Identifying Trends in Forecaster Behavior Over 35 Years at the Northwest Avalanche Center

Chris C. Morin

Northwest Avalanche Center, Crystal Mountain, and the Northwest Avalanche Institute, Seattle, Wa

ABSTRACT: The Northwest Avalanche Center has forecasted weather and avalanche hazard for the PNW for over 35 years. They have produced almost 100,000 point forecasts, each of which is targeted towards an area adjacent to a weather station. This provides the unique opportunity to match each of these forecasts with the actual conditions observed. In this paper the biases by forecaster, changes over time, and bias introduced by corresponding weather patterns are identified. Auto-Correction of future forecasts using the statistical analysis of past biases is also discussed. This is the first time such a complete analysis has been able to take place of an avalanche center from birth to present day, and we hope the learnings uncovered will be an example of the power of data within avalanche forecasting in general.

KEYWORDS: Avalanche Forecasting, Forecast Accuracy, Weather Data, Weather Forecasts

1. INTRODUCTION

An accurate quantitative precipitation forecast is often the most important factor in forecasting avalanche stability over the next 24-48 hours. On the Northwest Avalanche Center's website, it is the most viewed metric on the most viewed page (Kenny Kramer, personal communication 2014). Recreational users apply the water amounts to their daily travel plans, while professional avalanche organizations in the Northwest schedule their workforce and resources around the amount and timing of the precipitation expected (Baugher, 1984).

A number of historical weather forecasting accuracy studies have been done, most focusing on model results, temperature, and the probability of precipitation. These are often done over short time frames, on the order of tens or hundreds of days (Wilks 1998, Mittermaier 2010, Mason 2008). Attempts to identify heuristic differences between forecasters is rare, as the data many times does not exist to visualize trends over time between participants. The few studies with actual forecasters that have been done (Wilks 1995) are often with participants that are forecasting for the study at hand. Also rare is defining the accuracy of quantitative precipitation forecast by human forecasters, as most quantitative precipitation forecasts are being produced directly from computer models.

The Northwest Avalanche Center has produced weather and avalanche forecasts since it's inception in 1976. Historical forecast data is available starting in the 1980-1981season. During this time over 4600 weather forecasts have been created, producing almost 100,000 snow water equivalent forecasts for ski areas and mountain passes. Each of these forecasts is in the unique position of being targeted at a weather station with a heated precipitation can. Further, a forecasters name has been attached to each forecast since 2000, and work logs can match a forecaster to a forecast starting in the year 1990. This provides a unique, rich dataset that can be used to identify trends over time, biases between forecasters, differences between areas being forecasted, and patterns between precipitation accuracy and other meteorological elements such as wind direction.

2. METHODOLOGY

Weather Forecasts from the Northwest Avalanche Center were produced in a flat text format from 1980-2010. From 2010-2014 they were saved within an html template. The forecasters name was included within the forecast starting in 2000. Each forecast contained a Snow Water Equivalent (SWE) forecast over the next 24 and 24-48 hours for a number of specific areas (fig. 1).

```
24 HOUR FORECAST OF PRECIPITATION IN INCHES OF WATER EQUIVALENT
(OR MM WATER--FOR BC SITES) ENDING AT 4AM
                                             TUE           WED

         MT BAKER                            .75           2
         STEVENS PASS                        .5            .75-1
         SNOQUALMIE PASS                     .5            1-1.5
         WHITE PASS                          .5            1
         CRYSTAL MTN                         .5-.75        1-1.5
         PARADISE                            1             2-3
         MT HOOD                             1             2-3
         HURRICANE RIDGE                     .5            .75
         MISSION RIDGE                       .25           .5
         WASHINGTON PASS                     .25-.5        .75

         BOSTON BAR, BC                      6             13
         COQUIHALLA SMT, BC                  6-13          19-25
         ALLISON PASS, BC                    6-13          19-25

           [LT = LESS THAN      NR0 = NEAR 0]
```

Fig. 1: Example text forecast for SWE

Formatting varied within each forecast but was consistent enough to computationally parse through 4600 forecasts and develop a table with the date of forecast, the forecaster's name (if given), the area being forecasted for, and the day being forecasted for. Often forecasts were given for a range (.5"-.75" etc.). When a range was given the average of the two numbers was used. If the forecast was for "less than .1 " or similar, the assumption that zero was the start of the range was used. This produced 87,697 Snow Water Equivalent forecasts, 43,916 of which had a forecasters name attached.

