Problem Identification for **Alta Snowfall Analysis**

P

Context

Skiers love skiing in good snow conditions:

- **want fresh snow**
- **new snow covers more objects**
- **new snow softens up the surface they ski**
- **need to plan trips far in advance of forecasts**
- **skiers have heuristics to help them in this manner (Japanuary, March is Vail's snowiest month, Alta gets great snow all season)**
- **some ski areas do have snowfall patterns that could help plan a better vacation time for powder**
- **we examine the snowfall at Alta Ski Resort in Utah to look for snowfall patterns**

Specifics

- **Alta has operated since 1938**
- **NOAA snowfall data available from Nov 1944 onward from base area**
- **renown for its abundant and consistent snow**
- **We explore their data to see how consistent**

We turn to modeling for the solution….

P K M R S

Problem Statement for Alta Snowfall Analysis

P

How can we use daily snowfall data at Alta Ski Resorts to determine the best time to take a ski trip to either: A. maximize their chance for the most snowfall B. minimize their chance of having the worst conditions (no new snowfall).

Objectives

- Explore Snowfall at Alta to determine trends and variable effects
- Model snowfall at Alta based on available features
- Determine best times to ski Alta

Focus

- Time effect of seasonal snowfall
- Lag variables (temperatures and precipitation) that can be measured far enough to plan a trip
- Consider Pacific Decadal Oscillation and Atlantic Multi-Decadal Oscillation

Constraints

- snowfall and weather data available from Nov 1944 onward
- PDO and AMO only available monthly

Not considered:

- Short term variables (available close to measurement date)
- Southern Oscillation Index (El Nino / La Nina)
- Barometer readings
- Wind Data
- Sunshine measurements / cloud cover
- Salt Lake Levels / Temperature

P K M R S

Key Findings for Alta Snowfall Analysis

K

We found that snowfall at Alta was generally consistent throughout the ski season with our model performing only slight better than the dummy model of assuming the average. There is only a small advantage of planning a ski trip based on our model.

Model Accuracy

- Root Mean Squared Error for our model (avg for 3 years): 13.30in
- Root Mean Squared Error for dummy model: 13.90in
- Improvement: 0.60 in

Models used

- Final Model: Facebook's Prophet with Random Forest on Residuals
- Also tried:
  - Prophet with Regression
  - Prophet alone
  - Prophet with Regression and more variables

Possible Issues

- There is a lot of variability in the snowfall making it difficult to predict
- Snowfall is exponentially distributed, but we did not take the log before modeling
- Did not consider some lag variables that may improve our model
- Did include some short-term variables that could not be measured far enough from ski vacations
- Modeled on a 7-day snowfall sum; longer window may decrease variability

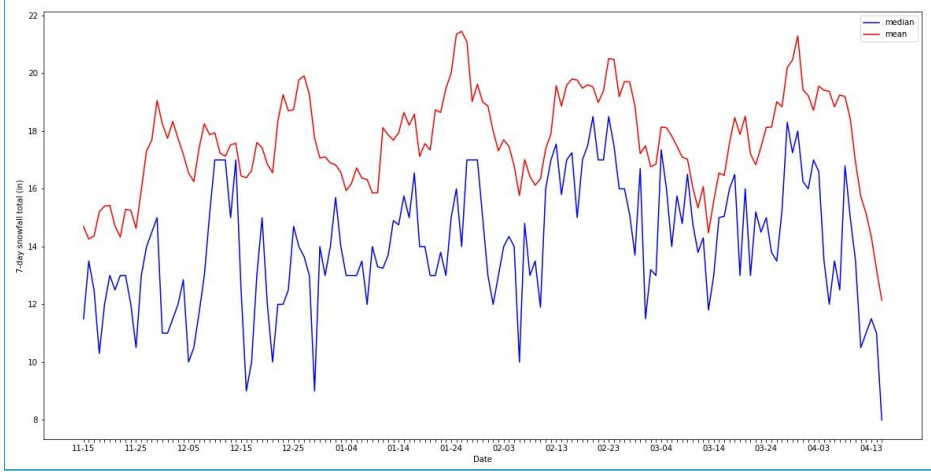
Exploratory Data Analysis for Alta Snowfall

K

Before we modeled Alta's snowfall over the season, we explored the data to find potential nuances.

