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

Recurrences

- Recurrence
  - Two nearby points in state space
- Recurrence Plots (RP)
  - 2-d plot of where (in time) recurrences occur
    - Axes the values of the time-series, in temporal order
  - Visual
  - Qualitative
- Recurrence Quantitative Analysis (RQA)
  - Numeric, inferential
  - Quantitative

5

RP & RQA Warning

RECURRENCE PLOTS AND RECURRENCE QUANTIFICATION ANALYSIS ARE COMPUTATIONALLY INTENSIVE

(They can take your computer a long time)

6

An Aside: Recommendations

- Computer
  - Fast processor
  - Lots of processors
  - Lots of RAM
- "Just because something new is available doesn't mean that what you have is suddenly broken." (Weinreich)
  - Corollary (Ricca): Wait as long as you can to upgrade your technology, then get as much as you can possibly afford
    - It'll be obsolete in a few months anyway
- HPC on campus?

7

Recurrence Plot Features

1. Diagonal line segments (parallel to LOI): recurring behavior (number, distribution)
2. Diagonal line segments (perpendicular to LOI): palindrome behavior
3. Vertical or horizontal line: stuck-ness
4. Blocks (with jagged edges): multiple consecutive recurrences
5. Density of points: higher density indicates relatively fewer unique values
6. Mesoscale patterns: units of analyses, slowing/speeding
7. White bands, isolated points: random data
8. And more…

8

Recurrence Quantification Analysis

Quantify what we see in RP

9

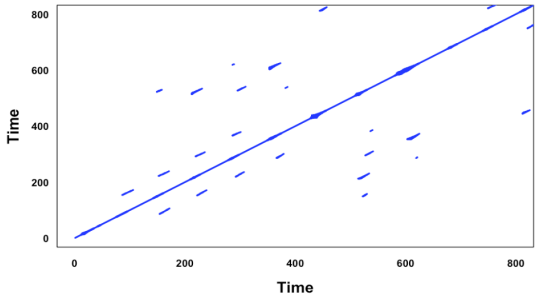
Radius: How close is close?

- Continuous data won't be exactly equal
- Choose how close is close
  - A.k.a., the "radius," a.k.a., "threshold"
  - Trial and error, generally: Choose a value that brings insight

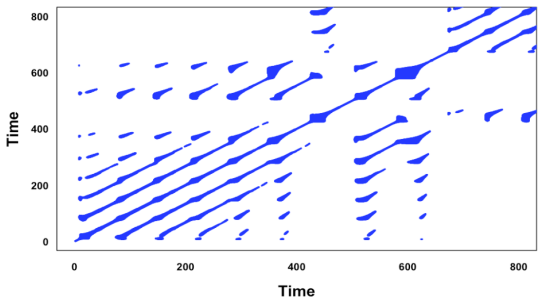
10

Radius example

Radius = 0.01



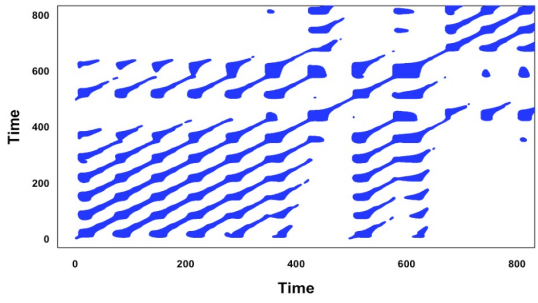
Radius = 0.05



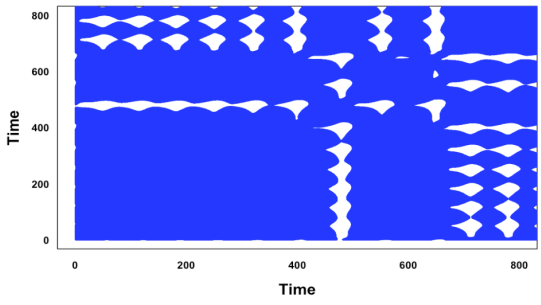
11

Radius example (continued)

