STAT GU4221/GR5221 Project
Forecasting Spatial-Temporal Climate Data

1 Introduction

Forecasting windspeed and solar radiation has become increasingly important in fields such as engineering and finance. In the Wind-Energy Marketplace, the forecasted windspeed becomes highly valued information, especially to the governing companies who own the windmills. In short, companies must project how much energy output a potential buyer requires before setting up an energy deal. This could then lead to the windspeed futures being above or below market expectations.

2 Data Description

The data set of interest (`WindSpeed_Month_Ave.csv`) consists of monthly spatial-temporal observations for windspeed over the span: January, 1979 to December, 2018. The training data is constructed from averaging daily measurements, i.e., one case is based on roughly 30 days. Averaging the cases helps to smooth the data and reduces the computational burden of fitting the full set of observations. The resulting dataset consists of $n = 480$ cases or 480 months over the timespan.

Consider defining each case $X_t(l_1, l_2), t = 1, 2, \ldots, 480$, as a function of latitude (l_1) and longitude (l_2). The spatial domain is observed primarily over Texas, New Mexico and Oklahoma. As displayed in Figure 1, the full domain spills over to Mexico and several surrounding states. Figure 1 also shows case 469, which includes both the wind speed and the spatial domain in a 3-dimensional scatterplot.

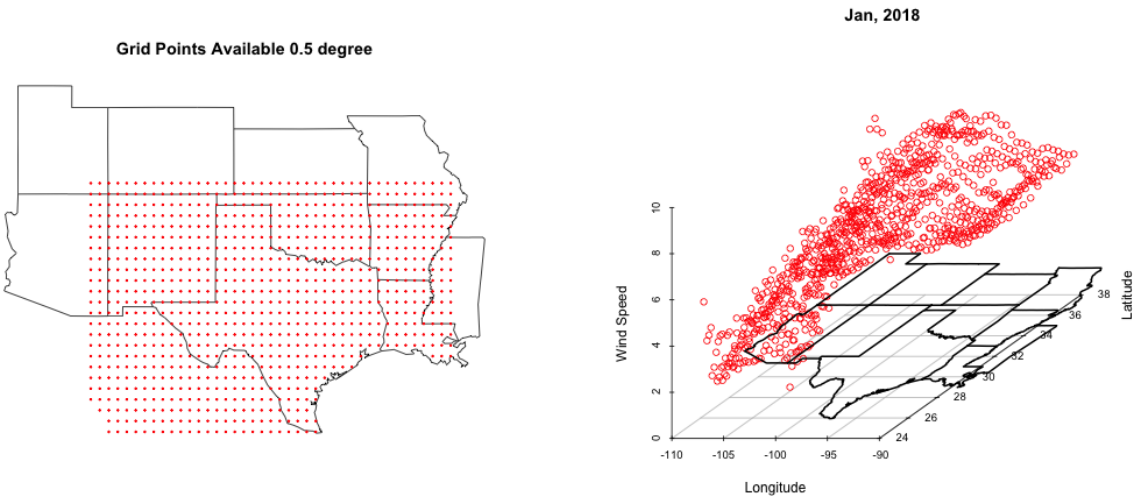

Figure 1: The left panel displays the spatial domain of the observed data set. The right panel is a 3-dimensional scatterplot of the 469th case, which represents the average wind speed for the month January, 2018. The wind speed is a function of both spatial coordinates (longitude, latitude).

To visualize the temporal aspect of the dataset, Figure 2 shows a sequence of three graphics over time. The three cases displayed are 469,470 and 471, which respectively correspond to months January - 2018, February -2018, and March-2018. To visualize the full dataset, imagine 480 scatterplots plots placed in a sequence similar to Figure 2.