Actual precipitation and other meteorological sensor data was obtained from the saved datalogger files for each area being forecasted. Sensor data was available dependent on location as far back as 1988. Each datalogger collected hourly measurements from a variety of sensors specific to each weather station. Over 1000 files of logger data with over 50 million sensor readings were aggregated to be matched to the forecasts. Considerable difficulty was found in the weather data collection including the complete absence of column headers to identify what each number represented, missing data, subtotaling within the files, and shifting sensor locations within the data files changing the position of the readings attempting to be deciphered. This contributed to a reduction in the total number of forecasts able to be matched to actual precipitation. Enough data was found to confidently match 36,625 snow water equivalent forecasts to the logger data.

To match a portion of forecasts back to the start of forecasting in 1980, manual observations from Crystal Mountain were added to fill in the years previous to 1988 for this area. This added 1,340 data points to the weather record for a total of 37,965 matched forecasts.

3. ANALYSIS AND RESULTS

3.1 Overall Forecasts Accuracy and Bias

The linear response between forecasts and the water amount being forecasted for provides the most basic to measure the correlation between forecast an occurrence (fig. 2)

Averaging over 35 years, Forecasts for the 24 hour time period tended to bias to an under forecast for water amounts below .25". Above .25" an over forecast bias is present, returning to a null bias near a 1" forecast. The 24-48 hr forecast shows a tendency to under-forecast small water amounts, and a significant tendency to over forecast larger water amounts. Surprisingly, this systemic error is worse in the 24-48 hour forecasts created after the year 2000 than before. In general these systemic errors are less on the 24 hour forecasts prior to 2000.

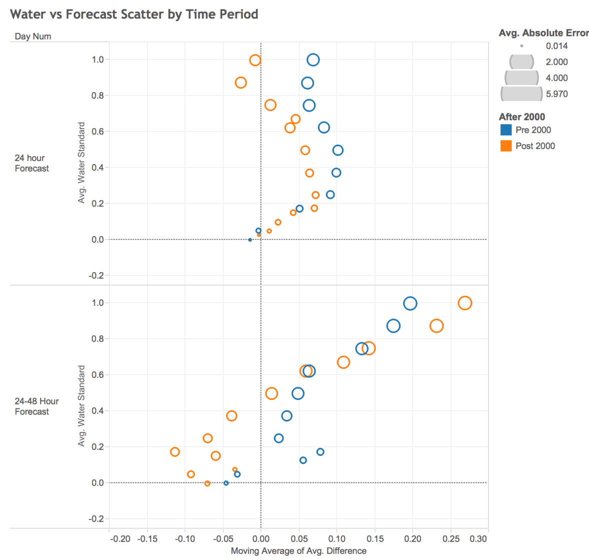

Fig. 2: Pre 2000 forecasts in blue, Post 2000 forecasts in orange. Points to right of center are over-forecasted, to the left are under-forecasted. Y-Axis is increasing SWE forecast.

Standard Deviations of forecasts increased with increasing forecasted water amounts, along with associated absolute errors. A 24hr forecast standard deviation for 0.01-0.25" SWE was .125", vs .68" for 1-1.25" SWE forecast. Figure 3 gives box plots and standard deviations of actual snow water equivalents for binned forecast amounts for 24hr and 24-48hr forecasts (fig. 3)