Radius = 0.1



Radius = 0.5



12

Aside: Categorical vs. Continuous Data

- Radius
  - Doesn't have meaning: Categorical data aren't "near" each other in any meaningful way; they are identical or not
  - Maybe, if you want to construct RP from spell-check (ordinal "distance")
- You can embed categorical data in multiple dimensions, but why?
  - Embedding depends on false-nearest neighbors
  - Probably don't get anything

13

Theiler Window

- If I were to measure the temperature outside my office window 1000 times per second, we'd get a lot of recurrences.
  - These really aren't recurrences that tell us much about the dynamics
  - Measured much, much faster than the dynamics
- Theiler window
  - How many nearby (in time) points do we exclude because we're measuring faster than we expect the dynamics to change?
    - If you have a theory about characteristic times, use that
    - Otherwise, trial and error. Choose a value that brings insight
  - Thieler window affects the region immediately around the LOI

14

Density of Points

- Recurrence rate (RR)
  - Density of points on the RP.
  - Probability that a state recurs.
- Impacted by choice of radius!
  - Like everything else
  - And the Theiler window (a little bit)
- Convention:
  - Choose radius (and Theiler window) so the RR is near 0.1
  - *Downsample*

15

Diagonal Lines (parallel to LOI)

- Deterministic Structures
  - Repeating patterns of behavior in the system.
- Metric: DET
  - Proportion of recurrent points forming diagonal lines
- Metric: NRLINE
  - Total number of diagonal lines in the RP
- Examples: Choruses in *Let It Be* and *Revolution* RP

16

Diagonal Line Lengths

- Metric: DMAX and <D> for line length
  - Histogram of diagonal line length (excluding the LOI)
  - maxL or DMAX: The longest diagonal line (excluding the LOI)
  - L or <D>: The average diagonal line length (excluding the LOI)

17

Vertical (or Horizontal) Lines

- State does not change
  - Stuck-ness
- Metrics:
  - LAM: proportion of vertical lines (laminarity)
  - TT: Average length of those lines
- Example: The top and right sides of the Revolution RP
- Can be influenced by the radius/Theiler combination
  - E.g., LAM = 0 for sin(x), but large radius with small Theiler window may make it look otherwise

18

Periodic patterns

- Periodically repeating dynamics
  - Recurrences, not repetitions
- Can be sensitive to radius choice

19

Entropy

- A measure of surprise
  - How much information is in each data point (on average)
- Less surprise, less entropy
  - Fewer options to choose from
  - More regular pattern of data
- Equation:

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

- Where p_i is the probability of the i^{th} state.
  - For continuous systems, we estimate by binning the states

20

Entropy

- ENTR (or sometimes ENT)
  - Entropy of diagonal lines longer than the threshold parameter, *mindiagline*
- rENTR (or nENTR)
  - Entropy normalized by the number of lines observed in the RP.
  - Useful for comparing across RP
- catH
  - Entropy based on rectangular block structures

21

Diagonal Lines (perpendicular to LOI)

- Palindromic in time
- No standard metric
- Not something you'll see a lot, probably
  - Strict sinusoidal relationships (or linear combinations of those) will demonstrate this

22

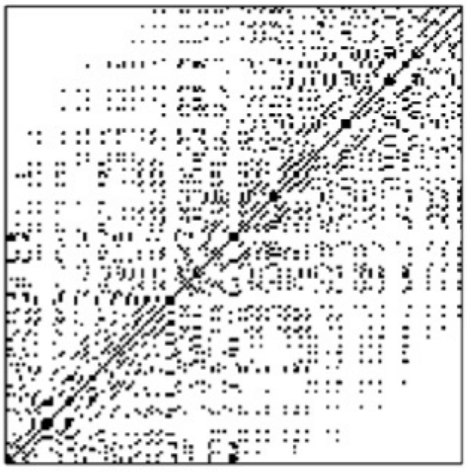
Homogeneity

- Pattern seems the same over large blocks
  - Statistically similar, not exactly the same
    - E.g., all red oak trees are statistically similar
- Indicates stationary dynamics
  - Phase identification!
- Diagonal white bands
  - Transitions between phases
- Vertical White bands
  - Stochastic, random (“fluctuation”) dynamics
  - A relatively few, isolated points

