Planning “Sick” Days using Oceanic Buoys
DSCI 478: Final Paper
Trevor Isaacson and Evan Kessler
5/4/22

Introduction:

As a student, there is no better feeling than waking up to a snow day. Classes and responsibilities for the day are canceled and it's a great opportunity to build a ski ramp in your backyard. We both enjoy skiing and there’s no better feeling than skiing on a couple inches of fresh snow. However, it is impossible to drive to any ski resort on a snow day. What if we could plan ski outings, days or weeks in advance?

We recently found a blog called PowderBouy [1]. Powder Buoy is a blog in which the writer has found a correlation between oceanic patterns off the coast of Hawaii and snowfall in the Rockies, especially in Utah, Wyoming, and Colorado. It describes how low-pressure systems and winter storms tend to bring moisture east from the coast.

But how is this correlation measured? The National Oceanic and Atmospheric Administration, or NOAA, has observation buoys sitting in oceans all around the world. These buoys send data back to the Nationa Buoy Data Center around the clock. NOAA has various types of buoys that collect data relating ocean climates, tsunamis, acidification and even monsoons. For example, Peggy has been taking observations in the Bering Sea since 1995 [2]. By observing variables such as wave height and air pressure, PowderBuoy can track potential snowstorms almost two weeks away. Imagine planning the days you wish to ski two weeks in advance!

However, is there an actual correlation or is this simply a string of good luck? Can a small weather buoy thousands of miles away actually help us determine the best days to ski? In this paper, we attempt to verify this blog’s past research while also attempting to build upon it.

Data:

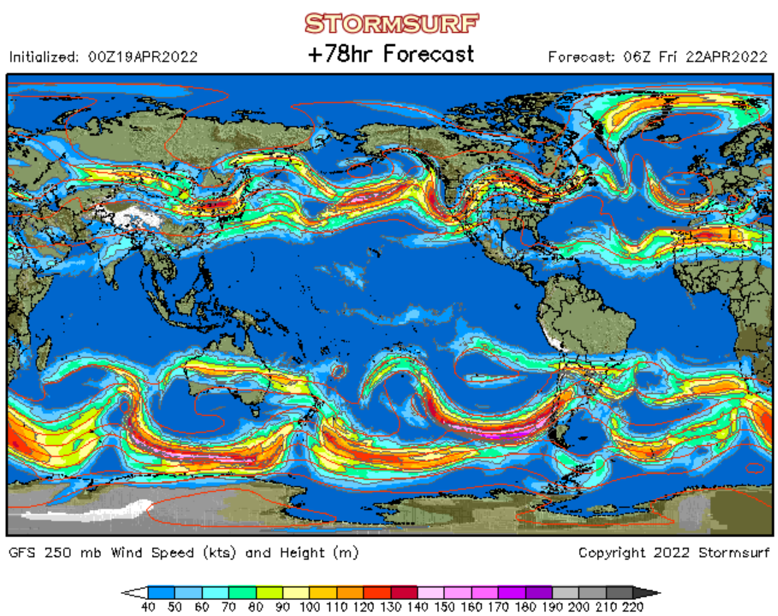
The data for the NOAA buoys is free and public and can be downloaded from the National Data Buoy Center [3] website. There are hundreds of buoys with available data all across the globe. Data is collected every ten minutes and supply data regarding wind, waves, temperature, and pressure.

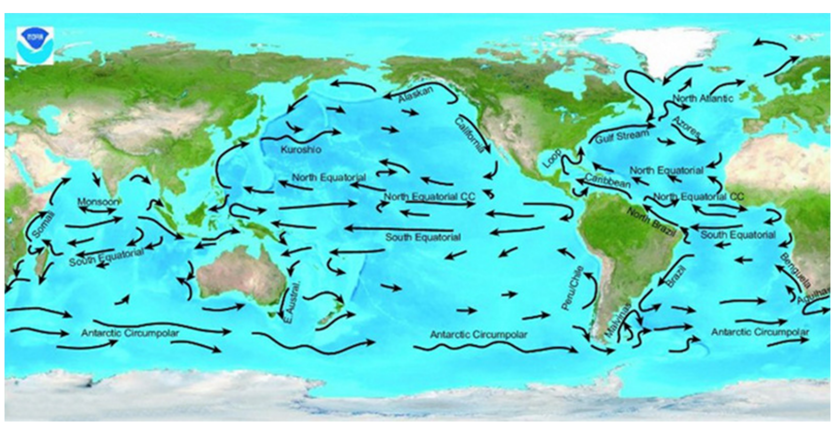
Our snowfall data was downloaded from the Applied Climate Information System or SC-ACIS [4] and it provides access to daily climate weather for various weather stations around the country. Information includes climate normals, daily almanacs, and daily summaries of precipitation, snowfall, temperature, and other measurements. For our research, we used the daily snowfall summaries.

