Applying The Kalman Filter to SNOTEL Snow Depth Data

Authors: Joa Ladwig and Sandra Francisco
Sponsor: Mike Cooperstein, Colorado Avalanche Information Center

May 18, 2019

1 Abstract

The Colorardo Avalanche Information Center forecasts avalanches in 3 square kilometer regions, but they would like to forecast more locally in some areas. We apply data wrangling techniques and a Kalman filter to noisy snow depth data to create a finer resolution model of snow depth for usage in the SNOWPACK avalanche model. The cleaning methods replace impossible values, and the Kalman filtered data provide a starting point for further experimentation with SNOWPACK.

2 Introduction

The latest advancement in avalanche forecasting is using the SNOw TELemetry (SNOTEL) network input with the Snowpack model. In order for the Colorado Avalanche Information Center (CAIC) to use SNOTEL to its benefit, a few corrections need to be made. First the data collected by CAIC need to be put through an error-correction process. A model by Avanzi, et al. allows for the processing of SNOTEL data; however, it does not do any work to correct the errors in snow depth measurements. Our goal is to use a data assimilation method known as a Kalman Filter to decrease the error in these measurements. This will improve avalanche forecast correctness to some degree, which will benefit those who use CAIC forecasts in determining whether or not to travel into the back-country at some specific time. This can help reduce the amount of people caught in avalanches in Colorado and, hopefully, save lives.

Physical snow models are used widely for forecasting avalanches, floods, water availability, and more. Physical models like SNOWPACK use data like temperature, precipitation, relative humidity, wind speed, etc. to predict changes in energy and mass and therefore forecast future events [3]. SNOWPACK is currently the most used physical snow model, and improvements for using SNOWPACK focus on data manipulation rather than improving the model itself. Typically, modern users of SNOWPACK use SNOTEL data as input [2]. However, CAIC currently uses High Resolution Rapid Refresh, or HRRR, data from the National Oceanic and Atmospheric Administration (NOAA). The SNOTEL data are noisy and requires cleaning, as described by Avanzi, et al. [2].

However, even after cleaning and processing, the data need further improvement. As mentioned above, snow depth is one of the most important measures for running SNOWPACK and is prone to error. Reducing the error in the measured snow depth before inputting the data in to SNOWPACK will help SNOWPACK provide more accurate forecasts, which is of specific interest to the Colorado Avalanche Information Center, or CAIC. CAIC uses SNOWPACK forecasts to predict avalanche likelihoods in various regions and uses those predictions to warn those going into the backcountry and even those using the mountain-lined Interstate 70 [1]. Accurate predictions help reduce harm to people and property.

We are using a data assimilation method based on probabilities to reduce the error in snow depth measurements and improve the simulation accuracy. In 2006, Andrew G. Slater and Martyn P. Clark of the Cooperative Institute for Research in Environmental Sciences at University of Colorado Boulder used the Kalman filter (KF) to assimilate snow-water equivalence, SWE, data from SNOTEL sites [6]. This method produced optimal weighting between a modeled and observed state given estimates of the errors in the model and observations. [1]

[1] In addition to implementing the project's main Kalman filter algorithm, Joa performed and prepared all data wrangling, code reviews, and visualizations. Sandra researched the methods to be used, worked on a machine learning algorithm and particle filter algorithm that we opted not to use, corresponded with Coop for project updates, and kept the team on track to complete the project in time. Reports were worked on together.

3 Methods

The available snow depth data are hourly ultrasonic depth sensor readings at weather stations throughout Colorado. The sensor works by bouncing sound off of a surface and recording how long it takes to return. The sensor is shown in figure 1.

