

Creating a Maritime Weather Forecast Model Using NOAA Buoy Climate Measurements

Austin Paczkowski

January 3, 2025

1 Introduction

Oceans have played an integral part in human history for thousands of years. From using it as a method of transportation, transporting goods, and collecting seafood, humans depend on the ocean for survival. At the same time, maritime activities have caused countless fatalities due to the rather unpredictable and extreme weather conditions involved. As such, this human-ocean relationship creates a necessity to predict future weather phenomena, birthing the field of meteorology. By applying statistical methods and regression models to meteorological theory, it is possible to forecast weather conditions such as air temperature, sky clarity, precipitation, air pressure, wave height, and countless other factors to increase confidence in maritime navigation, thereby reducing fatalities. In this paper, we will transform a collection of ocean weather measurements, create a linear model-based weather forecast using this data, perform simulations with this weather forecast, and discuss avenues for further exploration as well as ideas for improvement.

2 Transforming the Data

2.1 Introducing Buoy Data

To begin, we will introduce our data *marine-sample*, a collection of $n = 55$ weather measurements of buoy D5GN6 from the National Oceanic and Atmospheric Administration (NOAA) between January 12th and January 31st 2015[2]. Of the 55 observations listed, each one contains 33 variables with various types of data. Each observation stores ID, time, and coordinates of the observation. In addition, each observation stores categorical and numerical variables providing information about ice, air pressure, temperatures, wave measurements, swell measurements, cloud coverage, visibility, precipitation, and wind. While the data may be dense with information, it is messy and primitive. As such, we need to transform the data into a useful and intuitive form for analysis and modeling.

2.2 Cleaning the Data

Upon viewing the *marine-sample* data, each variable is recorded in one of two ways: as numerical values for objective measurements and as categorical letters or integers. Fortunately, there is a documentation file included with the data at NOAA [2] that explains the meaning of the values in each column as well as the corresponding WMO Table Code [4] to use when necessary. As such, we will use these to interpret, understand, and transform the data for analysis and modeling.

Removing Useless Columns

Nine of the 33 variables contain NA values for each observation. These include Ice Accretion, Thickness of Ice Accretion, Rate of Ice Accretion, Wet Bulb Temperature, Sea Temperature, Wave Direction, Low Cloud Type, Cloud Height Indicator, and Visibility Indicator. In addition, both Identification and Pressure Tendency have zero variability in their observations (all of them are the same). As such, we will remove these 11 columns, leaving us with 22 variables.

Latitude and Longitude

The Latitude and Longitude columns contain coordinates from (0,0) where negative values are South and West respectively. As such, these values are stored as desired and require no transformation.

Timestamp

The timestamps of the observations are stored as YYYY-MM-DDThh:mm:ss. Due to the sparsity and time frame of collected observations, only hour and day values are of any use for analysis. As such, we parse the timestamp into an hour column with the numeric hh value, a day column with the numeric DD value, and remove the timestamp column from our data.

Air Pressure

There are two columns involved with air pressure: Sea Level Pressure and Characteristics of Pressure Tendency. We will transform Sea Level Pressure into millibars from inches of Mercury by multiplying each value by 33.8639. Upon viewing WMO Table 0200 [4], there are 9 different classifications for the pressure. As such, we will transform these 9 classes into two major groups: Increasing Air Pressure and Decreasing Air Pressure.

Temperature

There are two columns involved with temperature: Air Temperature and Dew Point Temperature. For both Air and Dew Point temperatures, the values are initially in Fahrenheit. As such, we will subtract by 32 and then multiply by $\frac{5}{9}$ to get the Celsius value of each observation.

Wave Measurements

There are two columns involved with wave measurements: Wave Period and Wave Height. The Wave Period measurement is a categorical variable corresponding to a range based on WMO Table D5a [4]. For simplicity, we will record each observation as the mean value of these ranges and omit observations without a valid range. Wave Height is recorded as how many half meters high the waves are. As such, we will multiply these values by 0.5 to get the full meter values.