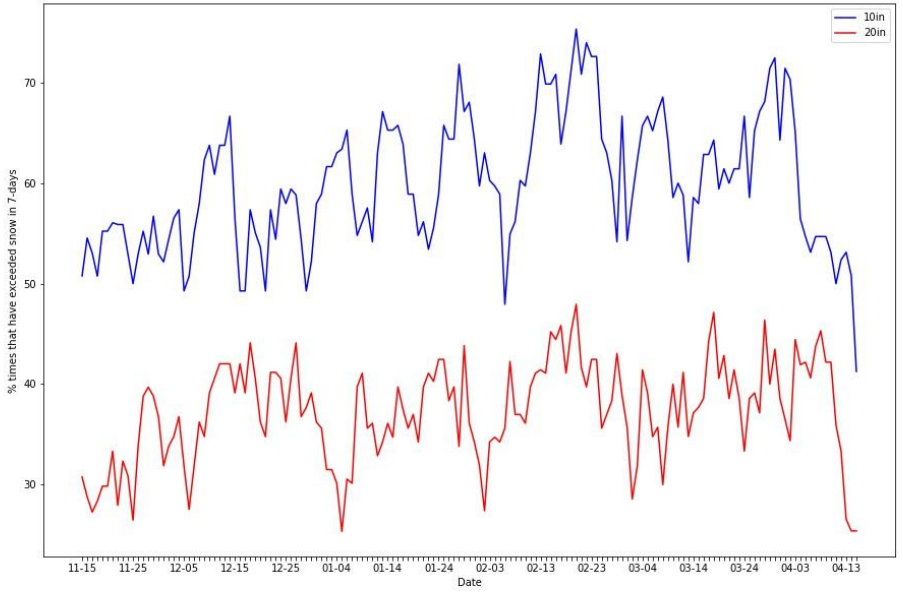
7-day mean & median

- We computed the 7-day rollaway sum of snowfall for reach day of the winter
- Below are the mean and median of those 7-day sums
- These values are still extremely variable, but we can see times of greater snowfall: late January, mid to late February, late March



7-day chance of accumulation

- We look at snowfall for consistency: in what percentage of years did 7-day snowfall exceed 10 or 20 inches?
- Similar results with February shining more in this approach



P K M R S

Modeling for Alta Snowfall Analysis

M

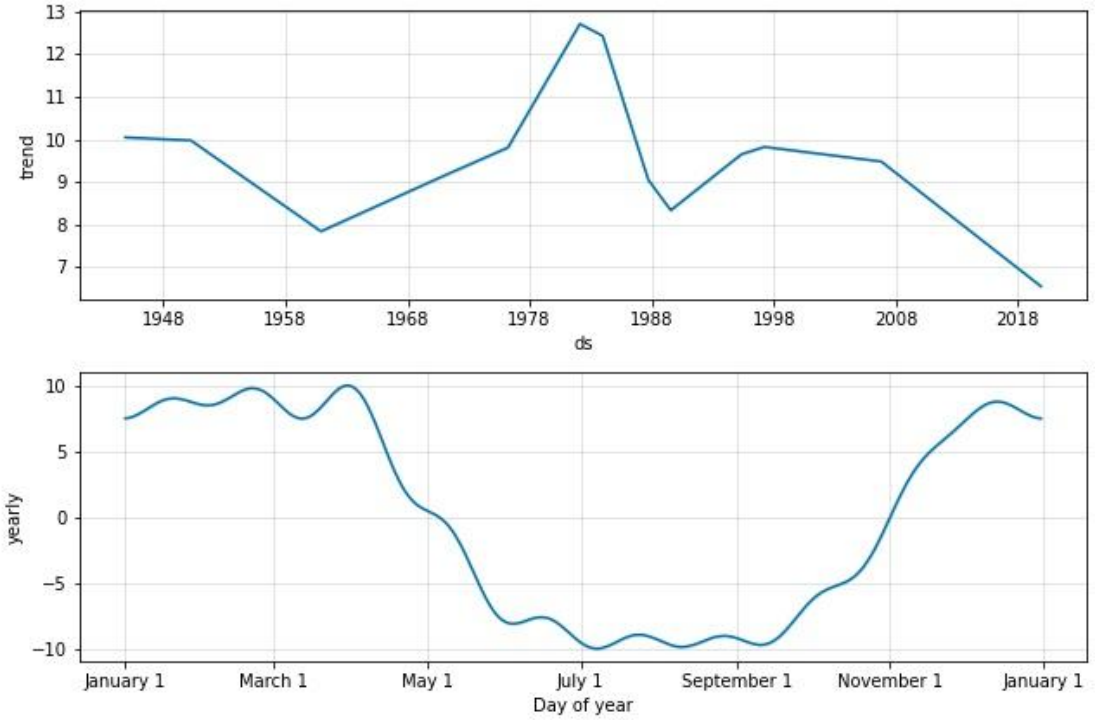
Since we have such variability in snowfall, we used the 7-day snowfall total as the target value in modeling and composed a time-series model using Facebook's Prophet.