Climate:

Before getting too far into our research, there needs to be a general understanding of how climate works and where potential storms may originate. The article, “How Does the Ocean Affect Climate and Weather on Land” by NOAA Ocean Exploration [5], takes the reader through the process on how ocean plays a role with weather. To start, most radiation form the sun is absorbed by the ocean, especially the pacific as it covers over 30 percent of the earths surface [6]. According to NOAA, “Ocean water is constantly evaporating, increasing the temperature and humidity of the surrounding air to form rain and storms that are then carried by trade winds. In fact, almost all rain that falls on land starts off in the ocean”, and snowfall is no different. These storms are then carried across the ocean largely due to ocean currents. In the image below, we see the typical ocean current systems across the world. Focusing mainly along the northern pacific ocean, we can see most ocean currents approach the western United States from

Alaska/Canada. The other image depicts the weather patterns in a 78 hour period earlier this year (April of 2022). Again, as we can see storms seem to travel along the coastline of Canada and California to then be swept upwards into the Colorado/Utah/Wyoming region. Understanding this, it seems best to use buoys located between Alaska and Califronia than the one initially proposed by Powder Buoy off the coast of Hawaii. Based on ocean currents and weather patterns surrounding Hawaii, data from Powder Buoy would most likely not travel towards the United States.





Continued Research:

We first wanted to verify there is some sort of correlation between Hawaii oceanic patterns and Utah mountain snowfall roughly two weeks later. The blog details how the main data point of interest is wave height. Wave height tends to spike or jump during low-pressure systems as these are typically systems that bring higher winds and precipitation. These same systems eventually bring snow to the Rocky Mountains. Along with wave height, other potential variables of interest are wind speed and temperatures including sea surface temperature, air temperature, and dewpoint temperature.

To begin our verification research, we found PowderBuoy uses Buoy Station 51001, which is located 188 nautical miles northwest of the Kaui. This is the same buoy PowderBuoy and the University of Utah use to track potential storms. We also found snow totals for a few ski resorts including Park City, Utah, and Jackson Hole, Wyoming. These are two very popular resorts in the Utah, Wyoming, and Colorado tristate area. By using the original buoy and ski resorts within the area, we can build upon past research and conduct our research.

Regression and Decision Trees:

To conduct our research and build upon past results, we decided to incorporate linear regression and decision trees. The purpose of using these techniques is to verify wave height is a significant predictor of future snowfall. For example, for snowfall data regarding Jackson Hole and Park City, wave height was a statistically significant variable in our linear regression models. Other significant variables include air temperature, dewpoint temperature and sea level pressure. When attempting to predict the amount of snowfall using linear regression, we saw RMSE values around 1 inch. This initial verification showed this research has merit and we can continue with more sophisticated models for more accurate snowfall prediction.

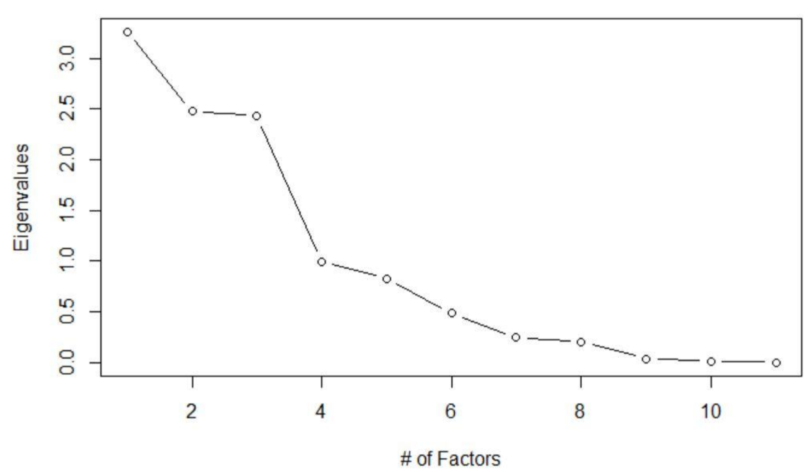
After linear regression, we progressed our research using decision trees like random forests. Using random forests, we were more concerned about predicting snow happening and not the amount of snow. For both Jackson Hole and Park City, we saw classification errors between 0.11 and 0.22, depending on the resort. Given the variability of weather, we feel this coincides with ourl linear regression results. However, when attempting to predict actual snowfalls using random forests, we saw RMSE values similar to linear regression.