Swell Measurements

There are three columns involved with swell measurements: Swell Direction, Period, and Wave Height. Swell Direction is a numerical measurement of how many degrees clockwise north the swell is traveling divided by 10. As such, we will multiply each value by 10 to get the full degree values. There is also 1 outlier of 380 degrees that will be removed. The Swell Period is recorded as is in seconds and requires no transformation. Swell Height is also recorded as how many half meters high the swells are. As such, we once again multiply by 0.5 to get the full meter values.

Cloud Measurements and Classifications

There are six columns involved in classifying and measuring clouds: Total Cloud Amount, Low Cloud Amount, Cloud Height, Visibility, Middle Cloud Type, and High Cloud Type. Viewing WMO Table 2700 [4] for Total Cloud Amount and Low Cloud Amount, the values correspond to how many oktas the sky is covered in clouds. Because an okta is $\frac{1}{8}$ of the sky, we multiply each value by 0.125 to get a proportion for sky coverage for these variables. For Cloud Height, we assign each class to the corresponding mean value in the range given by WMO Table 1600 [4] to get numerical representations. We omit class 9 and A observations as they are not numerically representable. For Visibility, there are 10 classifications from WMO Table 4377 [4]. However, only classes 97 and 98 appear in our data. As such, we will transform these into their corresponding kilometer values 10 and 20 respectively. For the Middle Cloud Type, there are 11 classifications from WMO Table 0515 [4]. As such, we will simplify this to 3 classes by omitting class A and blank observations, assigning class 0 observations to **No Middle Clouds**, classes 1 and 2 observations to **Altostratus** or **Nimbostratus**, and classes 3, 6, 7, and 8 to **Alto cumulus**. For the High Cloud Type, there are also 11 classifications from the sister table of WMO Table 0509 [4]. As such, we will again simplify this to 3 classes by omitting class A and blank observations, assigning class 0 observations to **No High Clouds**, classes 1 and 4 observations to **Cirrus**, and classes 5 and 8 to **Cirrostratus**.

Weather

There are two columns involved in classifying weather: Present Weather and Past Weather. Present Weather uses WMO Table 4677 [4], involving 100 unique weather classifications. However, only a handful of them are used in our observations. As such, we will simplify this to 3 classes where we assign classes 1, 2, 3, and 13 to **No Precipitation**, class 80 to **Light Precipitation**, and classes 62, 82, 90, and 92 to **Moderate to Heavy Precipitation**. In addition, we will create an additional column called Present Weather Numeric* and assign values 0.01, 0.51, and 1.01 based on the 3 Present Weather classes respectively to numerically represent rain probability. This is important as we will use this numeric in creating our forecast later. Past Weather uses WMO Table 4561 [4], involving 10 unique weather classifications. As such, we will simplify this to 3 classes where we assign class 0 to **Cloud Cover 0.5 or Less**, class 1 to **Mixed Cloud Cover**, and classes 2 and 9 to **Cloud Cover More Than 0.5**.

Wind

There are two columns involved with the wind: Wind Direction and Wind Speed. Wind direction is a numerical measurement of how many degrees clockwise north the wind is traveling from and requires no transformation. Wind speed is recorded in knots. Because 1 knot is $\frac{1}{10}$ of a meter per second, we will multiply by 10 to get the full meter per second values.

Resulting Data

With these transformations applied, our measurement variables are uniformly in the metric system and our categorical variables are heavily simplified from their table counterparts. As such, we will create models using the cleaned data.

(* : NOTE: Present Weather Numeric is initialized in the Data Analysis file, not the Data Cleaning File)

3 Creating a Forecast Model

3.1 Removing Outliers

Before we create models, we need to clean and remove significant outliers in the main regressors of our data (excluding the Precipitation factor). As such, we have our main regressors scatterplot matrix below:

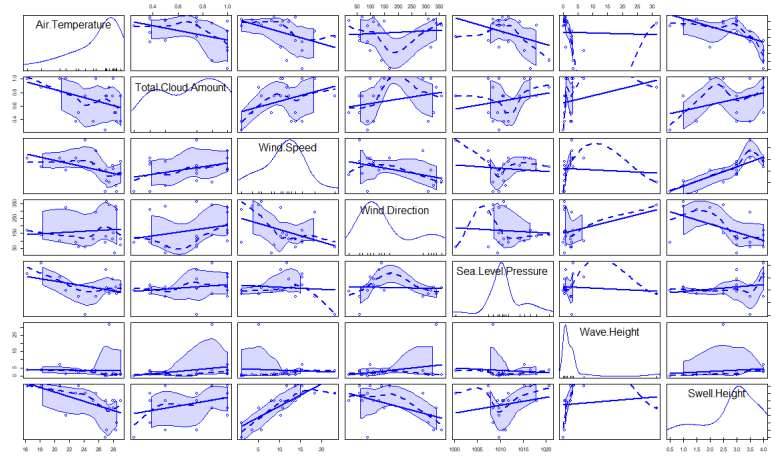


Figure 1: Scatterplot Matrix of Main Regressors

As such, there are two major outliers: one in Wave Height at an observed value of ~ 31.5 and one in Wind Speed at an observed value of ~ 46.2 . Removing these values gives the resulting scatterplot matrix:

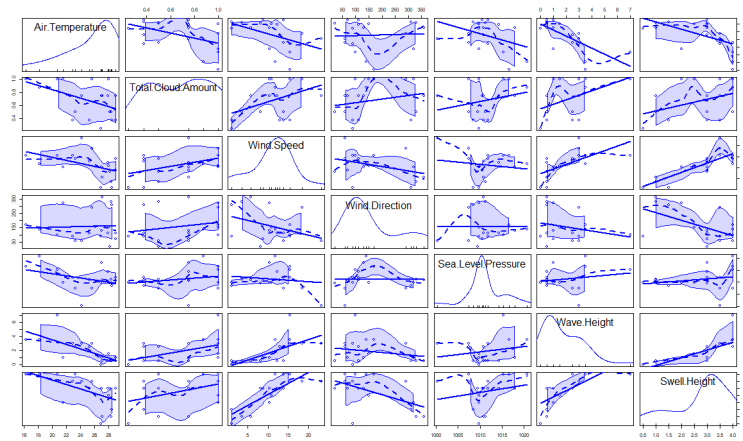


Figure 2: Scatterplot Matrix of Main Regressors (2 Outliers Removed)

3.2 Fitting Linear Models

After removing these outliers, we create linear models to predict the 7 main regressors from the scatterplot matrix as well as precipitation likelihood. As such, we will generate an n-day forecast using the following algorithmic approach:

1. Create a linear model for each of the 8 main regressors using all non-omitted non-NA observations.
2. Use the most recent observation (initially January 31st) as well as the corresponding daily Latitude and Longitude values as our initial input to the mean function of the regressor models.
3. Generate the next day values for each of the 8 main regressors using the most recent day observation values as inputs to the mean functions. Store these values in the succeeding column.
4. Repeat steps 2 and 3 n-times to get an n-day forecast prediction.

3.2.1 Reasoning Behind The Models

For each regressor, we will introduce the model fitted in R, run an overview of the factors contributing to regressor choices in each model [1], and explain the reasoning behind regressor and response transformations using simplified logic from more advanced atmospheric modeling methods [3].

Sea Level Pressure

```
> summary(pressureModel)

Call:
lm(formula = Sea.Level.Pressure ~ Air.Temperature + wind.Speed +
    I(wind.Speed^2), data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-11.6322  -1.3563   0.3619   2.1599   6.5942

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1030.88605    5.31582  193.928 < 2e-16 ***
Air.Temperature  -0.78292    0.17251  -4.538 3.69e-05 ***
Wind.Speed      0.79552    0.34528   2.304 0.02551 *
I(Wind.Speed^2) -0.05468    0.01581  -3.458 0.00113 **
```

Figure 3: Sea Level Air Pressure Model (millibars)

Air Temperature is included in the model as warmer air has a lower density, creating lower pressure conditions. Wind Speed is included in the model as air flows more aggressively in areas with larger changes in pressure. In addition, the quadratic term for Wind Speed is included to represent the two-dimensional effect of wind and reduce bias towards more extreme wind conditions, typically with wind speeds over 30 meters per second.