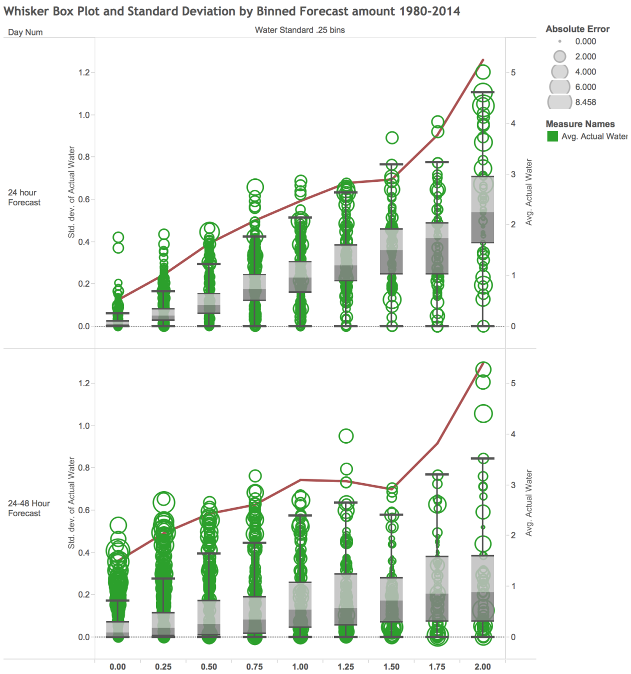

Fig. 3: Whisker Box plot with percentiles for increasing forecasted amounts vs observed totals. Red line is progressing Standard Deviation.

3.2 Forecast Evolution over Time

Mean absolute error increased during years of higher precipitation and lowered in drier years, but 24hr forecasts improved significantly over the time period studied. Mean absolute error decreased from an average of .25” prior to 2000, to .17” after 2010. 24-48 hour forecasts saw no improvement over time (fig. 4).

Grouping forecasts by water amount essentially normalizes the correlation between MAE and observed water totals. Figure 5 shows the obvious improvement across all one day out forecasts starting in 2004. Again, little to no improvement is seen in the 24-48hr forecasts.

3.3 Forecast vs Observed Precipitation and Wind Direction

Observed wind direction was matched to forecasts for Crystal Mountain and Mt Baker. Period of record for Mt Baker wind direction was 1988-2002. Period of record for Crystal Mountain wind direction was 1990-2009. For sake of simplicity both forecast days have been averaged in the wind study analysis.

Mt Hood Meadows saw significant over-forecasts from a 200-215 wind direction, averaging .15 -.3 above the observed SWE on 24hr forecasts and .3 above the observed SWE on 24-48hr forecasts. Severe under-forecasting occurred from Wind Direction 110-130, averaging .45 below the observed SWE on 24 hours forecasts and .03 below the observed SWE on 24-48hr forecasts (fig. 6).

