ENVIRON 790.30 Time Series Analysis for Energy Data Spring 2023

Course Overview

Class Hours

T-Th 12:00 to 1:15 LSRC A247
To join class on Zoom click here.
Passcode: F2022

Instructor

Luana Medeiros Marangon Lima
Office: Gross Hall - 102K
E-mail: luana.marangon.lima@duke.edu
Office hours: Thursdays 10:00-11:00am (Gross Hall 102K or Luana's Zoom), or by appointment.

Teaching Assistant

Yu Hai
email: yu.hai@duke.edu
Office hour: TBD

Communication

We will use Slack for communication. I will add all students to the slack workspace I've created for the class. Using slack will assure I never miss an email from you and will also keep us one text message away! You may use slack on your computer and/or phone.

Click here to join our slack workspace.

Course Description

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 R, therefore the course will also cover a short introduction to 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. The goal of teh course is to enable students to learn from data in order to gain useful predictions and insights.

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 assignments, a forecasting competition and a final project. Grades will be based on:

	Percentage
Assignments - A1 to A7	70%
A8 - Forecasting Competition	10%
Final Project	20%

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.

Policy on late submissions: Assignments are due at 11:59pm. Assignments submitted at least 2 hours after the deadline will have 1 point out of 100 deduction by hour. After that, there will be a 5 points out of 100 deduction per day.

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:

	Percentage
Proposal Presentation	20%
Final Presentation	40%
Report	40%

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.”

Course Modules

The class topics are divided into twelve modules. There will be readings and/or recording associated with each module.

M1 - Getting started, Intro to TSA, R and RStudio
M2 - Autocovariance and autocorrelation
M3 - Trend and seasonality
M4 - Missing data and outliers
M5 - ARIMA Models
M6 - Seasonal ARIMA Models
M7 - Intro to forecasting
M8 - Model Performance
M9 - State space models
M10 - Advanced forecasting models
M11 - Model based scenario generation

Class Proposed Schedule

The proposed schedule below is subject to change. My initial plan is to cover all the material listed here but I might modify it if extra time is needed for some particular topics. I will update this table as needed during the semester.

Lecture	Module	Date	Topic	Homework
L1	M1	Jan 12	Introductions and Course Overview	Join Slack Workspace
L2	M1	Jan 17	Intro to Time Series Analysis, Intro to R and RStudio, Github, R Markdown	A01
L3	M2	Jan 19	Autocovariance and autocorrelation function	
L4	M2	Jan 24	Partial autocorrelation function ACF and PACF in R	
L5	M2 M3	Jan 26	ACF, PACF and plots in R Trend Component Estimation	A02
L6	M3	Jan 31	Seasonal Component - Stochastic vs Deterministic Trend	
L7	M3	Feb 2	Trend and Seasonal component estimation in R	A03
L8	M3	Feb 7	Stationarity Tests: Mann Kendall, Spearman, Augmented Dickey Fuller	A04
L9	M4	Feb 9	Outlier types, detection, how to handle missing data	A05

Lecture	Module	Date	Topic	Homework
L10	M4	Feb 14	A3 Solution Finish outliers in R	
L11	M5	Feb 16	Intro to the Traditional Box & Jenkins Models - ARIMA family Stationary Models: AR and MA process	
L12	M5	Feb 21	A4 Solution AR and MA order (poll) ARIMA(p,d,q) Models	
L13	M5	Feb 23	ARIMA(p,d,q) Fitting ARIMA Models in R	A06
L14	M6	Feb 28	A5 Solution Seasonal ARIMA and Periodic ARMA Models	A07
L15	M6 / M7	Mar 2	Finish SARIMA in R Intro to Forecasting Averaging Techniques	
L16	M7	Mar 7	Forecasting with ARIMA Models Forecasting in R	
L17	M8	Mar 9 (remote assynchronous)	Watch recodings for M8 Work on Project, Team building	Project Proposal (2-3 slides)
-		Mar 14	**Spring break no class**	
-		Mar 16	**Spring break no class**	
L18	M8	Mar 21	Review Model Diagnostics Review Residual Analysis and Model Selection Model Performance in R	A08
L18	M9	Mar 23	Model Performance in R State-Space Models Bayesian Statistics	A08
L20	M9	Mar 28	State Space Models in R Go over Forecasting Competition Forecasting higher frequency time series	A09 - part I
L21	M10	Mar 30	Advanced Forecasting Models in R	A09 - part II
L22	M11	Apr 4	TBATS models in R Scenario Generation	Work on project/competition

Lecture	Module	Date	Topic	Homework
L23	M11	Apr 6	Scenario Generation in R Course Recap Course Evaluation	Work on project/competition
L24	-	Apr 11	**Final Project Presentations** Presentation Schedule	Work on project
-	-	Apr 13	**MEM Symposium - no class**	Work on project
L25	-	Apr 18	**Final Project Presentations** Presentation Schedule	Submit project and competition knitted files

Assigments Schedule

You should use Sakai to submit your work. Assignments should be submitted using the *Assignments* tab. This is a tentative schedule.

Assigment	Due Date
A01 - Course Set Up Sakai link GitHub link	Jan 24
A02 - ACF & PACF Sakai link GitHub link	Feb 3
A03 - Trend and Seasonality Sakai link GitHub link	Feb 10
A04 - Stationarity Tests Sakai link GitHub link	Feb 20
A05 - Decomposition Sakai link GitHub link	Feb 27
A06 - AR and MA models Sakai link GitHub link	Mar 6
A07 - ARIMA models Sakai link GitHub link	Mar 20

Assigment	Due Date
A08 - Model performance Sakai link GitHub link	Mar 27
A09 - Forecasting competition Sakai link GitHub link Kaggle link	Enter competition and create Github repo by Apr3rd (Part 1) First model by Apr 7th (Part 2) Final submission + knitted file by Apr 28 (Part 3)
Final Project Presentation & Short report	Apr 28