Time Series Analysis for Energy Data

ENV 790.30

Nicholas School of the Environment - Duke University

Spring 2021

Instructor

Luana Medeiros Marangon Lima

Visiting Assistant Professor of Energy and Environmental Sciences and Policy

luana.marangon.lima@duke.edu

Office hours: TBD

Teaching Assistant

Zoe Yang

xiaoyueyan.yang@duke.edu

Office hours: TBD

Lectures: On-line asynchronous (pre-recorded classes will be posted on Thursdays)

Lab Section: Wed 1:45 - 3:00pm

Course Description and Objectives

Time series and forecasting methods continue to improve due to the enhancements in computing power and capability of dealing with larger data sets. This course will focus on time series analysis, modeling and forecasting, with emphasis on energy and environment applications.

Throughout the course we will use real data sets from the US Energy Information Administration (EIA), National Oceanic and Atmospheric Administration (NOAA) and the National Renewable Energy Laboratory (NREL). This course will use R for most statistical analysis. Lectures will feature R syntax and/or demonstrations using the R Studio user interface. Note that R and R Studio work on Windows, Linux, and Mac operating systems.

Energy Analytics usually involve getting data, parsing the data and transforming the data to a state where you can actually apply time series analysis. This work is better done in Python, therefore the course will also cover a short introduction on Python.

Upon completion of the course, the students will be able to use R to carry out basic statistical modeling and analysis and fit a model to data. Our goal is to enable students to learn from data in order to gain useful predictions and insights.

Course Platforms and Communications

We will use **R** and **RStudio** to develop our codes. Recorded classes, additional resources and announcements will be posted on **Sakai**. We will use a **Slack** workspace for communication. That way you are just a text message away from instructors and TAs. We will use **Github** to share the lessons, scripts developed in lectures and lab sections as well as Assignments.

Course Prerequisites

Applied Data Analysis for Environmental Sciences (ENVIRON 710) is a prerequisite. The students are expected to know the concepts of probability, random variables, probability distributions, hypothesis test and basic regression analysis. Prior R programming is preferable but not mandatory.

Course Format and Grading

The course consists of lectures at which we will discuss theory and applications. We will learn the time series concepts through data analysis projects. During the classes we will also dedicate some time to learn the statistical packages in R related to the topic as well as small group problem solving. Aside from the in class problems, there will be a set of four assignments and a final project. Grades will be based on:

- 7 homework assignments (70%);
- 1 final project (30%).

Homework Assignments

The assignments involve applying concepts and tools learned in class to an specific data set or problem. Students might work together and help each other. However, the assignments are to be submitted individually. The table below shows possible due dates for the assignments.

	Post Date	Due Date
Assignment 1	1/21/21	1/27/21
Assignment 2	1/28/21	2/10/21
Assignment 3	2/11/21	2/24/21
Assignment 4	2/25/21	3/8/21
Assignment 5	3/11/21	3/24/21
Assignment 6	4/25/21	4/7/21
Assignment 7	4/8/21	4/21/21

Final Project
The final project could take several forms. If you have an interesting dataset, you may choose to work with it using existing methods and software tools to run your time series analysis. Another idea is to take some previously published data and analysis and use it as a starting point. You could simply take the data and do your own analysis. Or you may reproduce part of the published analysis, but in this case you will need to go further and try different models and analysis with the data. Make sure you clearly state the difference between what you have done and what was done previously. Students are encouraged to work in teams of two or three for a project.
There will be two short presentations of your final project. For the first you will present the data set you will use, what you plan to do with it and the project motivation. For the second presentation you will show the class the main results obtained throughout the analysis. Aside from the presentations, you are required to submit a final report as if you were writing a research paper. Describe the data sets, tools used and results. If the data set has been used before show what else you have done with it and compare with previous published results.
The final project grading will be weighted as follows:

- Proposal Presentation (20%);
- Final Presentation (50%);
- Report - paper style (30%).

Class Topics

Here is a list of topics on time series that have been widely applied and are relevant for the energy & environment field. We may or may not be able to cover all of them depending on the class flow and the remote learning structure.