Air Temperature

```
> summary(temperatureModel)

Call:
lm(formula = Air.Temperature ~ I(Latitude^2) + Sea.Level.Pressure +
    I(Total.Cloud.Amount^2) + Wind.Direction, data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4533 -0.9917 -0.2000  0.9312  5.1283

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.096e+02  6.319e+01   3.317  0.00178 **
I(Latitude^2)  -5.039e-03  6.814e-04  -7.395  2.33e-09 ***
Sea.Level.Pressure -1.771e-01  6.262e-02  -2.829  0.00690 **
I(Total.Cloud.Amount^2) -1.994e+00  7.976e-01  -2.500  0.01604 *
Wind.Direction   -6.487e-03  3.158e-03  -2.054  0.04567 *
```

Figure 4: Air Temperature Model (Degrees Celsius)

The quadratic term for Latitude is included in the model as air temperature generally gets colder the farther in magnitude you are from the equator. As such, Sea Level Pressure is included as warmer air is associated with a reduction in air pressure. Total Cloud Amount is included as the sun contributes to heating air particles in the sky. This term is quadratic to account for the two-dimensional coverage clouds have in the sky. Wind Direction is included as depending on geographic location, certain wind directions have stronger average wind speeds that create colder air.

Wave Height

```
> summary(waveHeightModel)

Call:
lm(formula = I((Wave.Height + 0.01)^0.5) ~ I(Latitude^2) + Wind.Direction +
    Wind.Speed, data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-0.53052 -0.16991 -0.00765  0.13075  0.74977

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.0879315  0.2000523   0.440  0.66456
I(Latitude^2)  0.0008069  0.0001579   5.110  4.04e-05 ***
Wind.Direction 0.0016177  0.0006527   2.479  0.02134 *
Wind.Speed     0.0421038  0.0145661   2.891  0.00849 **
```

Figure 5: Wave Height Model (Meters)

The quadratic term for Latitude is included as waves are typically more aggressive towards the poles due to the Coriolis effect. This effect also directly ties in with the other two regressors for this model: Wind Speed and Wind Direction.

Swell Height

```
> summary(swellHeightModel)

Call:
lm(formula = I(Swell.Height^2) ~ I(Latitude^2) + Wind.Speed +
    I(Wind.Speed^2), data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-2.8124 -1.4108 -0.0772  1.4510  3.6455

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.424209   1.300711  -0.326  0.74808
I(Latitude^2)   0.006101   0.001197   5.095 7.57e-05 ***
Wind.Speed      0.852208   0.268320   3.176 0.00523 **
I(Wind.Speed^2) -0.022426   0.011022  -2.035 0.05690 .

```

Figure 6: Swell Height Model (Meters)

The quadratic term for Latitude is included as swells are more aggressive towards the poles due to the Coriolis effect. Once again, this effect directly ties into the Wind Speed regressor for this model. The quadratic term for Wind Speed is included as swells also concern length, making them more two-dimensional types of observations. Consequentially, this also reduces bias towards more extreme wind conditions, typically with wind speeds over 40 meters per second.

Total Cloud Amount

```
> summary(cloudModel)

Call:
lm(formula = Total.Cloud.Amount ~ Air.Temperature + Wind.Direction,
    data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-0.5339 -0.2280 -0.0216  0.2233  0.5048

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.2741035   0.2874192   4.433 5.24e-05 ***
Air.Temperature -0.0319095   0.0111985  -2.849 0.00639 **
Wind.Direction  0.0007638   0.0003728   2.049 0.04586 *

```

Figure 7: Total Cloud Amount (Proportion)

Air Temperature is included as the reduced air temperature is associated with clouds blocking the sun at times. As such, colder air may signify an increased proportion of cloud coverage. Wind Direction is included due to Hadley Cells, a global atmospheric phenomenon that plays a critical role in the flow of Earth's atmosphere.

Wind Direction

```
> summary(windDirModel)

Call:
lm(formula = I(Wind.Direction^0.5) ~ I(Latitude^2) + Air.Temperature,
    data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-8.3472 -1.5511 -0.1851  1.7612  7.0238

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   35.923191    5.727474   6.272 7.67e-08 ***
I(Latitude^2)  -0.007259    0.001378  -5.267 2.81e-06 ***
Air.Temperature -0.763121    0.199986  -3.816 0.000368 ***
```

Figure 8: Wind Direction (Degrees Clockwise from North)

The quadratic term for Latitude is included as the direction of the wind is heavily dependent on the Hadley Cell airflow pattern of Earth's atmosphere. Air Temperature is included as air always travels from cold air to warm air areas due to the pressure difference. While Sea Level Pressure could be included or substituted, Air Temperature captures this change more aggressively in our model.