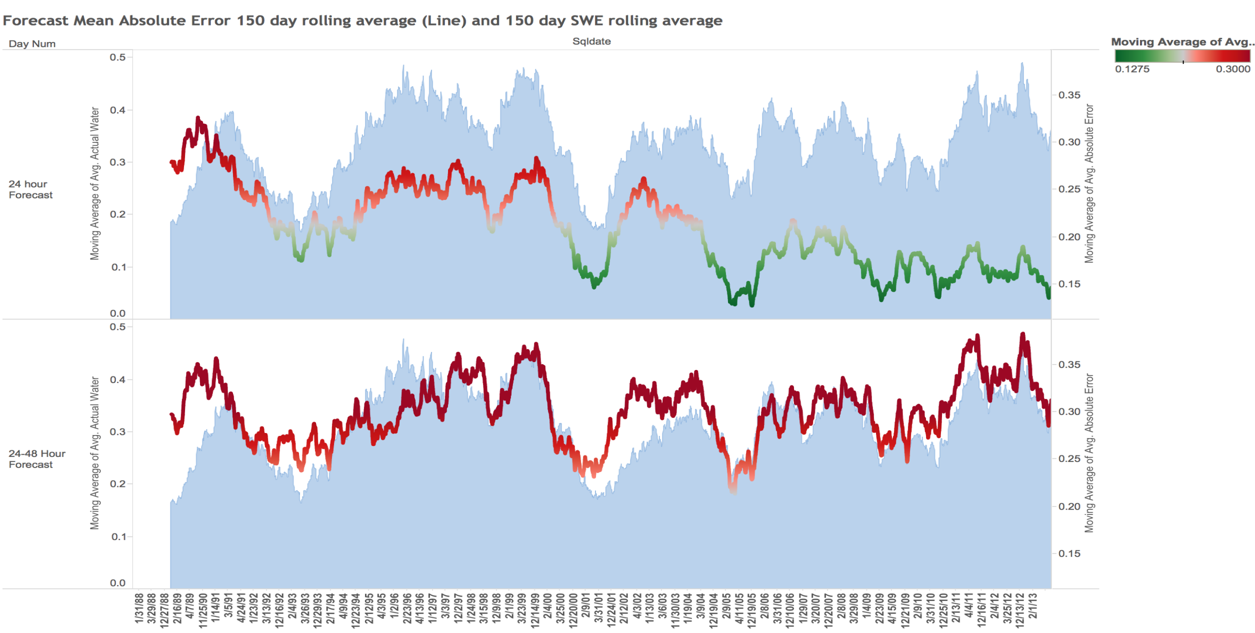

Fig. 4: Line is 150 day (approx. 1 season) moving average of Mean Absolute Error. Blue area is 150 day moving average of observed SWE by day.

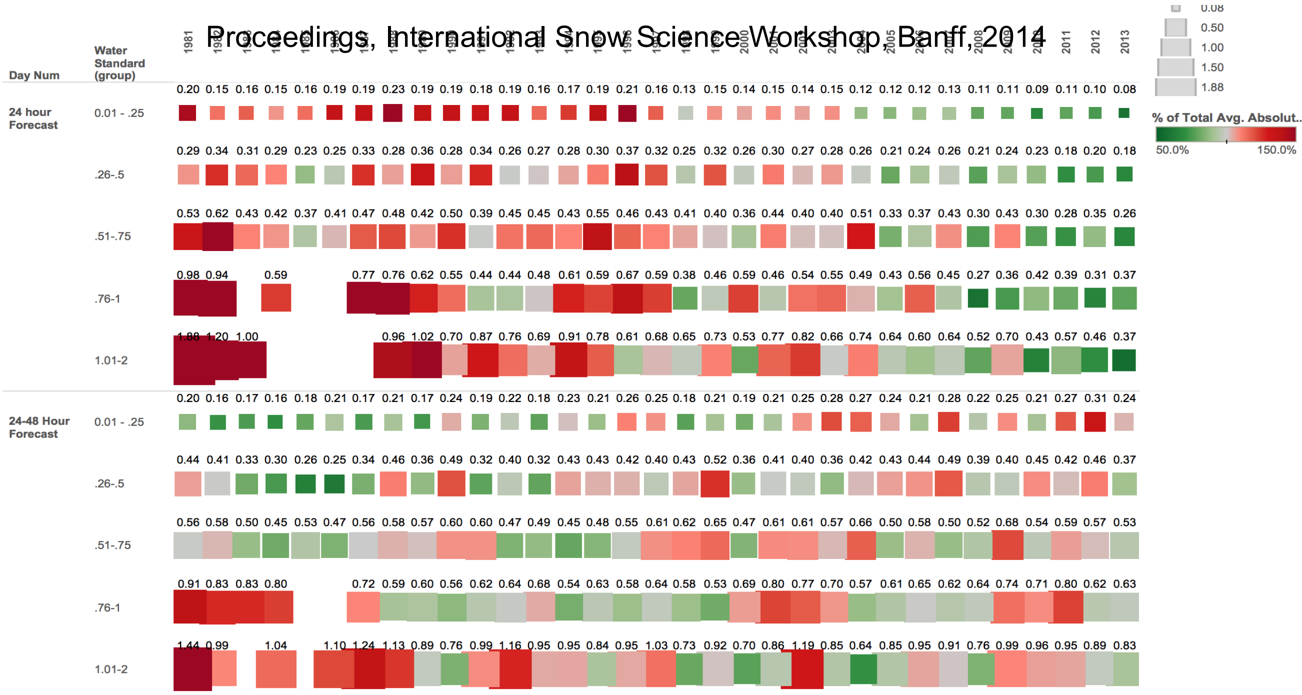

Fig. 5: Mean absolute error (Box Size, text, and MAE to mean within forecasted water group color) for grouped forecasted water amounts from .01-2”.

