Utilizing Meteorological Data with Supervised Learning to Predict Snowfall Amounts at Copper Mountain Ski Resort

Dustin Rapp

1

Introduction

- Prognostic models have difficulty predicting snowfall for specific mountains and slopes
- More accurate site specific snowfall forecasts could potentially improve ski resort operations
- Copper Mountain unique as multiple official meteorological stations are near or onsite, each with freely available data
- Could supervised learning be used to site specific snowfall?

2

Overview

› Outline
- Data Utilized
- Modeling Methodology
- Feature Selection
- Model Performance
- Model Diagnostics
- Recommendations

› Goals:
- › 1.) Run a supervised learning model to predict snowfall using readily available meteorological data
- › 2.) Make recommendations based on findings

3

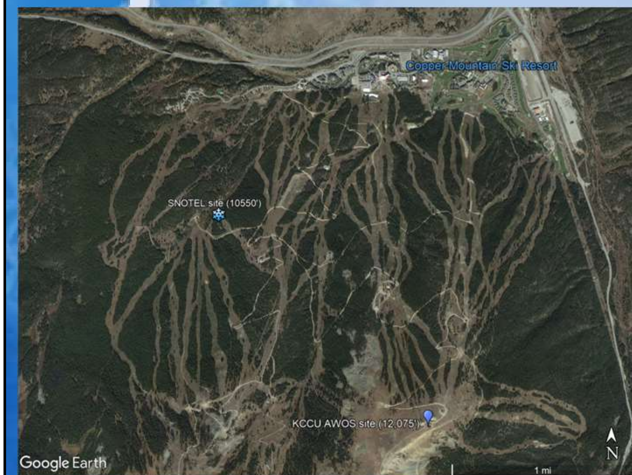
Surface Data

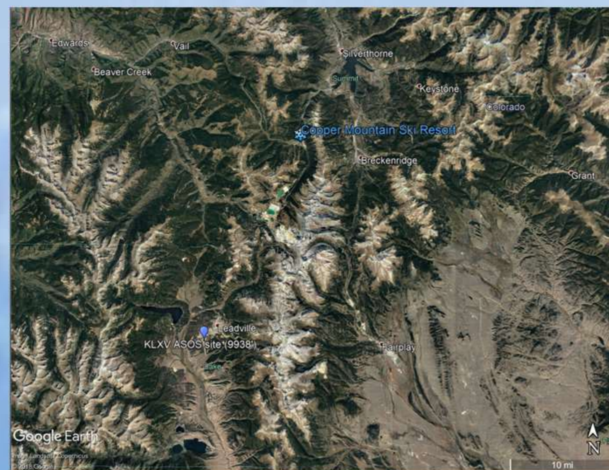
Station ID	Station Type	Location	Elevation	Meteorological Measurements Utilized	Temporal Resolution
SNOTEL 415	NRCS SNOTEL	Popular Ski Runs midway up Copper Mtn	10550’	• Temperature • Snow Depth	Hourly
KLXV	NWS ASOS	~25 mi SW of Copper Mtn	9938’	• Temperature • Dewpoint • Wind Speed • Wind Direction • Pressure • 12-hr Pressure Changes*	Hourly
KCCU	CDOT AWOS	Near top of Copper Mtn	12075’	• Temperature • Dewpoint • Wind Speed • Wind Direction	Hourly

*Calculated feature

4

Surface Data Monitor Locations





5

Additional Upper Air Data

- Snowfall is highly dependent on physics of crystal development, which is affected by upper air physics. Therefore, upper air data was also included

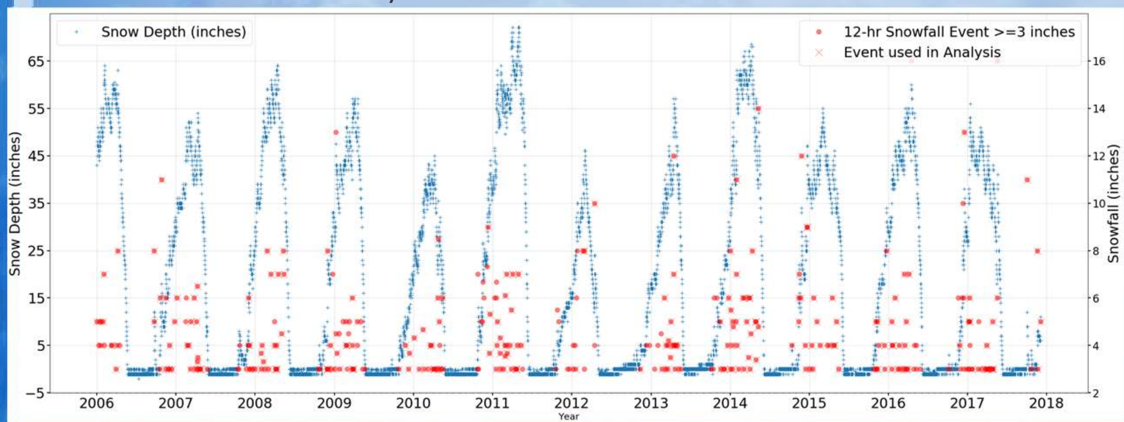
Station ID	Location	Pressure Levels	Measurements at each Level	Temporal Location
KJCT	Grand Junction, CO	200mb, 250mb, 300mb, 400mb, 500mb, 700mb, & 850mb	• Height • Temperature • Dewpoint • Wind Speed • Wind Direction	12 hour