Wind Speed

```
> summary(windSpeedModel)

Call:
lm(formula = Wind.Speed ~ I(Latitude^2) + Sea.Level.Pressure +
    Air.Temperature + Wind.Direction + Sea.Level.Pressure:Air.Temperature,
    data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-7.2889 -2.0777 -0.1537  1.4192 10.3812

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.782e+03  9.990e+02   2.784 0.00770 **
I(Latitude^2)  4.866e-03  1.872e-03   2.599 0.01246 *
Sea.Level.Pressure -2.725e+00  9.843e-01  -2.768 0.00803 **
Air.Temperature -9.951e+01  4.022e+01  -2.474 0.01702 *
Wind.Direction -1.167e-02  6.020e-03  -1.938 0.05860 .
Sea.Level.Pressure:Air.Temperature  9.779e-02  3.966e-02   2.466 0.01738 *
```

Figure 9: Wind Speed Model (Meters per Second)

The quadratic term for Latitude is included as winds tend to be more aggressive toward the poles. Sea Level Pressure and Air Temperature are included as air wants to flow from colder, higher-pressure areas to warmer, lower-pressure areas. As such, we will include an interaction term between these two regressors to model this interaction. Wind Direction is included as certain directions have higher average Wind Speeds depending on geographic location.

Precipitation

```
> summary(precipModel)

Call:
lm(formula = Present.Weather.Numeric ~ I(Latitude^2) + Wind.Direction +
    Wind.Speed, data = marine)

Residuals:
    Min       1Q   Median       3Q      Max
-0.38592 -0.13316 -0.04453  0.06366  0.62743

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.2563413   0.1294706  -1.980  0.05358 .
I(Latitude^2) -0.0003342   0.0001000  -3.341  0.00164 **
Wind.Direction  0.0006810   0.0004053   1.680  0.09951 .
Wind.Speed     0.0452378   0.0094765   4.774  1.8e-05 ***
```

Figure 10: Precipitation Likelihood (Proportion)

The quadratic term for Latitude is included as precipitation tends to be greatest at and around the Equator. Wind Speed and Direction are included as both are affected by Hadley Cells in the Earth's atmosphere, affecting the flow of air and rain clouds. Sea Level Pressure should be included, but it was deemed to not be statistically significant in our model, evident in the ANOVA test below:

```
Analysis of Variance Table

Model 1: Present.Weather.Numeric ~ I(Latitude^2) + Wind.Direction + Wind.Speed
Model 2: Present.Weather.Numeric ~ I(Latitude^2) + Wind.Direction + Wind.Speed +
  Sea.Level.Pressure
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      47 2.7475
2      46 2.6170  1    0.1305 2.2938 0.1367
```

Figure 11: ANOVA Table AH: + Sea Level Pressure

3.2.2 Model Generalization Matrix

To represent all of these linear models cleanly, we can transform all 8 model predictions into the inverse standard regression form equation below

$$((\beta_0 + \sum_{i=1}^{11} x_i^{\beta_i})^{-\beta_{12}}) - \beta_{13} \quad (1)$$

and use the row values in the table below (12) for values $\beta_1, \dots, \beta_{13}$ for each model. Hypothetical zeros require log transformations, hyphens omit the observation, and commas include multiple powers of a regressor.

	Latitude	Longitude	Sea.Level.Pressure	Air.Temperature	Wave.Height	Swell.Height	Total.Cloud.Amount	Wind.Direction	Wind.Speed	Present.Weather.Numeric	Sea*Air	RespPower	RespOffset
Sea Level Pressure	-	-	-	1	-	-	-	-	1.2	-	-	1	0
Air Temperature	2	-	1	-	-	-	2	1	-	-	-	1	0
Wave Height	2	-	-	-	-	-	-	1	1	-	-	0.5	0.01
Swell Height	2	-	-	-	-	-	-	-	1.2	-	-	2	0
Total Cloud Amount	-	-	-	1	-	-	-	1	-	-	-	1	0
Wind Direction	2	-	-	1	-	-	-	-	-	-	-	0.5	0
Wind Speed	2	-	1	1	-	-	-	1	-	-	1	1	0
Present Weather Numeric	2	-	-	-	-	-	-	1	1	-	-	1	0

Figure 12: Generalized Power and Offset Matrix for Each Linear Model

3.3 Inverse-Response and Box-Cox Plots

With our models established, we must justify our response transformations by analyzing the Inverse Response and Box-Cox method plots. As such, these must suggest no other transformation is necessary or can be performed. The IR-BC plots for each of the 8 models are on the next page in Figure 13.