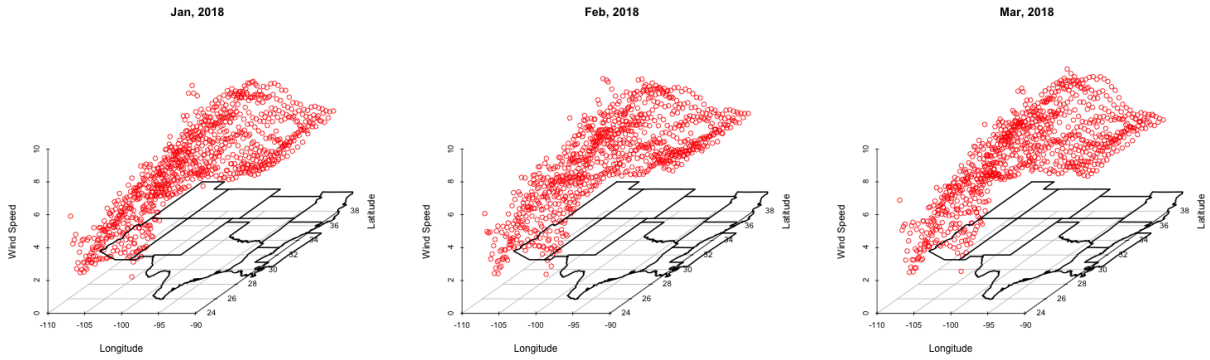

Figure 2: The above three graphics can be interpreted as a time series plot for cases $i = 469, 470, 471$, which correspond to the respective months: January, February, March-2018. In contrast to traditional multivariate time series, the above graphic preserves the spatial domain.

3 Project Objective

The goal of this project is to forecast windspeed one-, two-, three-, four-, five-, and six-months ahead using techniques from Time Series Analysis and related fields. Students can use standard models such as ARMA, ARIMA, GARCH, Linear Regression, .. etc. Or students can branch out and utilize more sophisticated models used to forecast multivariate time-series and/or spatial-temporal data. If students decide to use a more sophisticated model, please also choose a baseline model for comparison.

4 Considerations on Your Chosen Model(s)

As mentioned earlier, you can consider defining each case $X_t(l_1, l_2), t = 1, 2, \ldots, 480$, as a function of latitude (l_1) and longitude (l_2). The structure of this dataset is far more complex than the univariate setting introduced in GU4221/GR5221. Therefore students are allowed to make some major assumptions on their chosen time series model. For example, you can assume that the windspeed at each windmill cite (l_1, l_2) is independent, i.e., you can assume that each site j has it's own independent time series model $\{X_t^{(j)}\}$. This independence assumption is obviously false because we would expect for spacial observations to be correlated with each-other. The choice of your model is subjective but some will perform better than others. Make sure to be transparent on your chosen model.

5 Project Write-up

Students are required to type a final report. The structure of the report is subjective. Feel free to adjust your final report as you see fit.

I. **Introduction:** Include a brief description of the goals of your project coupled with some exploratory data analysis.

II. **Statistical Model(s):** Clearly describe your chosen model(s) and any major assumptions made on the data generating process. This section should also include how you trained your model of interest and any baseline models for comparison.

III. **Results:** Summarize your results, including model validation and forecasted results.

IV. **Appendix**

   a. **Model Selection?**

   b. **Model Validation?**

   c. **Selected Code?**

   d. **Other?**

6 R Code or Python Code

- Students should prepare an organized `R script` (or `Rmd`) file that complements the written report (or `Python` file)
- **Do not** copy and paste the code into your appendix. Only include very important code, or no code, in your final report. Please upload the script or similar on Canvas by the due date.

7 Grading

The Project is Due on May 5th, by 11:59pm.

- This project will be graded on:
   i. Completeness (don't forget to turn in your `R` file also)
   ii. Correctness
   iii. Organization/neatness
   iv. Creativity
- I want to see a nice organized final report. It must be typed with graphs labeled. Please do not make the report too long! No more than around 10 pages.

8 Other Project Types

If you have an idea for another project, please reach out to Prof. Young for approval. If you do not gain approval for another project, please stick with the windspeed data.