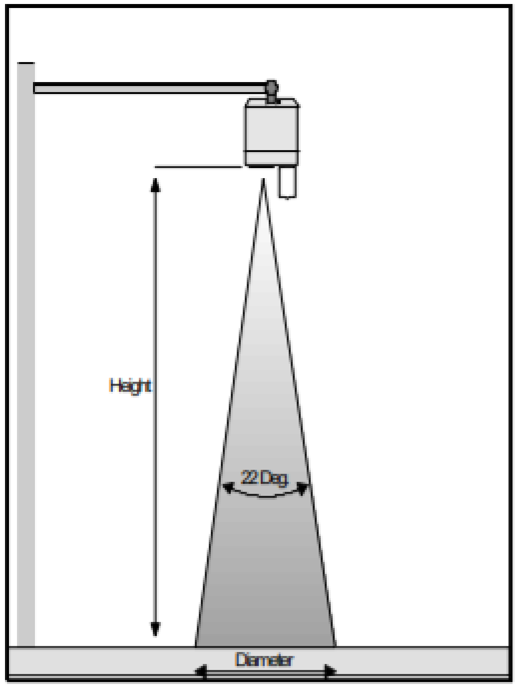

Figure 1: Visual of the Judd ultrasonic depth sensor used to measure snow depth at SNOTEL sites. [4]

These data date back as far as 2012, before which most stations only recorded daily values. SNOTEL has a website where these readings are recorded each hour, making it simple to obtain 24 values for each day for a given station. We can then build a data file over the snow season as each day passes. The data currently being used by the CAIC are from NOAA; This is known as HRRR, High-Resolution Rapid Refresh. According to our sponsor, Mike Cooperstein (hereafter Coop), the data from NOAA are less reliable than SNOTEL because the data are given for 3 kilometer regions rather than localized points. The variability of the measurements across such a large region are not typically accurate for the entire region, and thus the data are not particularly helpful in pinpointing regions that may have avalanche activity. The tools to use SNOTEL data have not been developed for use by CAIC yet. Our project aims to build at least some of those tools.

We used the following methods established by Avanzi et al. to correct transmission errors in hourly snow depth readings:

1. Filling in missing data - We forward-fill zero, negative, and Not-A-Number entries that are surrounded by non-zero entries with the previous value.

2. Removing data outside of the acceptable range - We replace entries that are greater than the determined maximum depth with the maximum depth.

3. Filter out data from Summer - Only dates between October and June are read and used.

4. Check for changes that are greater than $H_{\max}$ - We replace entries that have increased beyond the maximum hourly increase with the previous value.

We begin by checking for negative values and values greater than the maximum snow height, $H_{\max}$, which is determined separately for each site. Avanzi et al. also reduce any snow depth values recorded during summer to zero, which we have decided is unnecessary. Our model will only be run from October to June, a time-frame with no summer values, so the summer measurements will simply be filtered out. The ultrasonic sensors are prone to fluctuations between hours (due to high wind, low density snow, etc.), so our next step is to validate those fluctuations and eliminate them if erroneous. To do this, we first calculate the hourly difference in snow depth between two successive readings. Then we compare that incremental value to the hourly difference of the previous two readings; if the previous hourly difference is equal in absolute value, we assume the snow depth did not actually change and can thus replace that hour's value with the one before. Additionally, we have set a threshold for the maximum change in height between hours, ΔH_{MAX}, to be 0.6 meters. We leave the decision to increase or decrease this value to Coop's expertise. Once these processes are complete, the daily data are appended to the existing annual data and saved to be used in the following steps.

