Data Diagnostics: Detecting and Characterizing Deterministic Structure in Time Series Data

University of Bonn

Fall 2022 with a Follow Up class in Winter 22/23

Instructor

Ray Huffaker
Agricultural and Biological Engineering
University of Florida
rhuffaker@ufl.edu

Ray Huffaker works in the ecosystems complexity group in the Department of Agricultural and Biological Engineering at the University of Florida, USA. He specializes in empirical nonlinear dynamics techniques; biological and economic modeling of water and other ecosystem resources; economic and food system dynamics; and natural resource and environmental law. He teaches doctoral courses in nonlinear data diagnostics and biological modeling. He co-authored a book on Nonlinear Time Series Analysis: Huffaker, Bittelli, & Rosa (2017), *Nonlinear Time Series Analysis with R*, Oxford University Press.

Prerequisites

While there are no formal prerequisites, experience working with **R** programming language is helpful.

Course Overview

The scientist Clifford Stoll observed that: "Data is not information, Information is not knowledge, Knowledge is not understanding, Understanding is not wisdom." Time series data provide an essential portal to understanding the systematic behavior of real-world processes. These data may be 'observational' (e.g., collected with direct/remote sensing instruments), 'experimental' (e.g., output of lab experiments), or 'simulated' (i.e., model output). However, 'data is not information' until it has been analyzed for behavioral patterns, and 'information is not knowledge' until detected patterns are explained with theory. The course objective is to cross the portal from data to information—to detect behavioral patterns in data that students can then explain with science from their disciplines. Crossing this portal is challenging because time series data often exhibit an irregular appearance that conceals behavioral patterns from a cursory inspection (Fig 1). The conventional view is that irregularity derives from the stabilizing responses of economic and biophysical processes to exogenous random shocks (linear-stochastic dynamics). However, mathematical breakthroughs demonstrate the surprising result that irregular and apparently random observed behavior can emerge endogenously in nonlinear-deterministic dynamic systems. Several recent papers detect these dynamics in real-world environmental systems. Detecting the source of irregularity in data is pivotal, for example, to understanding how to manage most effectively particular real-world dynamic systems: Can we rely on a system to self-correct in response to outside shocks, or must we take human-in-the-loop corrective actions?

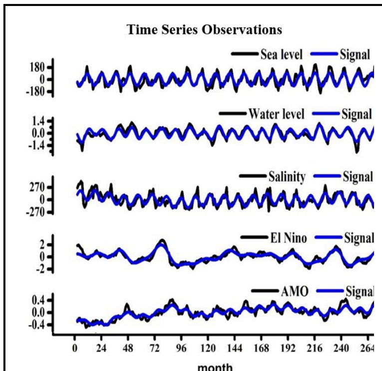

Fig 1

This course introduces Nonlinear Time Series Analysis (NLTS)—a collection of methods designed to reverse-engineer (*reconstruct*) state-space dynamics from sequential data, and thus distinguish between linear-stochastic dynamics or nonlinear-deterministic dynamics as the most likely source of irregular fluctuations in observational data. The empirical power of NLTS lies in its ability to reconstruct real-world system dynamics from even a single variable. This is possible because each variable in an interdependent dynamic system encodes the history of its interactions with other system variables. Although NLTS is not yet widely applied in economics, applied science, and engineering (and thus may well be unfamiliar to most students), it is firmly established in high-impact science, physics, and mathematical statistics journals. **The course puts key concepts of NLTS within the operational reach of non-mathematicians without prior knowledge of nonlinear dynamics.** Consistent with modern trends in university instruction, the course makes students active learners with in-class computer experiments in R code directing them through NLTS methods. The computer code is explained in detail so that students can adjust it for their own work.

Course Format

Students are provided with four written modules. **Module 1: Getting Started** (to be completed before the course begins) guides students through a sequence of tasks that prepares them for subsequent modules covered in class. Modules 2-4 guide students through in-class exercises to gain practical experience with applying a sequence of NLTS methods. Each module introduces a method, provides (and explains) **R** code to run the method, and applies the method to a real-world time series. Each module is accompanied by a homework assignment designed to give students hands-on experience with applying the method/code to analyze other real-world time series.

Scheduled Topics

Module 2: Data Pre-processing. Can we reduce computational time with a shorter dataset that reflects driving dynamics rather than the peak capacity off measurement instrumentation? For example, resampling hourly observations over a year (8675 observations) by taking daily averages decreases the number of observations to 361. We are justified in re-sampling daily only if we are not averaging out substantial change every hour. We use the *Fourier Power Spectrum* to make sure this isn't the case.

Module 3: Signal Processing. We apply *Singular Spectrum Analysis* (SSA) to isolate *signal* (structured variation composed of trend and cyclical components) from *noise* (unstructured variation) in time-series data. SSA can be used to detrend or remove seasonality and noise from time-series data.

Module 4: Screen for System Dynamics. We apply *Nonlinear Time Series Analysis* (*NLTS*) to distinguish between linear-stochastic dynamics or nonlinear-deterministic dynamics as the most likely source of irregular fluctuations in isolated signals.

Follow up in Winter 22/23

Module 5: Surrogate Data Analysis. *Surrogate Data Analysis* statistically tests whether nonlinear-deterministic state space dynamics reconstructed with NLTS were fortuitously generated by a linear stochastic dynamic system.

Module 6: Screening for Nonlinear Stationarity: Stationarity requires that the duration of the measurement is long compared to the time scales of the dynamics. This means that historic records must be of sufficient length to provide an adequate sampling of important oscillatory patterns occurring at lower frequencies[2]. We can test for stationarity in signals with *Space-Time Separation Plots*, *Singular Spectrum Transformation*, or *Nonlinear Cross Prediction*.