23

Fading and Darkening

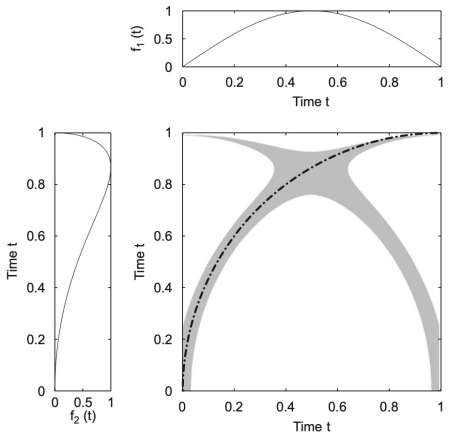
- Drift of dynamics
- Example from Marwan (2007)
  - Look at corners
- Trend (TND)
  - Regress RR of each diagonal (parallel to the LOI) as a function of the time (distance) between these diagonals and the LOI
  - Slope of regression is the trend (TND)
  - Fading away has a negative TND



24

Curves

- Some distortion (e.g., slowing down) of dynamics
- Example from Marwan (2007)
  - Really, this is a CRQA (for "cross")
  - Notice that DTW would remove this feature
    - Which is better for you?



25

…and more

- Marwan & Webber (2015) provide a very nice (and complete) introduction to RQA.
  - Although it is getting older, I think it is still the best place to start.
- Let's do some in R

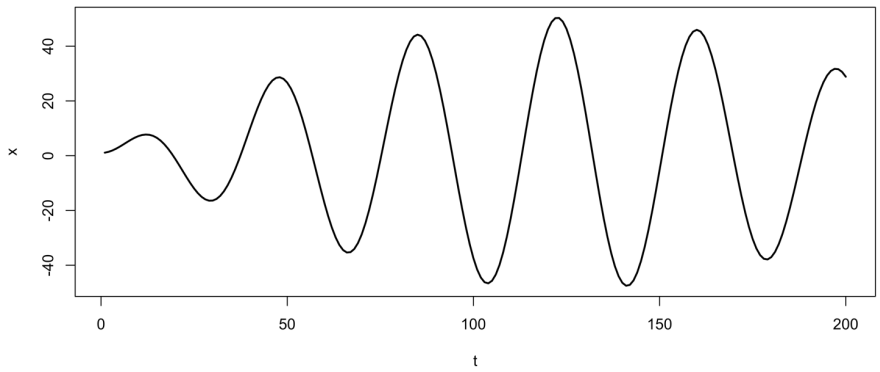
26

Distance Plots

- A.k.a., *global recurrence plot* or *unthresholded recurrence plot*
  - Not really a recurrence plot, but similar in interpretation
- Plot of the distance matrix from each point to each other point
  - Can have really long computational times, depending upon distance and embedding dimension
  - Scale the distances and convert to a color spectrum