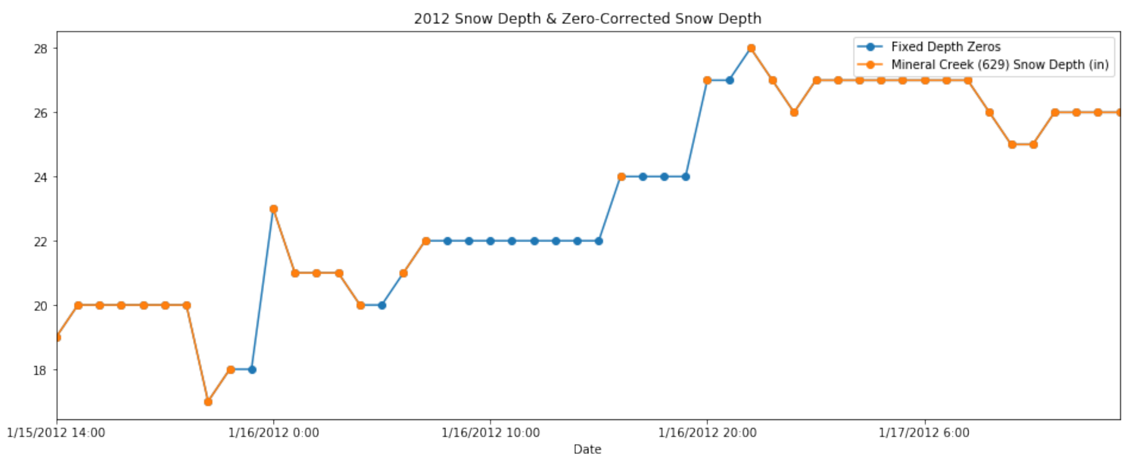

Figure 2: Example of cleaned snow depth data for three days, January 2012.

The Kalman Filter is a recursive solution to the linear filtering of discrete data; it uses the linear stochastic difference equation to estimate the state, x, of a discrete-time controlled system. The algorithm takes five steps, show in Figure 3 below3, to complete an iteration:

1. First, from the chosen starting point and error, the state is projected ahead by adding the previous state value to the control input.

2. Then the error covariance is also projected ahead by adding the process covariance, Q, to the previous error covariance.

3. Next we compute the Kalman Gain, K_k, by multiplying the projected error covariance by the inverse of the error covariance plus the measurement error, R, like so: $K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$.

4. then we update the estimate with the observed measurement, z_k, by adding the projected state the the Kalman Gain multiplied by the change in the observation.

5. the final step is to update the error covariance by multiplying the previous P_k^- by the identity minus the Kalman Gain, $P_k = (I - K_k H)P_k^-$.

See the table below for a list of variables for the filter:

k	Current time index
$\hat{x}_k^-$	Projected system state
A	State-transition Matrix
B	Control-input Matrix
P_k^-	Projected error covariance
K_k	Kalman Gain
H	Transformation Matrix
z_k	Snow Depth Observation Vector
Q	Process Noise Covariance
R	Observation Noise Covariance

In our case, the matrices A, B, and H are the identity matrix due to the simplicity of our problem. While A and B are likely to stay constant, future iterations could use further analysis to arrive at a non-identity H based on other conditions. We can establish R from the known tolerance of the sensor being used.

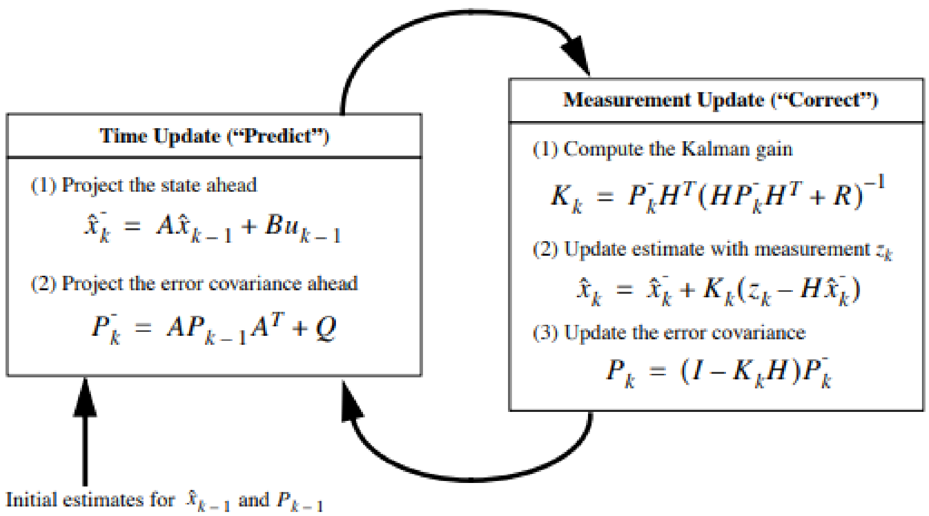

Figure 3: Steps of the Kalman Filter Algorithm
[7]

4 Results

The main result of our project is the hourly filtered depth data and the ability to run SNOWPACK on a significantly more local scale. Since the data are all from a specific site, the CAIC can build more specific, regional models of snow layers closer to areas of interest; this was more challenging before because the snow depth data from HRRR are at a 3km resolution, making it difficult to know the snow depth at a specific point within that region. Using a covariance of $Q = 0.1$, the result of applying the Kalman Filter is shown in figure 4.

