

# SYS 7030 Time Series Analysis and Forecasting

Instructor: Arthur Small

University of Virginia Engineering, Fall 2020

*Class meetings:* MW 09:30-10:45 a.m. online via Zoom

*Office Hours:* MW 11:00 a.m.-12:30 p.m. online via Zoom (subject to change). Sign up in advance for a 45-minute session via the Collab “Sign Up” tool. If you cannot make any scheduled time, please contact the instructor via email to schedule an appointment. Meetings online via Zoom: <https://virginia.zoom.us/my/arthursmalliii>

*Web Resources:*

- Collab class site, for basic course information, assignments, office hours sign-up, links to online textbook and other resources.
- Github class site (under development), for posting and sharing code: <https://github.com/uva-engineering-time-series>
- Zoom, for class sessions, recordings, and office hours.

## Course Description

The course is designed to introduce graduate students in engineering to time series analysis and forecasting. The course will not include a deep exploration of theory. Rather, the goal is for students by the end to be able to analyze time series data competently, as part of their work designing and working with engineered systems.

In addition to learning theory, each student will undertake a semester-long research project. Ideally, this project will relate closely to the student’s own dissertation research, professional practice, or other domain application that interests them. My hope is that these projects could form the basis for subsequent research papers, dissertation chapters, or other professional work products, for interested students.

The course will, therefore, be structured primarily as a *workshop*: the ultimate goal is to help you to create a new piece of potentially publishable research. Our workflow will, therefore, be subject to revision, according to my judgement of how best to use our time to help you with your research.

The course outline is divided into two major sections. First, we will introduce the theory, with examples. In the later part of the semester, we will focus on workshoping your projects in progress.

Important: class readings are subject to change, contingent on mitigating circumstances and the

progress we make as a class. Students are encouraged to attend lectures and check the course website for updates.

## Prerequisites

Students should have taken at least one rigorous intermediate course in probability and statistics. They should be comfortable with the representation of uncertain information in the form of probability distributions, with conditional probabilities, and with other such foundational concepts.

In addition, to carry out the data analysis, the student should be able to code, in some general-purpose language. They don't have to learn a new language for the course: any language the student wishes to work in (R, Python, Java, C++, Matlab,...) is probably fine. That said, the instructor works primarily in R, and will tend to use R when discussing code examples.

## Expectations

Each student will make a presentation on their data analysis project. Students will be evaluated based on their performance in these presentations and on their final project, on occasional short quizzes; and on their contributions inside and outside of class towards helping other students.

## Readings

The primary text for the course will be *Forecasting: Principles and Practice* by Rob J. Hyndman and George Athanasopoulos. This resource is available for free online and is linked from the Collab site. You may purchase a print copy if you wish. The text includes some example code in R, and covers several useful R packages related to time series and forecasting.

Additional readings including relevant articles will be provided as the course progresses. The choice of readings will depend in part on student interests, as conveyed through their choice of projects. Spatial and spatio-temporal analysis may be covered in class if interest is sufficient.

## Course Objectives

1. Students will learn the foundations of time series analysis and forecasting.
2. Students will gain the experience of building statistical models of time series, and models for forecasting, and will learn how to evaluate their performance.
3. Students will learn the concepts and practice of *reproducible research*, in the course of preparing a research paper.
4. Students will gain experience in making presentations and in preparing a polished research article.

## Grading Policy

- 10% of your grade will be determined by quizzes designed to test your understanding of the theoretical concepts introduced in class. This quiz will be delivered at roughly the mid-point of the semester. It will be open-book and open-notes, outside of class. You will have multiple days to complete it. The quiz will not be designed to be especially challenging: the goal is to give you the opportunity to synthesize your understanding of core concepts, in preparation for developing your data analysis for your research project.
- 10% of your grade will be determined by your performance in one in-class presentation based on your project. These presentations will be scheduled when you are, in the judgement of the instructor, far enough along to do so.
- 10% of your grade will be determined by your contributions to assist other students. These contributions can come through class participation, by making useful contributions in online forums (Github), or through other means that add value to the group experience.
- 70% of your grade will be determined by your performance on your final project. The development of the project will include multiple iterations, each with an associated deliverable:
  - An initial Concept Note.
  - A more developed Project Proposal.
  - A first working draft of your final project paper.
  - A final complete draft of your project paper.

Details of these staged intermediate deliverables will be forthcoming. The final product should be a polished professional paper that meets academic standards regarding format, quality, and integrity.

## Attendance Policy

*Showing up is 80 percent of life* – Woody Allen, via Marshall Brickman

Regular attendance is very much in your pedagogic interest. However, it is up to you whether to attend in person or to view recorded class sessions afterwards.

