**Project 3 Proposal**

Group members: Steph Abegg (I ran it by James, and he was okay with me going solo with no grade cut)

Track: Data Visualization

**The research question**

I work for [LongPath Technologies](https://www.longpathtech.com/), a Boulder-based company that has created revolutionary laser-based technology to monitor methane gas emissions. We monitor emissions at hundreds of sites across the United States. The technology works by measuring the methane concentration on either side of the site, and using the difference in measurements along with wind speed and wind direction data to compute the methane being emitted by the site. So site-specific accurate wind measurements are vital to accurate emission readings. LongPath has a 3D anemometer installed at each site to collect wind data.

Three-dimensional (3D) anemometers offer more comprehensive measurements than 2D anemometers by capturing wind speed and direction in three dimensions, providing a complete understanding of wind flow. This makes them ideal for complex environments and scientific research that require detailed wind analysis. However, they are typically more expensive and may be more susceptible to environmental factors such as precipitation and icing up. In contrast, 2D anemometers measure only the horizontal components of wind speed and direction, making them suitable for applications where vertical wind measurements are less critical or where there is a fair amount of precipitation.

A question is whether 2D anemometer data such as temperature, wind direction, and wind speed can be used as a substitute for the equivalent measurements from a 3D anemometer. This would be particularly beneficial during the times when a 3D anemometer is iced up. To address this question, I analyze data collected from a 2D anemometer and a 3D anemometer at the same location during the same 30-day time frame. The data from both anemometers is recorded on 5-second intervals. The data for each anemometer is first averaged over 15-minute windows (this smooths out the data as well as corresponds to how the wind data is used in practice), and then the 15-minute averaged temperatures, wind directions, and wind speeds are directly compared via time series plots, regression analysis, and binning.

**The data**

This study uses two datasets, both from anemometers located in North Dakota at 47.8437 N, 102.8524 W, elevation 2300 ft above sea level. The data from both anemometers spans 30 days from February 11, 2024 to March 11, 2024. One dataset is from is from a 2D anemometer and the other from a 3D anemometer. The anemometers measure on five-second intervals. There are two two-day gaps in the data, corresponding to when one or both of the anemometers was iced up: February 22 and 23 (both anemometers iced up) and March 3 and 4 (3D anemometer iced up). So there are a total of 26 days of data to compare. The raw 2D anemometer data contains 472,048 rows (pared down to 404,984 rows after cleaning the data) and the raw 3D anemometer data contains 420,917 rows (pared down to 420,911 rows after cleaning the data).

The relevant columns include:

* Date and time in UTC;
* Number of internal data points used to compute the measurements corresponding to a single time;
* Temperature in degrees Celsius;
* Wind direction in degrees (North: 0°, East: 90°);
* Wind speed in meters per second;
* Wind elevation in degrees (3D anemometer only).

In practice the anemometer data is averaged only 15-minute intervals. After doing so and removing rows where either the 2D or 3D data was not represented, there were 2344 rows of data.

Here are links to the data:

* [Link to 2D anemometer data](https://www.dropbox.com/scl/fi/g67lue6yzxxr5e13wjhmz/2d_anemometer.csv?rlkey=2v3isntsbl744gy09f7dayepu&dl=0).
* [Link to 3D anemometer data](https://www.dropbox.com/scl/fi/eryx8777pyz1fo945ufsh/3d_anemometer.csv?rlkey=n8pu010cpx6dhgm32qr1mpp59&dl=0).
* [Link to 15-minute averaged merged data](https://www.dropbox.com/scl/fi/uq6w39h5tk0fczfhj8kob/df_2d_and_3d.csv?rlkey=canv9uh7n8y55k7qkbsjmao99&dl=0).

**The elements of the dashboard**

The elements of the dashboard will be:

(1a) A radio button list with three options: Wind Speed, Wind Direction, Temperature. This allows different data to be to be selected and viewed. All of the plots and metadata update when a new option is selected. Wind speed is the default when the page first loads. (1b) A radio button filter with two options: All wind speeds, or wind speeds greater than 1m/s (2.2 mph). All of the plots and metadata update when a new option is selected. All wind speeds is the default when the page first loads.

(2) A time series that shows the 2D and 3D anemometer data over time, as well as the difference between the measurements corresponding to the same 15-minute window. The user can zoom into the plot using the Plotly zoom feature.

(3) A histogram for the differences between the measurements. The histogram also shows the average and median of the differences over the 30-day span of data. The binning is in set increments for each variable, so that integers always fall between two bins.

(4) A scatterplot of 2D vs. 3D data. A regression line and R^2 value is shown on the plot as well. The points are colored by wind speed. The colorbar is fixed for all plots, even when low wind speeds are filtered out.

(5) Buttons that when clicked show popups with an image of the 2D or 3D anemometer with a brief description. The style of the buttons was set using css style file.

Here is a sketch of the planned layout of the dashboard described above.

Timeseries

Histogram

Scatterplot &

Regression

Title/Description

Button

Select metric (temperature, wind speed, wind direction) and Wind Speed filter (All wind speeds, or wind speeds > 1 m/s).

Buttons for popups describing anemometers.

**Instructions on how to use and interact with the dashboard**

The dashboard is a single page showing three graphs: timeseries, histogram of differences, and scatterplot with regression line. Use the radio buttons on the upper right of the page to display either wind speed, wind direction, or temperature data. Radio buttons can also be used to filter the data to points with wind speeds > 1 m/s. Click the buttons to read more about 2D and 3D anemometers.

**Visualizations by Python**

The dashboard will create the visualizations using Plotly and javascript. But as a future check of these visualizations, I used Matplotlib in Python to create examples of the desired visualizations.

Timeseries:

A graph of different colored lines

Description automatically generated

Histograms:

A graph of a wind direction difference

Description automatically generated

Scatterplots with regression lines:

A diagram of a graph

Description automatically generatedA graph of a graph with numbers and a red line

Description automatically generated with medium confidenceA diagram of a graph

Description automatically generated

**Link to GitHub Repository**

Here is the GitHub Repository where I will be storing my project work:

<https://github.com/sabegg2/Project3.git>