We are working with Coop and his colleagues at the CAIC to see how our data compare to HRRR. We explored how the data assimilation using KF is affected by the choice of process covariance Q. In figure 5, we visually compare the data provided by the KF while varying Q. This image shows the algorithm's increasing hesitation

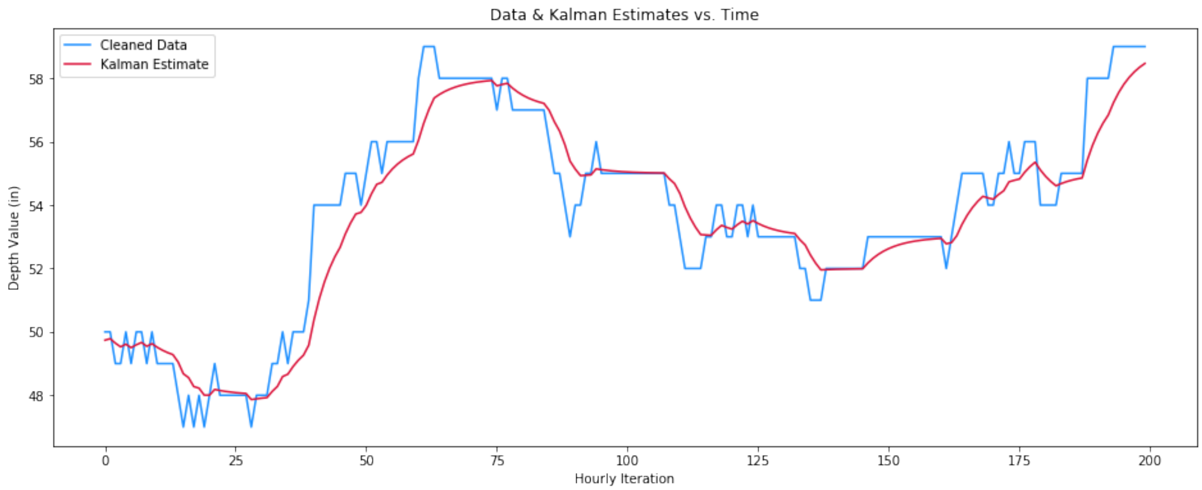

Figure 4: Applying the KF with $Q = 0.1$

to change the estimated state of the model as Q gets smaller. The "correct" Q value is elusive and will require more testing–perhaps on a per-site basis–to pin down; this is being worked on with Coop's colleagues.