So far, our research has shown that the weather around this Hawaii buoy can help determine if snow will fall about two weeks later in Wyoming and Utah. However, these methods haven't showed the ability to accurately predict the amount of snowfall. This leads us to our last method of Factor Analysis.

Factor Analysis:

Factor analysis holds two different phases, exploratory and confirmatory. Both mean what their respected names seem to entail; exploratory looks at data without any prior hypothesis. P.H.D Diana D. Suhr describes the differences between exploratory and confirmatory factor analysis in her paper, "Exploratory or Confirmatory Factor Analysis?" [8]. Suhr states that confirmatory factor analysis is used best to dictate an entire study including the data gathering. Exploratory factor analysis however is used as a dimensional/variable reduction technique similar to PCA, in fact one of the methods used for factor analysis is the principal component method [8]. One of the assumptions underlying Exploratory Factor Analysis that would apply to our dataset are all variables holding a normal distribution.

After several attempts with machine learning using decision trees and not the greatest outputs, we decided to pivot our strategy. That pivot was towards instead analyzing the covariance relationships among our observed buoy variables and their unobservable underlying variables which we will call factors. These factors can be regarded as how similar our variables are to each other based on their correlations. Variables held in one group will be very highly correlated while they may hold very small correlations with variables in other groups/factors. Some variables will also hold equal correlation between factors, meaning they are equally similar to multiple different groups of variables. Based on the 11 variables we can expect to see anywhere around 1-5 factors, as most factors will contain 2-3 related variables. To find the best number of factors, we will need to build a scree plot [7]. This can be compared to our Gini coefficient plots used in decision trees to find the most relevant variables. Instead of checking individual variables, we are looking at the eigenvalues of each factor. On our scree plot, we are looking for a 'bend' in the graph, someplace where the eigenvalues explained by the factors take a dip and levels out. According to Suhr, "Kaiser's criterion, suggested by Guttman and adapted by Kaiser, considers factors with an eigenvalue greater than one as common factors" [8]. This means we will be looking for the number of factors where we see a bend and is greater than a value of one on our y-axis.



Looking at our eigenvalues and the number of factors graph, we can see that our bend occurs around two to four factors. According to our quote above from Suhr's paper, the largest factor we would consider is four as that is the last factor holding an eigenvalue over one [8]. This is also our last factor in which our 'bend' seems to occur. In this case, we would intend on testing each number of factors from 1-4 to see which may yield the most accurate response.

Before we jump into testing each different factor, we would like to see our 4 different factors in action. We will do this by looking at a matrix of loadings and communality. This matrix checks each one of our variables and for each factor PC1-4, how much each is responsible for a variables variance. The range of values for our loading matrix is -1 to 1, a value of -.6 would be equivalent to a value of .6 just in different directions. The closer the number is to plus or minus 1 the more that factor explains the variables variance. Our final column in the matrix is responsible for reporting each variables communality, meaning how well represented each variable is by all factors combined (out of 1.0) [7]. Looking at our factors it seems when we use 4 factors only two variables are explained by our 4th factor, one of those being very close to 0 (-.108) meaning it hardly is explained. However, it seems that our fourth factor is responsible for almost all of our variance in SeaSurfaceTemp. Due to the fact that our fourth factor in our scree plot was on the cusp of 1 eigenvalue and post bend as well as our fourth factor only explaining one or two total variables it would be beneficial to extradite our fourth variable when doing our analysis'.

	PC1	PC2	PC3	PC4	
WindDir	-0.37103680	-0.07948679	0.40027169	-0.107904822	0.3158473
WindSpeed	0.52970016	0.79766749	-0.21457302	-0.003254933	0.9629079
GustSpeed	0.46134835	0.84506510	-0.20703167	-0.003687392	0.9698530
WaveHeight	-0.20813277	0.91861826	0.11591634	0.009201649	0.9007000
DominantWavePeriod	-0.64627031	0.30982293	0.46956565	-0.004226561	0.7341653
AverageWavePeriod	-0.69740806	0.31856952	0.51922657	0.039775864	0.8590429
DirWavePeriod	-0.64184449	0.15869387	0.47565611	-0.035198642	0.6646358
SeaLevelPressure	0.58375097	0.11572354	0.70916744	-0.007439006	0.8571309
AirTemp	0.67923508	-0.14038157	0.69621362	-0.030589703	0.9667164
SeaSurfaceTemp	0.02149988	-0.02671528	0.09427557	0.990534823	0.9912231
DewpointTemp	0.68564719	-0.14301833	0.67643267	-0.041502366	0.9498499

