *The Quark and the Jaguar*. NY: Freeman.
- Johnson, T., and K. Dooley (1994). "Looking for Chaos in Time Series Data," *Proceedings of the Society for Chaos Theory in Psychology and the Life Sciences*, Baltimore.
- Peitgen, H-O., Jurgens, H., and D. Saupe (1992). *Chaos and Fractals: New Frontiers of Science*. NY: Springer-Verlag.
- Peters, E. (1991). *Chaos and Order in the Capital Markets*. NY: John Wiley and Sons.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and J.D. Farmer (1992). "Testing for Nonlinearity in Time Series: the Method of Surrogate Data," *Physica D*, 58, 77-94.
- Weigend, A. S., and N. A. Gershenfeld (Eds.) (1993). *Time Series Prediction: Forecasting the Future and Understanding the Past*. Reading, MA: Addison-Wesley Publishing Co.