Addressing The IR-BC Plots

Analyzing the IR-BC matrix, six of the eight models require no further transformation as suggested by both a Box-Cox plot containing $\lambda = 1$ in the 95 percent confidence interval for log-likelihood maximization and $\hat{\lambda}$ estimates in the Inverse Response plots fitting relatively close to the linear transformation $\hat{\lambda} = 1$. However, we still have two cases to address. For the Present Weather Numeric (Precipitation) variable, these values were artificially generated by converting a previously categorical variable into a numeric one. As such, neither plot provides relevant information for this model. For Sea Level Pressure, it appears that the RSS does not converge to a value for $\hat{\lambda}$. However, $RSS(\lambda = 1) \approx 265.9$ and $RSS(\lambda = 10) \approx 263.3$. Because $RSS(1)$ and $RSS(10)$ are quite close, it is reasonable to assume that no further transformation is needed, even in the absence of convergence in the Box-Cox log-likelihood maximization.

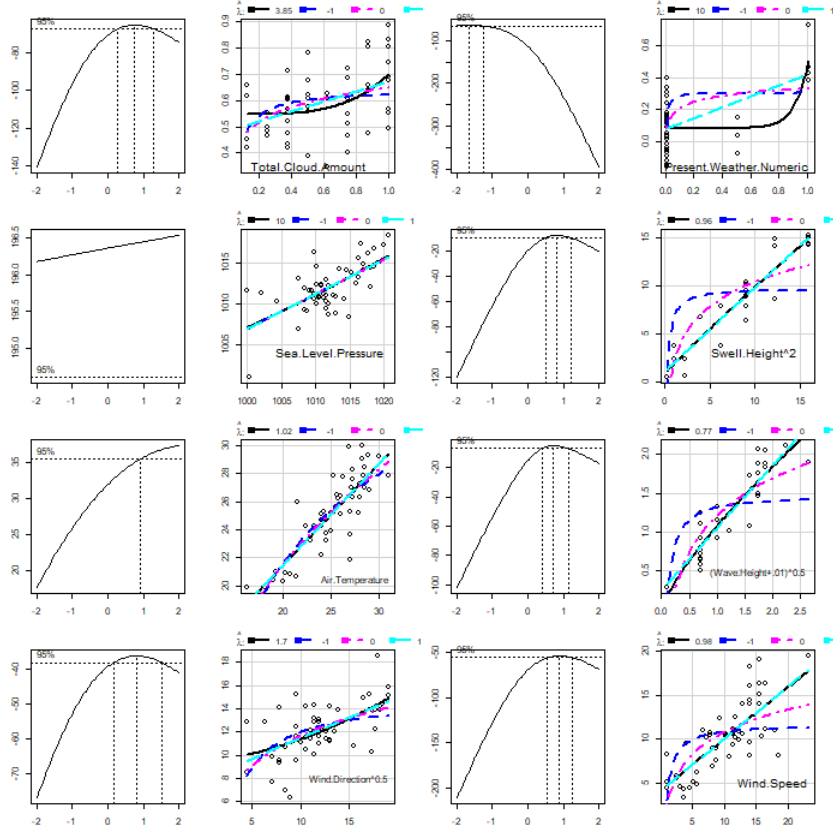


Figure 13: Inverse Response and Box-Cox Matrix

With our statistical methods and linear models established and verified, we have the framework to create maritime weather forecasts.

4 Simulating Forecasts

To put these models in action, we will perform two forecast demonstrations starting at coordinates (6.1, 94.6). These demonstrations will involve a realistic simulation of traveling to Mayabunder, India (12.9, 92.9) over 5 days and an unrealistic simulation of traveling to Augusta, Western Australia (−34.3, 115.2) over 5 days. The red line on the world map graphs indicates the initial path of the vessel with buoy D5GN6 while the blue dots represent the daily Latitude and Longitude values to input for future weather forecasts. As such, the path maps and forecast tables are generated below.