## Communications protocols, including emails and office hours

I prefer to avoid using email to communicate with students about class matters. For substantive questions about course materials and concepts, please use class time, office hours, or meetings by appointment. Please use email only for brief clarifying questions, or to set up appointments.

## Academic Dishonesty Policy

Don't cheat. Don't plagiarize. Don't present someone else's work as your own.

## **Disabilities Policy**

Together with the University of Virginia, I am committed to assuring that all students have the full opportunity to benefit from the course regardless of their disability status. If you have a disability that may require accommodations, please see the instructor early in the semester to work out appropriate arrangements.

# Course Outline

## Course Introduction

### Theory

#### Introduction

A simple example.

The general form of the problems we'll take up: There is a decision-maker who must choose one option from a menu of options. The decision-maker has objectives. These objectives can be quantified in the form of an objective function (equivalently, a loss function). The payoffs to different choices depend on the values of certain state variables. The decision-maker is in general uncertain about the values of the relevant state variables. When feasible, the information the decision-maker has can be expressed in the form of probability distributions over the relevant state variables. Typically, options are available to apply analytic methods to data to reduce these uncertainties. Many, many problems correspond to this general form. More examples:

Note, though, that the way the problem is presented may hide that it shares the above structure. The outline of topics for this course is structured to provide insight into these concepts, and guidance for the construction of usable computational decision models

Readings: Berger 1.1-1.3, Hoff Ch. 1

#### Formalism for statistical decision models

The payoff to a decision depends on the value of some unobserved parameter  $\theta$ , which could be a vector. We have some system for generating a probability distribution  $\pi(\theta)$  over  $\theta$ . In Bayesian approaches, the probability distribution  $\pi(\theta)$  used is the posterior distribution of  $\theta$  given the observed data:  $\pi(\theta|y)$ . Other approaches exist for creating probability distributions over states.

#### Objective functions, loss functions

- Loss functions that depend only on the magnitude of estimation error: MSE, MAE
- Asymmetric loss functions

#### Relationships between frequentist statistical estimation procedures and loss functions

Hypothesis testing; p-testing; tests of statistical significance. Statistical significance is not the same as practical significance. There is nothing sacred about the 95% confidence threshold. What matters for decision-making: the decision-maker's degree of reasoned belief in the values of relevant state variables. Generally, it does not matter whether or not the variable is statistically significant, or whether a given hypothesis test is passed or failed. These test results are irrelevant to the decision. Strictly irrelevant.

## Predictive Modeling: Estimation, Forecasting, Nowcasting

**Introduction: The Big Picture.** “Many roads to the same destination”: Many analytic techniques are available. The common goal: probability distributions over relevant state variables. Typically this means: conditional probabilities, given the How to choose a predictive modeling approach. Which predictive modeling approach you use will depend on many things: the specifics of the prediction problem, the data you have or can gain access to; your own skills; how much time, budget, expert assistance you have available; whether the problem is important enough to bother getting fancy; whether someone else has already created a good-enough model that you can just use, or use with light modification. How much value does your predictive model add?

**Machine learning:** Classification. Confusion matrices: Converting to conditional probabilities. Q: Should classification algorithms be designed to be conservative or “risk averse”, allowing more false positives in order to avoid false negatives when the consequences of a false negative are severe? Ans: No! Separate the analysis from the decision. Job of the analysis is to provide the best possible information, not to prejudge how the information will be used. (Will return to this question when we take up scoring rules.)

## Estimation, Forecasting, Nowcasting

Examples. The goal for us: probability distributions over relevant state variables. Some predictive modeling tools may produce scores that are not easy to interpret in terms of probabilities. These must be handled with specific care.

**A tour of forecasting and estimation models and methods** Linear regression models. Time series forecasting models. Various machine learning models. Models that incorporate structural knowledge.

## Forecast Evaluation

**Forecasting Evaluation I: Deterministic Forecasts** Example: US AID Famine forecast US Terrorism Alerting System: “The threat level has been raised to ‘Orange’ ”.

**Forecasting Evaluation II: Probabilistic Forecasts** Example: IRI climate prediction system Observation: Scoring rules Strictly proper scoring rules

## Other topics

Multi-period forecasting.

## **Topics**

### **Time series decomposition**

- See if inflow data is non-stationary (e.g. trending from climate change/exhibiting non-constant mean/variance with climate indices like PDO)

### **Detecting regime changes**

- See if there are any change points due to land use/policy change
  
- Try to build a predictive model of releases at different reservoirs as a function of storage and past flow conditions • Build a multivariate simulation model of reservoir inflows • Build a multivariate time series model of the “natural” inflows (i.e. flows that would be there in the absence of human intervention)

### **Handling problems with the data**

#### **Missing values, data gaps**

#### **Censored data**

### **Student Projects**

In the later part of the semester, we will focus on interactive work with students on projects.