Initial model results

- You can see the annual trend in snowfall over the years as well as the seasonal trend
- Dates that we highlighted before have higher snowfall in the model - as expected
- When tested versus last 3 years of snowfall, the model had a 0.02 in RMSE better than the dummy model of assuming the average
- Maybe we can improve these results by considering other features

7-day accumulation Prophet Model

- To get the expect snowfall sum the trend and yearly results



P K M R S

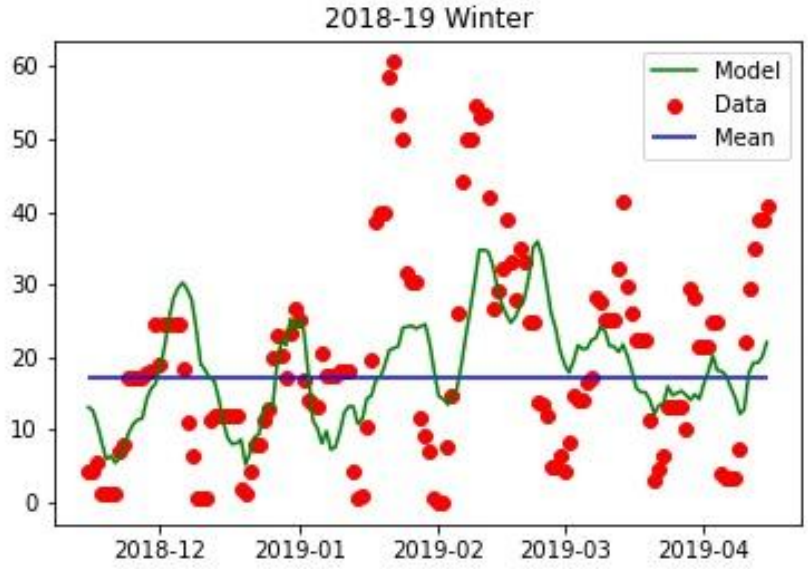
Modeling for Alta Snowfall Analysis

M

We also modeled using a Time Series plus Linear Regressor and Time Series plus a Random Forest Regressor; both of these produced much better results.

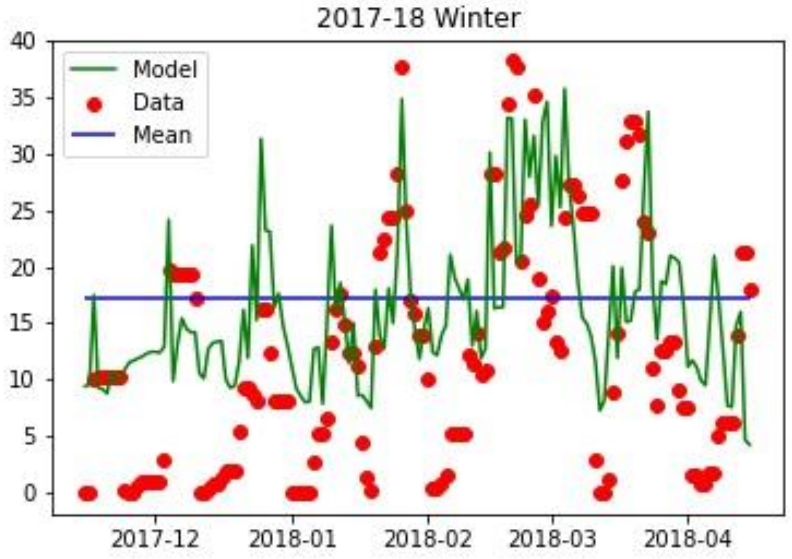
Time Series plus Regression

- **This improved model performance to an average of 11.07 RMSE over 3 years**



Time Series plus RF Regressor

- **Time series then Random Forest Regressor on the Residuals; RMSE of 10.78 in over 3 years**



