Welcome to instats

The Session Will Begin Shortly
(At the top of the hour, Eastern USA time)

1

START

2

Nonlinear Time Series Analysis, Part I: Detecting Nonlinearity

Barney Ricca
Lyda Hill Institute for Human Resilience
University of Colorado Colorado Springs

3

Seminar Overview

- Day 1
  - Session 1: Introduction to Nonlinear Time Series (NTLS)
  - Session 2: Behaviors and State Spaces
- Day 2
  - Session 3: State Spaces (continued)
  - Session 4: Recurrences
- Day 3
  - **Session 5: Tests**
  - Session 6: Singular Spectrum Analysis and Noise
- Day 4
  - Session 7: Surrogate Data
  - Session 8: Convergent Cross Mapping

4

Eliminate all other factors, and the one which remains must be the truth.

Arthur Conan Doyle

The Sign of the Four

5

Cautionary Tale

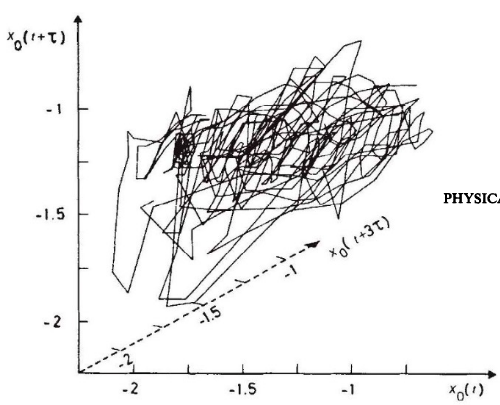

Fig. 3 A view of the climatic attractor embedded in a three-dimensional phase space. The result utilizes about 500 equidistant values of X_0. These are inferred from the oxygen isotope record obtained from the V28-238 deep sea core of Shackleton and Opdyke[4] and extended over the past million years, following an interpolation available from the data bank of the University of Louvain[7]. The value of the shift τ adopted in drawing the figure is $\tau = 2$ kyr.

Do climatic attractors exist?

Peter Grassberger

Physics Department, University of Wuppertal, D-56 Wuppertal, FRG

PHYSICAL REVIEW A VOLUME 34, NUMBER 3 SEPTEMBER 1986

Spurious dimension from correlation algorithms applied to limited time-series data

James Theiler

Department of Physics, California Institute of Technology, Pasadena, California 91125

(Received 7 May 1986)

LETTERS TO NATURE

Dimension of weather and climate attractors

Edward N. Lorenz*

6

What must we eliminate?

- We try to eliminate the following possibilities
  - That our time series data are simply random (noise)
  - That the data may be equally well modeled by something other than a nonlinear system (nonlinearity)
    - The signal must be strong enough to be modeled
  - That the system changes over time in a way that different segments of the time series should be modeled with different models (nonstationary)
  - In the case of vector time-series data, that the various components belong to different, and unconnected, systems (separable)

7

Procedure

- (Linear) Noise
  - Spectrum signal analysis (SSA; Session 6)
  - Brock-Dechert-Scheinkman (BDS) test
  - Surrogate test
  - Various entropy-related tests
  - Maximum characteristic Lyapunov exponent (MLCE)
- Nonlinearity
  - Nash-Sutcliff efficiency (NSE)
- Nonstationarity
  - Nonlinear cross-prediction
- Separability
  - Convergent cross-mapping (CCM) – session 8

8

Brock-Dechert-Scheinkman (BDS) Test

- Is the series signal or noise?
- H_0: The series is independent and identically distributed
- BDS test
  - $BSD(m, \epsilon) = \frac{\sqrt{n}C(m)}{s}$
  - (*n* is the series length, *C(m)* is related to the correlation integral for various embedding dimensions, and *s* is the asymptotic standard deviation)
- Embed the series in *m* dimensions (try several)
- Consider various levels of "closeness"

- Tests.Rmd

9

Surrogate Test

- Is the series signal or noise?
- H_0: The series is independent and identically distributed
- We'll say much more about surrogates in session 7, so for now, we will just note that the one of the surrogate tests is a resampling alternative to the BDS test.

10

Entropy

- Surprise!

$$S = -\sum p_i \log(p_i)$$

- where:
  - p_i is the probability of an event
  - *Log-base* determines the units (*nats* for natural log, *bits* for log-base-2)

Entropy: Alternate interpretation

- Uncertainty and information
- "I'm thinking of a number 1 and 128"
  - Uncertainty (in bits): $\log_2(128) = 7$
    - We need 7 bits to write each of the 1^{st} 100 integers
  - Optimal strategy: Guess the midpoint each time
    - Gains 1 bit of information
  - Because 7 bits is the information we must gain to know the number, the optimal strategy for [1,100] takes at most 7 guesses

- Entropy calculations

Entropy Rate

- Many flavors of entropy; we want *entropy rate*
- Divide state space into boxes ("hypercubes") with edge ε
- Then
  - Deterministic: Entropy rate as ε approaches zero approaches a constant
  - Stochastic: Entropy rate as ε approaches zero diverges

13

Randomness v. (Deterministic) Chaos

- Randomness: No model
- Chaos: Model, but...
  - "Sensitive dependence on initial conditions" (SDIC)
  - Lyapunov exponent > 0
    - Maximum characteristic Lyapunov exponent (MCLE)
- Lack of infinite precision in measurement limits forward prediction/modeling

- In R: Lyapunov.Rmd

14

Nonlinear Serial Dependence

- Reconstruct state space
- Nash-Sutcliffe coefficient of efficiency (NSE)
  - NSE > 0.65 is good evidence of a signal
- *Simplex prediction*
  - Essentially, fancy autoregression
  - Weighted average of the predictions of nearby neighbors
- Aside: Autoregression
  - $x_{n+1} = \alpha_i x_n + \epsilon_i$
  - Not i.i.d., so NOT linear regression
- Tests.Rmd

15

Nonstationarity Tests

- Linear TS
  - “Weak” stationarity: Mean and variance, for example, do not fluctuate
  - Divide TS into segments and compare across segments
  - Doesn’t work for nonlinear
- Nonlinear TS: Cross-prediction stationarity test
  - Divide TS into segments and *predict* across segments
  - Matrix of NSE values
    - Each segment predicts each other segment
    - Entries should be large and roughly constant

16

And around again…

- Tests
  - Noise
  - Nonstationarity
  - Linearity
- Next session: Singular Spectrum Analysis
  - Removes (some) noise
  - Removes (some) nonstationarity
- Repeat tests
  - Why? Stay close to original data and make certain nothing changed

17

Questions

18

STOP

19

Next session @ UTC 1900

20