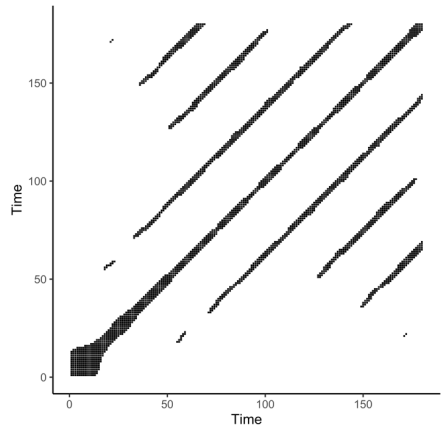
27

Distance Plots: Data



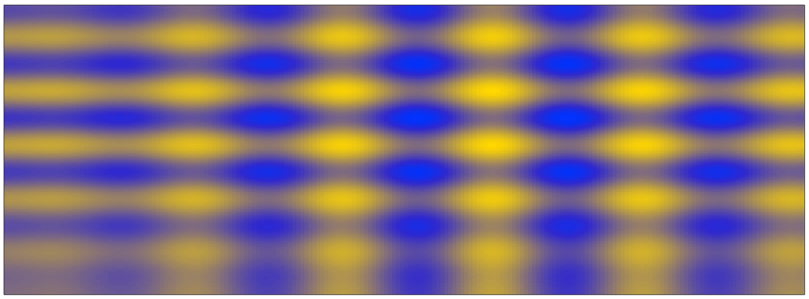
28

Distance Plots: Recurrence Plot



29

Distance Plots: Distance Plot



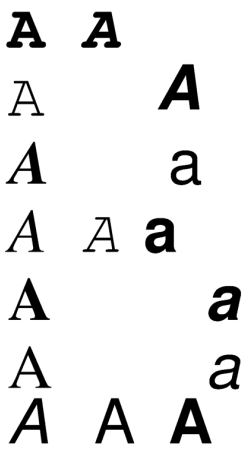
30

Topological Data Analysis (TDA)

- Topology
  - What features are invariant under stretching and bending?
    - No cutting or gluing is allowed
  - Objects, holes, and so on are the invariants
- Topological Data Analysis
  - Identify natural scales in the data
    - *Persistent homology*
  - Heighten the contrast in signals

