Computers & Geosciences Vol. 19, No. 1, pp. 41–52, 1993
Printed in Great Britain
0098-3004/93 $6.00 + 0.00
Pergamon Press Ltd

DIMENSIONALITY ANALYSIS OF TIME-SERIES DATA: NONLINEAR METHODS

A. D. FOWLER and D. E. ROACH
Ottawa Carleton Geoscience Centre and Department of Geology, The University of Ottawa, 770 King Edward Avenue, Ottawa, Ontario, Canada K1N 6N5

(Received 22 November 1991; accepted 17 August 1992)

Abstract—The field of nonlinear dynamics has resulted in the development of several techniques aimed at determining the dimensionality of strange attractors underlying time series outputs from chaotic systems. Knowledge of the system dimension allows estimation of the number of independent variables governing it. The techniques also allow one to distinguish between data sets produced through either random (i.e. high dimension) or deterministic processes (i.e. low dimension). Unlike physics and geophysics where data strings may be large ($n \sim 1000$s) those in geology are generally short ($n \sim 100$s), and there is associated noise. A new algorithm modeled after that of Sugihara and May is presented which in comparison to the correlation function technique works on short and discontinuous data strings.

Key Words: Time-series, Nonlinear, fractal, chaos, deterministic attractor.

INTRODUCTION

Earth scientists have been concerned with the analysis of series data for decades. The data generally are in two forms; temporal and spatial, represented by such diverse phenomena as the records of eruptions, tremors, earthquakes, rhythmic sequences of sediments, mineral zoning, and metamorphic and igneous rock banding. An understanding of the underlying processes controlling these events and sequences is best approached through a combination of theoretical and empirical methods. Time-series analysis is useful for comparing the nature of signals with theoretical models. The ultimate goal is to understand systems to the degree that short-term accurate predictions can be made.

Many classical approaches, such as Fourier analysis, are not particularly useful for characterizing even the simplest of nonlinear systems. This is because nonlinear systems can produce chaotic signals lacking periodicity.

Perhaps the best known method of discriminating between stochastic and deterministic time series is that used by Farmer in his now famous dripping-tap experiment. He showed that construction of plots of differences of successive values in the time domain t (i.e. $t_x - t_{x+1}$ vs $t_{x+1} - t_{x-2}$, where t_x is the time of an event), conserve the characteristic behavior of the system. This demonstrates that one can analyze the temporal aspects of a system (i.e. time between events) in the same manner as used to analyze time series data of a system variable. This is particularly important for geological systems where the system variables cannot be observed and the data consists of a sequence of events.

Figure 1 shows plots of the variable x_n for the recursively generated time series of the logistic map,

$$x_{n+1} = Cx_n(x_n - 1) \quad (1)$$

for a value of C, the control parameter = 3.999, n from 1 to 300, and a seed value of $x_1 = 0.5$. This is compared with data from random numbers; data are presented in Table 1. Qualitative observation shows that both time series seem to be uniformly distributed (Fig. 2). Output from the logistic map for values of $C > 3.57$ are chaotic (e.g. May, 1976) indicating that the data distribution is characterized by what we term deterministic randomness. The plots of Figure 1 show that despite the apparent uniform distribution of the logistic data, and unlike the random data, the time series when plotted as a return map (i.e. $x_{n+2} - x_{n+1}$ vs $x_{n+1} - x_n$) approaches a continuous smooth curve. This simple analysis shows that the data output from the logistic map are deterministic, that is, each successive value depends upon the value of its predecessor. Below, more sophisticated numerical techniques used in nonlinear dynamics are documented and compared. We also present a new and simple method of time-series analysis based upon the technique of Sugihara and May (1990). The algorithms discussed are listed in the Appendix as programs written in Think Pascal. System requirements and typical execution times on a Macintosh IIfx are given as comments at the start of each program.

THE CORRELATION FUNCTION PLOT

The correlation function plot has been used for the analysis of time-series data in physics (e.g. Malraison, Bergé, and Dubois, 1983), meteorology (e.g. Essex,