Collaborators: Peter Kinder, Byanka Estudillo, Michael Ghattas
Domain expert name: Serena DiLeonardo
Domain expert affiliation: Atmospheric and Oceanic Science (ATOC)
Domain expert contact info: Serena.Dileonardo@colorado.edu
Date of initial meeting: 03/03/23
Date(s) of additional meetings: N/A (Using Discord as main channel of communication)
Date of report: 03/15/23

Q1: Qualitative
Q1 is the project's initial Qualitative component, which sets the foundation for the Quantitative component of the project (Q2) and the implementation of the solution (Q3). Specifically, there are seven aspects of Q1 relevant for every project.

A: What is the domain problem? Be sure to write down their overall research, business, or policy goals and their specific scientific questions.

To find how downslope windstorms have changed from 2002 to 2022, and assess any trends in those changes. Domain expert has developed a code to identify how many windstorms are happening monthly, seasonally, and yearly.

B: Why is this problem important or interesting? This should be answered individually by each collaborator on the project. Why is your domain expert's research, business, or policy question interesting to you? If it's not interesting, make up a plausible reason.

Byanka: I think it's important to see how our environment is changing over time, to then make projections on what to expect in the future.

Peter: One important aspect that stood out to me was how knowledge about windstorms could potentially affect planning for wildfire mitigation efforts.

Michael: Investigating changes in the pattern and behavior of wind provides empirical key points as indicators that could be utilized in researching environmental subjects around pressure, humidity, temperature, and other topics, such as economics to agriculture.

C: How will the eventual solution be used? How they will use the answers to their research questions (i.e., what is their intended outcome of the research) and how will this help achieve the overall goal of the project?

Windstorms exacerbate wildfires, and it would be helpful to know how windstorms are changing, whether they are increasing or decreasing, or if there is expected variability to be better prepared for these potential situations.

D: What potential data could solve the domain problem? What data, if it were available and accessible, would help answer the underlying research questions or guide the business or policy decisions? This is an important hypothetical exercise.

Accounting for the potential shifting seasonal variability or trends.

E: The actual data (only if data have already been collected)

E1: What data have been collected?

Monthly wind data i.e. wind speeds, wind direction, humidity and temperature.

E2: Why were the data collected originally? (For what purpose?)

To develop predictive modeling and simulation capabilities to benefit the wind energy industry.

E3 and E4: When and where were the data collected?

Data has been collected since 2001 at NREL's campus.

E5: Who or what collected the data?

Data from National Renewable Energy Lab.

E6: How were the data collected? With what instrumentation/methods?

Data on wind speeds and direction is collected by an 82 meter high meteorological tower situated around 8 miles south of Boulder, right at the foothills of the mountains where wind speeds are usually the fastest.

F. What may be the qualitative relationships between variables, for those observed and unobserved?

There could be a relationship between the frequency, duration, and intensity of windstorms and certain seasons of the year (i.e. more windstorms during the winter months).

G: Which types of statistical analyses or techniques would be most useful to the domain expert? Which would not be useful?

Recurrent neural networks on sequences of data could be applied to the time series data if time permitted. Autoregression (moving averages) could also be a good fit for the model that would take seasonality into account. A linear model wouldn't be appropriate for the count data as it would indicate negative amounts of windstorms which would never happen.

Q2: Quantitative

Summarize the statistical collaborators' quantitative contribution or advice. Did the domain expert understand the statistics? This can be whatever (if any) quantitative contribution or

advice you provided during the initial meeting or in a subsequent follow-up email. If there has not been any Q2 advice so far, indicate your thoughts of potential directions for Q2.

The DE expressed that they tried using time-series analysis methods to identify changes in wind pattern behavior and feel they have been successful for the most part. However, Serena needed our assistance verifying her techniques by completing our own time-series analysis focusing on seasonality and trends in the data she cleaned and prepped. Since Serena is versed in mathematics and very familiar with most statistical concepts, it was easy to communicate our strategy, utilizing techniques such as Lag and Partial Autocorrelation functions to investigate the seasonality and trends in the data.