Initial Path (Jan 12 to Jan 31)

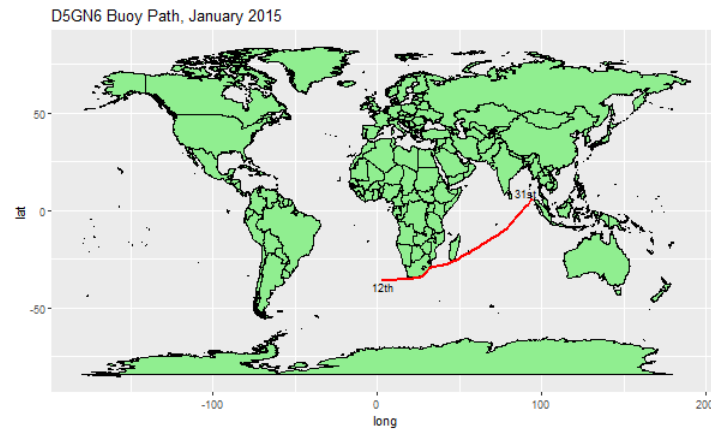


Figure 14: Buoy Path from original data, Jan 12th to Jan 31st

4.1 Simulation 1: Mayabunder, India

5-Day Path Visualization

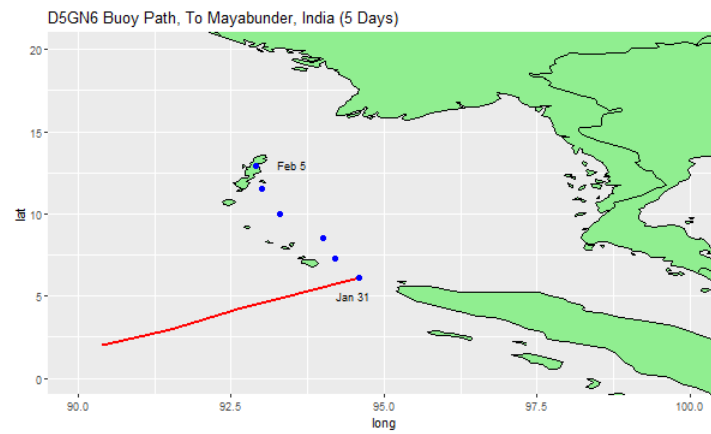


Figure 15: Hypothetical 5-Day Path to Mayabunder, India

5-Day Weather Forecast Table

	Jan 31	Feb 1	Feb 2	Feb 3	Feb 4	Feb 5	Units
Latitude	6.10	7.30	8.50	10.00	11.50	12.90	Degrees
Longitude	94.60	94.20	94.00	93.30	93.00	92.90	Degrees
Sea Level Pressure	1009.14	1011.57	1010.31	1011.25	1011.23	1011.60	Millibars
Air Temperature	28.22	29.96	28.48	28.77	28.26	28.15	Degrees C
Wave Height	0.50	0.33	0.59	0.38	0.55	0.54	Meters
Swell Height	1.00	2.35	2.27	1.96	2.19	2.20	Meters
Total Cloud Amount	0.38	0.42	0.47	0.49	0.50	0.51	Cloud %
Wind Direction	60.00	199.31	160.53	186.77	175.39	179.52	Degrees-CW from N
Wind Speed	8.70	7.69	5.20	6.55	6.25	6.86	m/sec
Present Weather Numeric	0.01	0.17	0.21	0.06	0.13	0.10	Rain Chance

