METR 5433 - Advanced Statistical Meteorology
Spring 2020 – Homework #3
Due: 12 March 2020

(#1) The file *NormanMinT.csv* contains daily minimum temperatures (in F) for Norman, Oklahoma from July 1, 2002 until February 29, 2020. The goal of this exercise is to make you think about how and what techniques to use that could make sense for ultimately producing a predictive model for minimum temperatures in Norman. I am leaving this question fairly open-ended. That means that you make the choices on how to proceed and justify those choices to me. The one caveat is that you may use **ONLY** the data in the file provided for this exercise to complete the task.

Think about concepts we discussed in class, including autoregressive modeling, covariance modeling, trends, and least-squares fitting. In designing your model, think about any relevant statistics (e.g., AIC, BIC, (auto-)correlation, etc.) to back up choices you make. Produce the necessary plots to show your logic. Some guidelines and tips:

- Make sure to take a look at your data before jumping into the modeling. Are there trends you need to consider? What should you do about that?
- You might find some interesting (conflicting?) results between different statistics, so don't just rely on one statistic!
- Temperatures have a long-term mean and a seasonal cycle. Leaving the seasonal cycle is probably OK (your predictive model should capture this, but if you want to justify removing the seasonal cycle, then do so). However, remember that AR models require centered data (i.e., the overall mean removed).
- You may or may not consider adding "white noise" to your predictive model when doing the reconstruction and predictions.
- For building the predictive model, split the dataset into a "training" and a "testing" dataset. That is, you will use your training dataset to build the statistics and the actual statistical model you think works best. Then, you will use data from the test dataset to "test" how well your model performs for 7-day forecasts of minimum temperatures. You might consider including relevant error statistics for sample forecasts.
- For this exercise, use 2002-2018 as your "training" dataset.
- For the forecasting element of the exercise (i.e., "testing"), I want you to make four 7-day forecasts. The start dates will be (a) April 10, 2019, (b) July 10, 2019, (c) October 10, 2019, and (d) January 10, 2020. For example, starting on January 10, 2020, you will use your model to forecast the minimum temperatures for each day from January 11-17, 2020. Remember, these forecasts are for actual temperatures, not anomalies or standardized anomalies.

In the writeup for this problem, include (at least) the following:

- A succinct description (i.e., the equation/definition of terms) of your statistical model (e.g., if your model has 100 terms, please don't print out all of those terms!)

- The reasoning behind how you chose your model, including relevant statistics and/or plots.
- A plot of the original data and the "reconstructed" minimum temperatures from your model overlaid on top from 2002-2018. Include the r^2 value of your model on the plot.
- A comparison of your predicted values with the actual values for those seven days, as given in the file. Include any plots and relevant statistics to illustrate your model's performance and error for each of the forecast periods.
- A discussion of the potential sources of error in your model and whether or not your model could be generalized to work for any period of time (e.g., all seasons, El Niño years, "extreme" periods (warm or cold), etc.).

Again, this exercise is for me to see your thought process and creativity in building this model. Explain your steps and reasoning and back them up with relevant statistics and plots. Even if your model doesn't work great for forecasting, that's OK, but try to tell me why that is the case.

(#2) It's time to start thinking about your class project. Remember, the aim of the class project is to put your data analysis skills into practice . The project needs to be based on research question(s) that you would like to answer, **not** on the type of data analysis techniques you would like to (or "have to") apply. Your project must also involve more than one of the data analysis techniques we have learned / will learn in class. Since we still have concepts to cover, it's OK if you don't know the exact techniques yet, but an idea of them would be good. You might consider talking with your advisor(s) or other people for ideas and datasets to use. One restriction is that this project cannot be something you have already done (for another class or for your own research).

Write a 1 page maximum, single-spaced proposal of your project. Include a short background, the research problem, hypothesis / research question(s), and potential techniques you want to use (you can look ahead on the syllabus to get ideas of upcoming techniques) and/or steps you plan to take for completing your project. By reading this document, I should know what you want to do, why it is important, and your plans for moving forward with the project. We may need to chat one-on-one about some more details of the project, but that is OK.