We don't stop at just testing the different number of factors either. In factor analysis, there are three different parameters in which we need to change. In R we will be looking at two functions, fa() and principal(). The reason we hold two different functions is based on the method we use to perform factor analysis. Fa() is used for maximum likelihood estimate (MLE) while principal() is used for the principal components method. Our last parameter that we will change throughout the iterative process is the rotation. There are two different rotations we will be using, one simply being no rotation while the other being a varimax rotation. This rotation is primarily used to combat interpretability by selecting the orthogonal transformation that maximizes the sum of our variance of squares of the scaled loadings [7]. Looking at these three parameters there will be 12 different analysis done on our data in order to acount for each iteration of nfactor (3 total), rotation (2 total), and method (2 total).

Application to Fort Collins and Breckenridge:

Because of the success seen with the original buoy using factor analysis, we decided to try and find our own buoy and predict snowfall totals for Fort Collins and Breckenridge using factor analysis. Based on our understanding of winter weather patterns, we decided to use data from buoys located off the coasts of Cape Edgecumber, Alaska, Tillamook, Oregon and San Francisco, California. Using locations along the Pacific coast and various time delays, we were able to predict snowfall totals for the new locations. Using RMSE, we found the California buoy and the Cape Edgecumber buoy to have the best prediction results with the Cape Edgecumber having the best predictions for both Breckenridge and Fort Collins.

Buoy <fctr>	Resort <fctr>	Rotation <fctr>	Factors <fctr>	RMSE_Inches <fctr>
West Cali	Breck	None	3	0.4621
West Cali	Breck	None	3	0.4642
West Cali	FoCo	None	2	0.4031
West Cali	FoCo	Verimax	1	0.4049
Cape Edgecumber	Breck	None	3	0.4075
Cape Edgecumber	Breck	Verimax	3	0.4092
Cape Edgecumber	FoCo	Verimax	1	0.3411
Cape Edgecumber	FoCo	None	2	0.3508

These predictions are almost twice as good as linear regression and random forests. Of course, weather is very hard to predict, so we think these results only being about 0.4 inches off from actual snowfall totals is impressive. This is definitely helpful in determining our ski and snow days and we cannot wait to apply this research next winter.

Citations/Sources:

[1] “Connecting Soul Riders from Hawaii to the Rockies.” *Powder Buoy - Connecting Soul Riders from Hawaii to the Rockies*, https://powderbuoy.com/.

[2] Valentine, Katie. “Meet 5 NOAA Buoys That Help Scientists Understand Our Weather, Climate and Ocean Health.” *Welcome to NOAA Research*, Welcome to NOAA Research, 3 June 2021, https://research.noaa.gov/article/ArtMID/587/ArticleID/2762/Meet-5-NOAA-Buoys-that-help-scientists-understand-our-weather-climate-and-ocean-health.

[3] US Department of Commerce, National Oceanic and Atmospheric Administration. *National Data Buoy Center*, 8 Nov. 1996, https://www.ndbc.noaa.gov/.

[4] “SC acis2.” *SC ACIS2*, http://scacis.rcc-acis.org/.

[5] US Department of Commerce, National Oceanic and Atmospheric Administration. “How Does the Ocean Affect Climate and Weather on Land?” *Ocean Exploration Facts: NOAA Office of Ocean Exploration and Research*, 21 June 2013, https://oceanexplorer.noaa.gov/facts/climate.html.

[6] US Department of Commerce, National Oceanic and Atmospheric Administration. “How Big Is the Pacific Ocean?” *Ocean Exploration Facts: NOAA Office of Ocean Exploration and Research*, 5 Mar. 2013, https://oceanexplorer.noaa.gov/facts/pacific-size.html.

[7] Richard A. Johnson, Dean W. Wichern. “Applied Multivarite Statistical Analysis Sixth Edition”. CH. 9. *Factor Analysis and Inference for Structured Covariance Matrices*. 2008.

[8] Suhr, Diana D. “Exploratory or Confirmatory Factor Analysis?” *Statistics and Data Analysis, University of Northern Colorado.* April 2022.