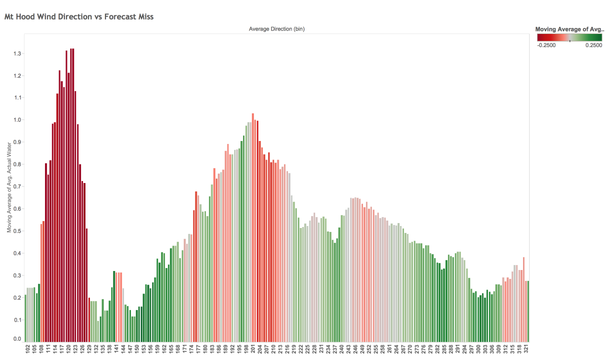

Fig. 6: Mt Hood wind direction vs average precipitation (bar) and average forecast error (color). Large red spike on L side is the under forecasted during an ESE wind direction.

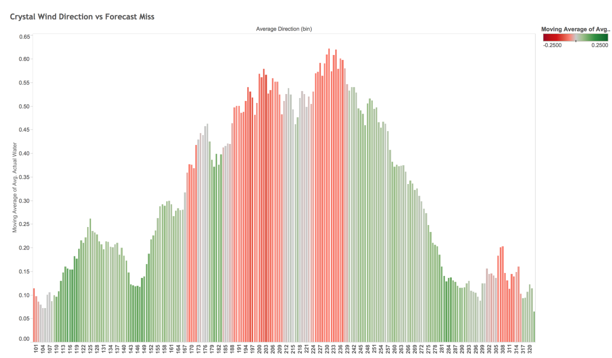

Fig. 7: Crystal Mountain wind direction vs average precipitation (bar) and average forecast error (color). Break in between red bars (under-forecast) is theorized due to the blocking action of Rainer.

Crystal Mountain forecasts from the 185-240 wind directions averaged -.05 under observed, with a break from 210-220 where forecasts showed little bias (fig. 7). This is likely due to forecasting a block from Mt Rainier over a wide wind direction for Crystal, but only having a sharp wind direction in which the block occurs.

3.4 Forecast Bias by Forecaster

All Forecasters over-forecasted water amounts when less than .5” on the 24hr forecast. Above a .5” forecast, Mark Moore strongly under-forecasted, Kenny Kramer slightly under-forecasted, and Garth Ferber over-forecasted. On 24-48hr forecasts, all forecasters under-forecasted when below .5”, and strongly over forecasted when above .5”. Garth showed the strongest over-forecast in this segment, while Mark Moore showed the weakest over-forecast (fig. 8).

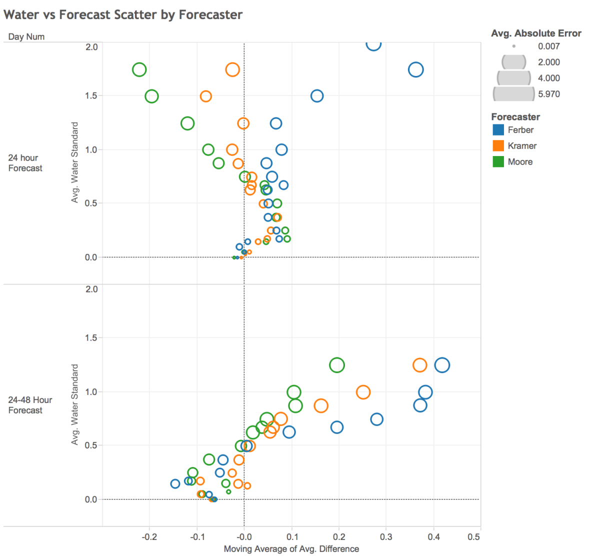

Fig 8: Forecast bias by forecaster. Plot is same as in Figure 2 but broken out by forecaster. Ferber-Blue, Kramer-Orange, Moore-Green.

3.5 Correction of Forecast using Bias Matrix

An attempt at auto-correcting a forecast based off a single forecasters bias was made. Mark Moore's 24-48hr forecasts from 1988-2005 were analyzed by forecasted water amounts, and table 1 was created showing the average bias by binned water amount forecasted.