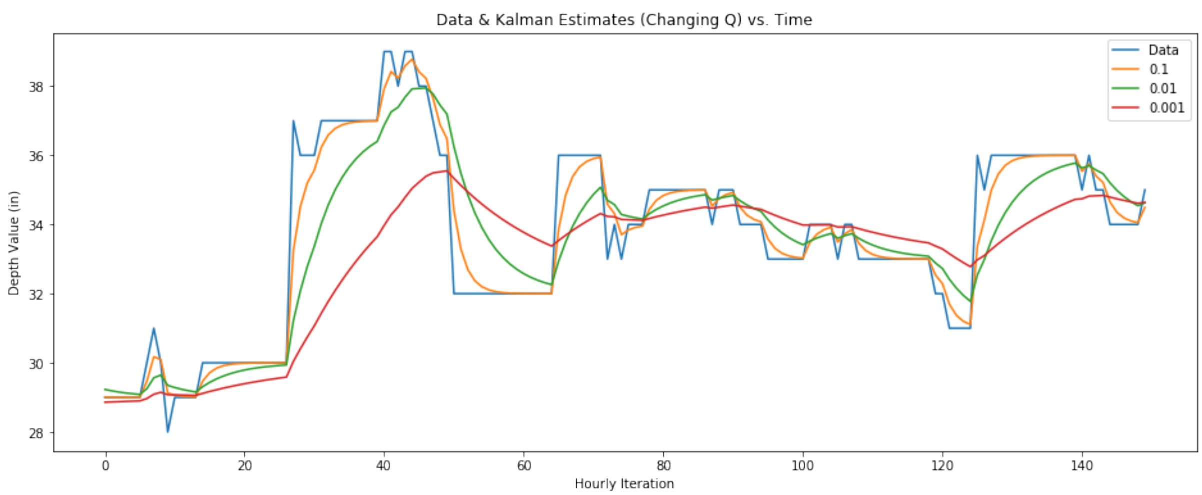

Figure 5: The adjusted data varies with process covariance, Q.

5 Discussion

When we began researching this project, all arrows pointed toward using machine learning to build a better model. As we got more familiar with the data, it became clear that we do not have nearly enough data points for machine learning to work. The estimated sample size we would need to achieve a 99 percentile confidence interval, if the population of avalanches is 25,000 and we allow margin of error of 5 percent, is 646[5]. Unfortunately we do not have 646 correct snow depth measurements to train a machine learning algorithm. The filter method we researched, Kalman Filtering, seemed to apply much more appropriately to our situation. In the end, the filter method allow us to build a more accurate, "living" model of snow depth at a weather station over the course of the snow year. Our work is certainly not a finished product, but rather a first step that can be used for further study and development. This project can be used as a starting point for others to expand on, such as an indexed process covariance, Q_k, analysis on the relationship between different SNOTEL sites, or further data validation through other variable observations at the same station (wind, air presssure, SWE, et cetera).

6 Conclusion

Even if the Kalman Estimates are entirely inaccurate and there isn't a way to fix them, we will still have made a tool CAIC can use to gather better snow depth data in a more refined manner than they are currently using. We think this will be relevant to our sponsor's work, but our results may be useful to other avalanche research groups as well. SNOTEL gathers data at thousands of weather stations across the continent, and an easy to use tool like ours could be helpful to many of them.

With more work on the process covariance and modeling of SNOTEL sites, our work could be expanded on to build a model of all SNOTEL sites within a given range; this would create a map of depths in an entire region that updates hourly, which could be used by CAIC to monitor extreme changes and alert them of potentially avalanche-triggering conditions.

References

[1] *About the CAIC*. URL: `https://avalanche.state.co.us/about-us/about-the-caic/`. (accessed: 03.29.2019).

[2] Francesco Avanzi et al. "A processing–modeling routine to use SNOTEL hourly data in snowpack dynamic models". In: *Advances in Water Resources* 73 (2014), pp. 16–29. DOI: `https://doi.org/10.1016/j.advwatres.2014.06.011`.

[3] Perry Bartelt and Michael Lehning. "A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model". In: *Cold Regions Science and Technology* 35 (3 2002), pp. 123–145. DOI: `https://doi.org/10.1016/S0165-232X(02)00074-5`.

[4] Judd Communications. *Judd Communications Depth Sensor*. URL: `https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdhttp://static1.1.sqspcdn.com/static/f/1146254/15414722/1322862508567/ds2manual.pdf?token=M60l3VwgYw7IJEW5Syn08azgrjE%3Df`.

[5] Qualtrics. *Sample Size Calculator*. URL: `https://www.qualtrics.com/blog/calculating-sample-size/`.

[6] Andrew G. Slater and Martyn P. Clark. "Snow Data Assimilation via an Ensemble Kalman Filter". In: *Journal of Hydrometeorology* 20.2 (2006), pp. 478–493. DOI: `https://doi.org/10.1175/JHM505.1`.

[7] Greg Welch and Gary Bishop. *An Introduction to the Kalman Filter*. URL: `https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf`.

A Required Files

1. kalmanSnowDepth.ipynb

2. WolfCreekSummit2018.csv

3. avanziExample.png