Q3: Qualitative
Did the contribution, advice, or solution answer the researchers' questions? Will it help the domain expert achieve his or her overall research goal(s)? Are there any practical constraints limiting the effectiveness of the proposed Q2 statistical solution? What is the answer to the research question(s)? Note: It is uncommon for an initial meeting and follow-up activities to result in Q3 conclusions or recommendations. If these have already occurred, please detail them here. If they have not occurred, just state that Q3 has not yet occurred.
Q3 has not occurred yet.

Yes. The collaborators' contribution validated the DE's findings and further elaborated on the nature of the wind trends and seasonality. The added information included identifying a SARIMA model that explains the wind trend and behavior. In addition to the recommendations for further analysis, the study also reported several necessary statistical measures for the ARIMA model. The model's log-likelihood was 158.57, indicating that the model fits the data well. The Akaike Information Criterion (AIC), corrected AIC (AICc), and Bayesian Information Criterion (BIC) values were also reported, with values of -311.13, -311.03, and -300.7, respectively. These values suggest that the ARIMA(0,1,0)(1,0,1)[12] model is a good fit for the data, as it has the lowest AIC and AICc values among all the candidate models considered. The relatively low value of the estimated sigma-squared parameter (0.01455) also indicates that the variance of the trend component of the series is relatively low, suggesting that the series is relatively stable and predictable. However, the study notes that caution should be exercised when interpreting this finding, as the ARIMA model assumes that the underlying data-generating process is stationary, which may not be the case for the original time series. Furthermore, the study also discussed the diagnostic plots for the ARIMA model. These plots showed that the residuals were approximately normally distributed and had constant variance over time, suggesting that the model fits the data well. The forecast from the ARIMA model predicted that the number of windstorms would continue to decrease slightly over the next 12 months. However, the study notes that there is still considerable variability in the data, and therefore, it is essential to continue monitoring the data and update the analysis as more data becomes available. Overall, the study provides a comprehensive analysis of the time series of the number of windstorms per month from 2001 to 2019. The study highlights the presence of a seasonal pattern in the data, a slight downward trend in the number of storms over time, and

a good fit of the ARIMA(0,1,0)(1,0,1)[12] model for the trend component of the series. However, the study also provides recommendations for further analysis to incorporate other relevant variables, test for seasonality, explore other time series models, extend the period, and investigate the impact of extreme events on the occurrence of windstorms.

Shared Understanding Statement:
After completing the previous report sections, you must send the report to the domain expert(s) for any edits or additions and obtain their concurrence on the report. After your team has reached an agreed upon version with the domain expert, you can proceed to submission.

This has been reviewed by the Domain Expert (Yes/No)
The Domain Expert made edits or additions (Yes/No)
The Domain Expert agrees that shared understanding has been created (Yes/No)

Meeting Notes:

LISA Meeting Notes

Attending: Serena DiLeonardo, Michael Ghattas, Byanka Estudillo, Peter Kinder
Date: 03/03/23
Location: Zoom
Video Recording: Yes
Time: 3:00PM - 4:00PM

Wants/Goals for the Meeting:

1. Review types of questions in the survey. Do the types of questions used elicit useful responses?
   Yes! The questions were well-received and instigated a deeper understanding of the project.
2. Discuss best methods for analyzing the results of each type of question (multiple choice, rank order, etc.).
   Discussed Linear Regression, ARMA, ARIMA, and RNNs.

Items Discussed/Decisions Made:

1. Complete a descriptive and diagnostic analysis on the data, and if possible consider building a predictive model that could be used for prescriptive analytics in the future..
2. Assess any changes in windstorms behavior
3. Testing for trends and volatility within seasonality and trend
4. Wind intensity based on velocity and duration are important metrics of the analysis

To do:

1. DE will provide the data via email as soon as possible
2. LISA collaboration team will complete an Exploratory Data Analysis (EDA) and revert back with findings while proceeding with the needed steps for the Statistical Data Analysis (SDA)

Timeline:

1. What are the near-term deadlines? *Nothing set yet, though we should aim to complete the EDA by Monday, March 13th.*
2. Any longer term deadlines? *Project completion* by *Monday, April 3rd*

Next Meeting: All meetings and correspondence will be via email, Discord, and in person during STAT 5680/4680 lecture.

This has been reviewed by the Domain Expert (**Yes**/No) [Initials: SD]

The Domain Expert made edits or additions (Yes/**No**) [Initials: SD]

The Domain Expert agrees that shared understanding has been created (**Yes**/No) [Initials: SD]