- Differences in meteorological measurements between Pressure Levels are common considerations when forecasting snowaffall so were also calculated and included (e.g. 500mb Dewpoint minus 400mb Dewpoint) as additional features

6

Snowfall vs Snow Depth

- SNOTEL stations report *snow depth*. Snowfall was calculated.
- 12-hr snowfall only. Events >=3" due to noise in data.



Modeling Methodology

- Ordinary Least Squares (OLS) model chosen as most basic supervised learning approach
- Linear Regression Analysis performed for all features
- Datasets divided into two Training/Test partitions
- Features for use in model chosen using forward stepwise approach
- Two model runs were fit on each of the two training partitions:
  - Surface Data Feature Only
  - Surface Data plus Upper Air Data Features
- Fitted models were then tested on respective test datasets to assess snowfall prediction

8

Test/Training Partitions

- Two partitions were developed as two modeling scenarios would give better assessment of real world model performance
- Priority was given to preserve seasonality in test/train set by choosing full year groupings which resulted 80/20 splits
- Partition A:
  - Test set: Years 2016 and 2017
  - Train Set: Remaining Years 2006-2015
- Partition B
  - Test Set: Years 2014 and 2015
  - Train Set: Remaining years - 2006-2013, 2016-2017
  - Each of the these divisions resulted in ~80% Training/~20% Test Datasets

9

Feature Selection

- Statsmodel using forward stepwise approach on Training Datasets
- Adjusted R^2 used as metric
- Optimized such that individual feature t-statistic confidence intervals (p values) were minimized
- Upper Air Data did improve Adjusted R^2

Partition A – Best Combination of Features

Dataset	Adj. R-squared:	F-statistic:	Prob (F-statistic)
Surface Data Only	0.045	2.818	0.0177
Surface+Upper Air Data	0.189	3.493	8.73e-06

Partition B – Best Combination of Features

Dataset	Adj. R-squared:	F-statistic:	Prob (F-statistic):
Surface Data Only	0.100	5.309	0.000137
Surface+Upper Air Data	0.303	1.38e-09	0.000229

10

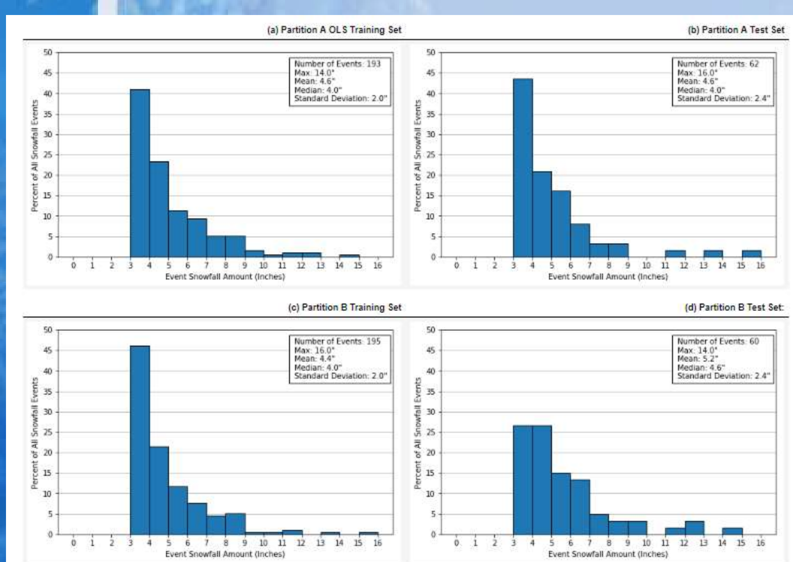
Model Performance of Snowfall Predictions

Model	Sci-kit R2 Score	RMSE (inches)
Partition A (Surface Data Features only)	0.018	2.392
Partition A (Surface+Upper Air Data Features)	0.052	2.35
Partition B (Surface Data Features only)	-0.164	2.622
Partition B Surface+Upper Air Data Features)	-0.293	2.764

- Partition A model performed better then Partition B model
- The Partition B models performed very poorly by all metrics
- Upper Air Data did improve performance in the Partition A cases.
- Why such bad performance for Partition B??