Tbl. 1: Mark Moore's 24-48hr forecast bias from 1988-2005. Positive numbers are an over-forecast

Forecasted Precip., 0.25 bins	Bias
0.00	-0.0608
0.25	-0.1136
0.50	0.0453
0.75	0.0696
1.00	0.1605
1.25	0.3333
1.50	0.6495
1.75	0.7600

This forecast bias was then subtracted from each of Mark Moore's 24-48hr forecasts from 2006-2013. The resultant plot of uncorrected vs corrected forecasts is seen in Fig. 9.

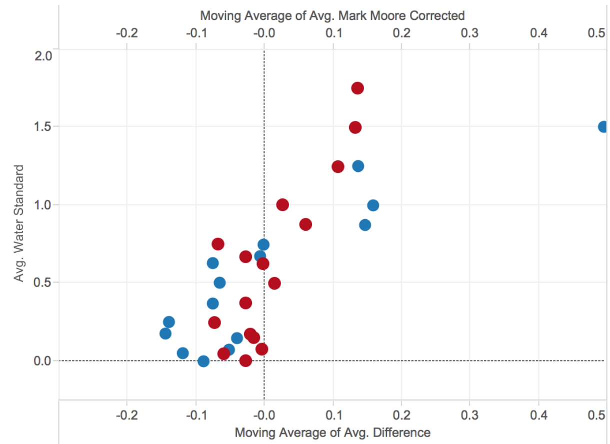

Fig. 9: Plot as in Fig. 2 and Fig 8. Mark Moore 24-48hr forecast bias from 2006-2013. Blue is uncorrected Forecasts. Red is corrected forecast. Obvious improvement is seen.

4.1 Conclusion

Using historical weather data it is possible to match Avalanche Weather forecasts with observed results. The specific forecasting system NWAC has developed provides a unique view into systemic biases that can be easily visualized using modern analytics packages. Some of these biases were already known but have been quantified for the first time. Southeast flow at Crystal has long known to "beat the forecast" by locals, but had not been previously numerically verified to do so.

Others are being understood for the first time. Little knowledge was found of the anomaly of heavy precipitation resulting from ESE flow at Mt Hood.

Still other long held perceptions are completely refuted when faced with a statistical analysis. Opinions garnered from a variety of audiences as to forecasters over or under forecasting often do not hold weight when analyzed.

The analysis confirmed the logical conclusion that 24hr precipitation forecasting has improved over the past 35 years. However, no significant changes in accuracy were observed in the 24-48hr forecast. With the amount of effort being put in to optimizing these forecasts, it should be asked if providing more general ranges and giving more time to other avalanche products would be a better allocation of forecaster time.

Personal forecaster development would be greatly aided by the utilization of a more strategic data model. Few if any avalanche organizations have aggregated all of there data into an easily usable format. Uncovering systemic errors can be exceedingly slow and inefficient if years of remembered antidotal evidence is all that can be relied on to improve forecasting accuracy.

Lastly, with the easily identifiable forecaster patterns, it is possible to make computer-generated improvements on the forecast using past biases. A rudimentary exercise was performed on Mark Moore's forecasts showing how a bias could be removed at an aggregate level. Further work needs to be done to refine an algorithm that can more uniformly improve precipitation forecasts. It is too early to make the assumption that an auto-corrected forecast would be a better forecast for the end user, as granular differences still need to be better understood and identified.

REFERENCES

Allan H. Murphy and Daniel S. Wilks, 1998: A case study of the use of statistical models in forecast verification: precipitation probability forecasts. *Wea. Forecasting*, **13**, 795–810.

S.J. Mason 2008: Understanding forecast verification statistics. Meteorol,15:,31-40

Hamill T M and and Daniel S. Wilks, 1995: A probabilistic forecast contest and the difficulty in assessing short-range forecast uncertainty. Wea. Forecasting, 10, 620-631

Paul Baugher, 1984: Crystal Mountain Avalanche Forecasting Program, 1-10.

Marion Mittermaier and Nigel Roberts, 2010: Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score. *Wea. Forecasting*, 25, 343–354.