1 Introduction to Time Series Data Analysis

2 Trend and Seasonality

3 Autocovariance and Autocorrelation Functions

4 Box & Jenkins Models - Autoregressive (AR) and Moving Average (MA) Processes
  - ARMA and ARIMA Processes
  - Seasonal ARIMA Models
  - Periodic ARMA Models

5 Time Series Modeling and Forecasting
  - Fitting Models to Data
  - Model Identification and Parameter Estimation
  - Model Diagnostics and Model Selection

6 State-Space Models

7 Introduction to Bayesian Statistics

8 Dynamic Linear Models (DLM)

9 Scenario Generation (SG)

Class Proposed Schedule

The schedule below is subject to change. I may modify it throughout the course if extra time is needed for some particular topics. I will provide updates in class and via Sakai.

Week	Module/Lessons	Recorded Videos Post Date	Lab Sec. (W)
1	Class Overview, Introductions, Intro to Time Series Analysis	-	20-Jan
2	Intro to R and RStudio, Github, RStudio basics, R Markdown	21-Jan	27-Jan
3	Autocovariance and Autocorrelation function, Trend and Seasonality	28-Jan	3-Feb

Week	Module/Lessons	Recorded Videos Post Date	Lab Sec. (W)
4	Stochastic vs Deterministic Trend, Stationarity Tests: Mann Kendall, Spearman, Augmented Dickey Fuller	4-Feb	10-Feb
5	Outlier Detection, Missing data	11-Feb	17-Feb
6	Intro to the Traditional Box & Jenkins Models, Stationary Models: AR and MA process	18-Feb	24-Feb
7	ARIMA Models in R	25-Feb	3-Mar
-	Spring Break	-	**10-Mar**
8	Seasonal ARIMA and Periodic ARMA Models	11-Mar	17-Mar
9	Model Diagnostics, Residual Analysis and Model Selection, Model Performance Metrics	18-Mar	24-Mar
10	Intro to Forecasting, Forecasting with ARIMA Models, Forecasting in R	25-Mar	31-Mar
11	State-Space Models, Bayesian Statistics	1-Apr	7-Apr
12	Dynamic Linear Models - Application to Inflow Forecasting	8-Apr	14-Apr
13	Scenario Generation for Uncertainty Modeling, SG Application to Hydro-thermal Scheduling	15-Apr	21-Apr

Class Etiquette

You should take responsibility for your education. I expect students to attend every class and get to class on time. If you must enter the class late, please do so quietly. Retain from using phones and tablets for social media during class. Some classes will involve coding on your laptop. I expect you to focus on the assignment and refrain from any web browsing that may disrupt the progress of your work.

Your classmates deserve your respect and support. We will likely have students from many different backgrounds and countries in this class and you should all feel comfortable and make each other comfortable while participating.

Nicholas School Honor Code

All activities of Nicholas School students, including those in this course, are governed by the Duke Community Standard, which states:

"Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and nonacademic endeavors, and to protect and promote a culture of integrity. To uphold the Duke Community Standard:

- *I will not lie, cheat, or steal in my academic endeavors;*
- *I will conduct myself honorably in all my endeavors; and*
- *I will act if the Standard is compromised."*

Please add the following affirmation to the end of all assignments, and sign your name beside it: "I have adhered to the Duke Community Standard in completing this assignment."

Land Acknowledgment

"What is now Durham was originally the territory of several Native nations, including Tutelo (TOO-tee-lo) and Saponi (suh-POE-nee) - speaking peoples. Many of their communities were displaced or killed through war, disease, and colonial expansion. Today, the Triangle is surrounded by contemporary Native nations, the descendants of Tutelo, Saponi, and other Indigenous peoples who survived early colonization. These nations include the Haliwa-Saponi (HALL-i-wa suh-POE-nee), Sappony (suh-POE-nee), and Occaneechi (oh-kuh-NEE-chee) Band of Saponi. North Carolina's Research Triangle is also home to a thriving urban Native American community who represent Native nations from across the United States. Together, these Indigenous nations and communities contribute to North Carolina's ranking as the state with the largest Native American population east of Oklahoma."