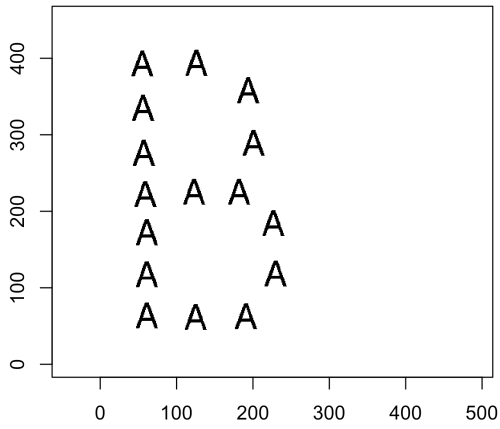
31

The “A-B” of TDA



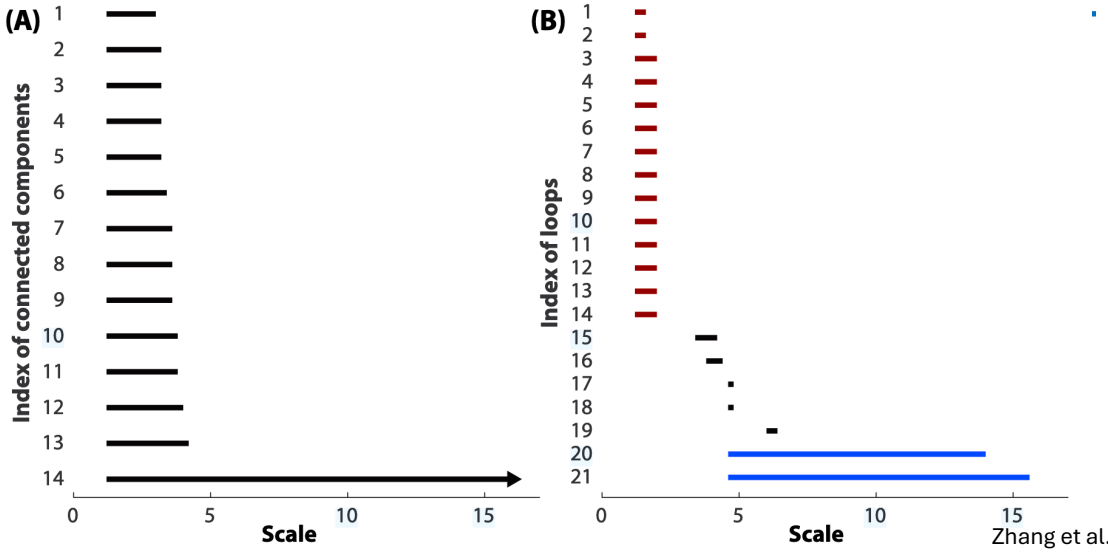
32

The “A-B” of TDA



33

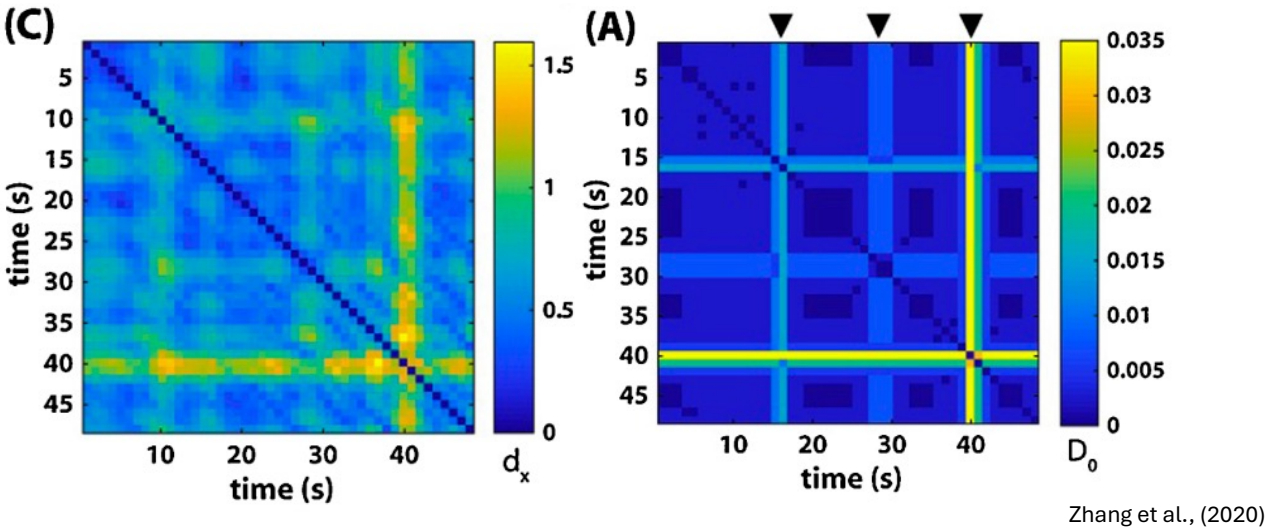
Persistent Homology



Zhang et al., (2020

34

TDA + Distance Plot



Zhang et al., (2020)

35

Inference

- Comparing 2 RP
  - Really, comparing the RQA metrics
  - Estimate standard errors
- Requires a resampling approach
  - Not trivial to resample time series!
  - Session 7

36

Cross Recurrence Quantification Analysis

- Compare two (or more) data streams
  - Must embed in similar ways
  - No longer symmetric about LOI
- Leads, lags, and leapfrogs
  - Chickens and eggs
  - USA Fed: Interest rates and inflation
  - Home: Outside temperature and heater/air conditioner
- Same function in R package:crqa
  - Use two streams, not one
  - Change method from 'rqa' to 'crqa'
- Back to R

37

Questions

Optional homework in "Day 2 Homework.Rmd" (Scripts folder)

Reconvene at 1600 UTC

38

STOP

39

Resume tomorrow

(in most time zones)

Reconvene at 1600 UTC

40