Figure 16: 5-Day Maritime Weather Forecast to Mayabunder, India

4.2 Simulation 2: Augusta, W. Australia

5-Day Path Visualization

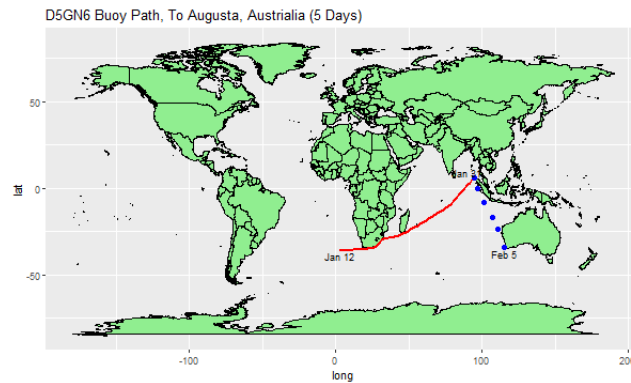


Figure 17: Hypothetical 5-Day Path to Augusta, W. Australia

5-Day Weather Forecast Table

	Jan 31	Feb 1	Feb 2	Feb 3	Feb 4	Feb 5	Units
Latitude	6.10	0.00	-8.50	-17.00	-23.50	-34.30	Degrees
Longitude	94.60	97.20	101.00	107.00	110.50	115.20	Degrees
Sea.Level.Pressure	1009.14	1011.57	1010.31	1010.97	1011.25	1012.33	Millibars
Air.Temperature	28.22	29.96	28.75	28.70	27.39	26.25	Degrees C
Wave.Height	0.50	0.33	0.53	0.39	0.76	1.13	Meters
Swell.Height	1.00	2.35	2.19	1.92	2.40	2.82	Meters
Total.Cloud.Amount	0.38	0.42	0.47	0.49	0.50	0.51	Cloud %
Wind.Direction	60.00	199.31	170.48	181.21	142.11	121.24	Degrees-CW from N
Wind.Speed	8.70	7.69	4.94	6.24	7.25	9.83	m/sec
Present.Weather.Numeric	0.01	0.17	0.23	0.06	0.05	-0.02	Rain Chance

Figure 18: 5-Day Maritime Weather Forecast to Augusta, W. Australia

4.3 Strengths and Weaknesses of Forecast

Comparing the generated maritime to our original data, three main strengths exist. First, our forecasts generated prediction values that are both realistic for the time of year and relatively responsive to sudden changes in Latitude as evident in the Augusta forecast. Second, these forecasts are extremely simple to interpret, especially in comparison to the original observations. Third, the forecast generation is relatively powerful compared to the relatively high-dimension, low-observation characteristics of the original data. Despite these strengths, there are also two major weaknesses in the forecast. First, the predictions generated for each daily weather column depend on the previous day's observations. As such, there will always be some amount of underlying colinearity between the values in each day column. Second, the forecast algorithm tends to be sufficient for roughly 7-14 iterations (days), representable of a real-life weather forecast. Beyond this time frame, some of the weather values may become unstable, trend towards undefined values like NaN and ∞ , or become uninterpretable. Even in our shorter 5-day simulation to Augusta, the Precipitation Chance on February 5th generated a negative probability.

5 Conclusion

In this paper, we analyzed maritime weather measurements from the NOAA, transformed these measurements for exploratory data analysis and modeling, created linear models based on maritime meteorological theory, transformed these models into a weather forecasting algorithm, and analyzed the major strengths and weaknesses behind our forecasting process. As such, there are several actions one could take for further data exploration and fitting. Some of these include applying more advanced regression or machine learning models, using bootstraps to increase the number and power of observations, and applying deeper meteorological and atmospheric theory to fine-tune the model. It is worth noting that some R packages exist specialized specifically for weather analysis and forecasting. As such, the proper application of relevant packages would fit a more accurate and cleaner model than what was created here. However, using R packages was not the objective of this project. Instead, we applied regression modeling techniques, statistical testing, and meteorological theory to create our own weather forecasting model from scratch.

References

- [1] Nathaniel Bowditch. *American Practical Navigator: An Epitome of Navigation*, volume 1. The Nautical Almanac, 1st edition, 2017.
- [2] National Centers for Environmental Information (NCEI). Climate data online: Dataset discovery, August 2023. Accessed 12-13-2024.
- [3] Mark Z. Jacobson. *Fundamentals of Atmospheric Modeling*. Cambridge University Press, Cambridge, May 2005.
- [4] World Meteorological Organization. *Manual on Codes Volume I.1 Annex II to the WMO Technical Regulations Part A – Alphanumeric Codes, 2011 edition, updated in 2017*. WMO No.306, Part I.1 Part A. World Meteorological Organization, Geneva, Switzerland, 2017.