Much room for improvement! But many opportunities..

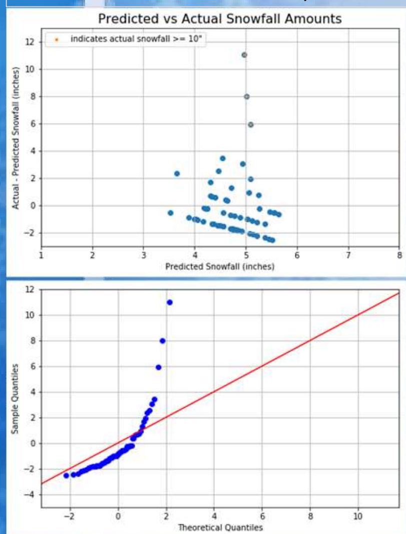
Snowfall Distributions between Training and Test Set



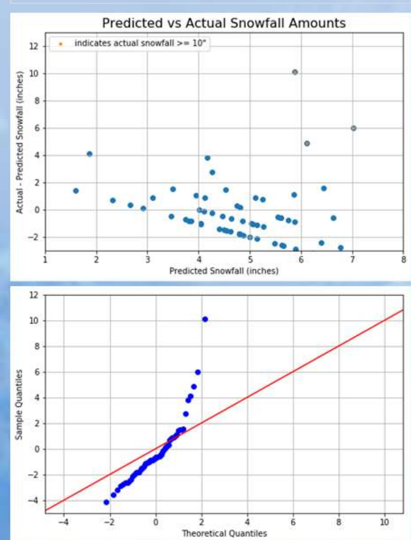
- Partition B Test and Training set snowfall distributions very different, likely causing poor performance
- Both Partition A Test/Train sets are heavily weighted by lighter snowfall – may tend to underpredict higher snowfall events.

Upper Air Data Inclusion (Partition A)

Surface Data Only



Surface+Upper Air Data

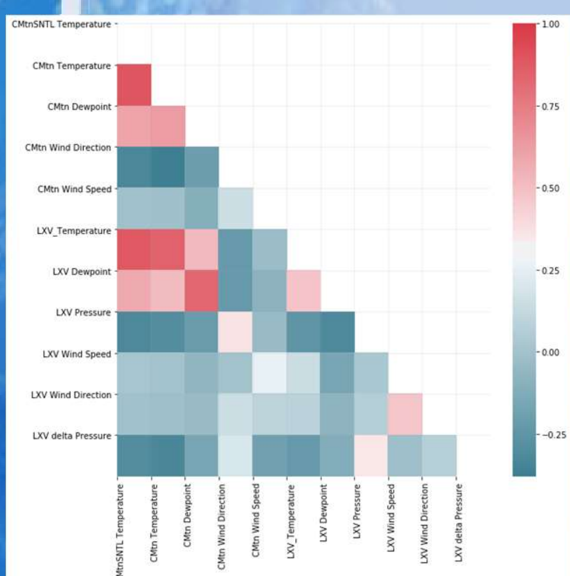


Large number of smaller snowfall events in training set caused lower underpredictions in higher snowfalls

Despite large number of small snowfall events, upper air data did improve predictions of these events

13

Multicollinearity and Non-linear features



› Multicollinearity
  – Due to proximity, temperature trends between sites were similar
  – In upper air data, heights were strongly correlated.

› Non-linear features
  – Though there was found to be predictive capability, Wind Direction is by nature not a linear variable and may not be beneficial in all cases, especially when winds are more northerly.

› Likely negatively influencing model

14

Non-linearity considerations

- Sometimes 1" of melted can equate up to a foot of snow. Finer snow could result in two feet or more
- Snowfall must consider temperatures where the snow develops (upper air), as well as amount of moisture available.
  - Breaking model down into finer components (e.g. try to predict amount of moisture to fall vs snowfall) may be better approach

15

Major Recommendations

- The top three recommendations to improve performance are:
  1. Investigate collinear features more, and consider eliminating some of the strongly collinear features
  2. Work to balance the distribution of snowfall event sizes in the training sets to improve prediction of larger snowfall events.
  3. Consider breaking complexity of model by predicting only the amount of moisture that is expected to fall at first - then add additional complexity of snow depth.

16

Acknowledgements

› Thanks to Dr. Guy Maskall of Springboard for mentorship during the preparation of this analysis

17