P K M R S

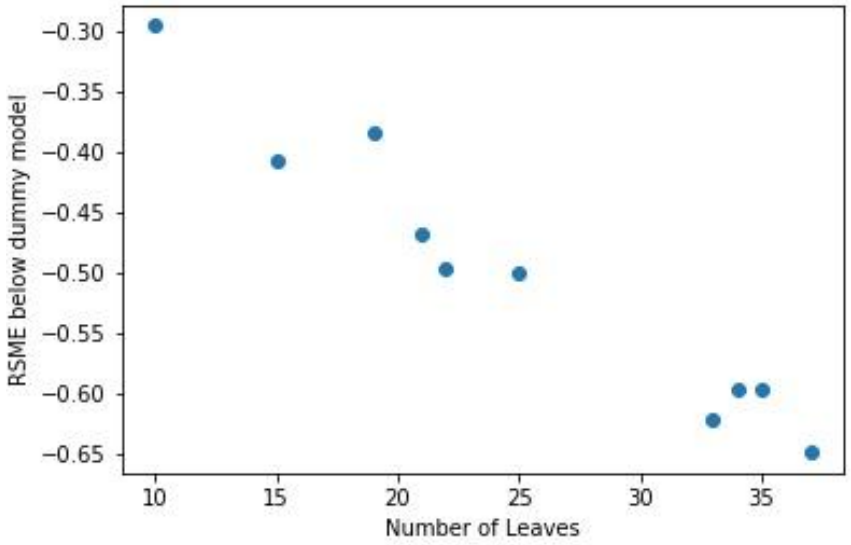
Modeling for Alta Snowfall Analysis

M

- **The previous model were great**
- **But relied on measurement (temperature and precipitation) available too close to the start of the model**
- **weather forecasts available by then**
- **we model again using just features available 30 days in advance**

Tuning the RF Regressor

Time series then Random Forest Regressor on the Residuals; "elbow" of 33 leaves looks the best



This long range model is better than the dummy, but not by much and in 2018-19, it performed worse.

Year	RMSE Model	RMSE Dummy
2016-17	16.00in	16.27in
2017-18	10.39in	11.38in
2018-19	14.68in	14.03in
Average RMSE	13.69in	13.90in

The "dummy" model of assuming the average is challenging to beat with long range models.

P K M R S

Recommendations for Skiers and Modelers

R

When should we ski Alta?!?
What action can we take to improve the model?

Best times to ski

- Differences are so slight
- Suggested times:
  - Late March (week of Mar 23-29 had snow every year; Mar 18 - April 9 performs well)
  - Mid-Late February (7- 28)
  - Late January (15-31)
- Only better than other dates slightly more than 50% of the time

Other potential model

Just based on historic: a skier select a week and we compare the rest of the season to that week to see in how many years other weeks have better snowfall

Improved Modeling Approach

- Since snowfall has an exponential distribution; model the log of snowfall
- Substitute values for zero snowfall to enrich the model
- Include Southern Oscillation Index (SOI) for La Nina / El Nino effects
- Include more lag variables
- Try different ensemble models
- Try other mountains with snowfall that varies more throughout their winter
- Model a larger window of snowfall accumulation
- Try this as a classification problem with “bins” of snowfall

Summary and Conclusion for Alta Snowfall Analysis

S

- Explored snowfall data for Alta
- Constructed a model for snowfall at Alta
- Model performed only slightly better than dummy (average)
- Ideas for future models

Data used in Model Building

- Snowfall data for Alta since 1944
- PDO and AMO monthly values
- Explored many aspects and effects on snowfall
- FB Prophet to model for time series
- Used both regression and RF Regressor added to time series

Model Performance

- Average RMSE: 13.69 inches
- Dummy RMSE: 13.90 inches
- Dummy (assume average all winter) is a pretty good model
- Would only depend on this model if no other contraints for vacation

Improved Approach Proposed

- Take log of snowfall
- substitute values for zero snowfall
- Include Southern Oscillation Index
- Try different lag variables (temp)
- Use different ensemble models
- Model a larger snowfall window (14 days instead of 7)
- Perhaps use classification to sort snowfall into “bins”

Special Thanks to

- Springboard TA’s
- Springboard Mentor, Raghu

P K M R S