ABSTRACT: The amount and timing of precipitation is one of the most difficult weather variables to forecast, especially in the Western United States. Safe and efficient operations of transportation organizations, emergency services, and avalanche forecasting rely heavily on accurate precipitation forecasts. Meteorologists use a variety of tools to help determine the amount, timing and type of precipitation but still struggle to produce accurate forecasts, especially in mountainous areas. Two new experimental ensemble forecast systems were studied during this project. The North American Ensemble Forecast System (NAEFS) technique is a high-resolution forecast produced by the University of Utah based on downscaling the low resolution North American Ensemble Forecast System forecasts with a climatological high resolution precipitation analysis from PRISM. The National Center for Atmospheric Research (NCAR) ensemble is based on 10 forecasts produced by the WRF-ARW model at 3km horizontal grid spacing. The NCAR and NAEFS produce multiple ensemble members that represent the spread of possible QPF results. This project is an attempt to compare a human quantitative precipitation forecaster (QPF) to these two new ensemble QPF forecasts for three locations in the western United States, Snoqualmie Pass (SNO30), Central Sierra Snow Lab (CSSL) and the Collins Study Plot at Alta Ski Area (CLN). After the two-week study period was complete, statistical validation was performed on the gathered data using absolute error, hit and miss rate and Heidke skill score. The outcomes from the validation process indicate that ensemble models outperformed the human forecasters in most instances during the study period.

KEYWORDS: NCAR, NAEFS, QPF, Precipitation

1. INTRODUCTION

Modern day weather forecasters are increasingly dependent on numerical weather prediction (NWP) models as a forecasting tool. Numerous NWP models produce deterministic forecasts at 0z and 12z each day. Depending on the audience that a forecast is produced for, forecasters attempt to provide the type and timing of weather events in a clear and concise forecast.

One of the most difficult weather variables to accurately forecast for is precipitation amount. Quantitative Precipitation Forecasts (QPF) have been produced by the National Weather Service (NWS) as an output from an NWP and then adjusted by a human forecaster. This process has the potential for many different bias errors depending on which forecaster is adjusting the QPF. Forecasters also use climatology information to help them better understand the weather patterns that are favorable for precipitation in their forecast area.

Two new experimental ensemble forecast systems are currently being produced by National Center for Atmospheric Research (NCAR) and the North American Ensemble Forecast System (NAEFS) downscaling technique. Ensemble forecasts are produced by adjusting a model's initial conditions for multiple model iterations to produce a spread of possible results. This project is an attempt to compare a human QPF forecast to these two new ensemble QPF forecasts for three locations in the western United States, Snoqualmie Pass (SNO30), Central Sierra Snow Lab (CSSL) and the Collins Study Plot at Alta Ski Area (CLN).
2. CLIMATOLOGY

The three forecast locations chosen for this study represent the diverse snow climate regimes within the Western United States as well as having quality controlled precipitation data available.

2.1 SNOQUALMIE PASS (SNO30)

The Cascade Mountains are a north to south oriented range located in the Pacific North West region of the United States. The western slopes of the range are characterized by warm and mild temperatures due to the close proximity to the Pacific Ocean while the eastern side is much colder and drier due to the continental air mass from eastern Oregon and Washington (Steenburgh, Mass, & Ferguson, 1996)

Elevations range from sea level to the west, up to the mountain passes that are below 1000 meters and high peaks up to around 3,000 meters. The annual amount of precipitation is between 60 and 100". (P. S. Hayes, L. A. Rasmussen, 2002) describes that the majority of precipitation falls during the winter months from land falling extra tropical cyclones. South westerly is the favorable flow direction in this area of the Cascade Mountains and generally produces higher amounts of precipitation. The Olympic Mountains to the west of the Cascades will block and divert the flow of moisture away from the Cascade Mountains in a west and northwest flow.

The differencing air mass characteristics between the east and western slopes of the Cascades have a lot of influence on precipitation type. The cross barrier temperature gradients can cause cold air to be pulled up through the lower elevation passes in the cascades, causing a dramatic change in freezing level (Steenburgh et al., 1996). In addition to being in a rain shadow, areas of the eastern cascades where the freezing levels are lower generally receive less precipitation than the western slopes.

2.2 CENTRAL SIERRA SNOW LAB (CSSL)

CSSL is positioned in the northern Sierra Nevada Mountain range within the Yuba River basin and Tahoe National Forest. The Sierra Nevada Mountains are oriented in a north/south direction with elevation gradients from close to sea-level up to 4000 meters. CSSL is located at 2,118 meters. Its proximity to the Pacific Ocean causes the Sierra Nevada Mountains to fall into a maritime type snow climate and can produce large variations in new snow densities. The average annual precipitation for CSSL between 1931 – 2006 is 59" and 397" of snow.

South Westerly flow patterns result in orographic enhanced precipitation events that produce the majority of the precipitation that falls in the Sierra Nevada Mountains. North Westerly flow directions result in a rain shadow at CSSL by the high peaks to the northwest.

2.3 ALTA – COLLINS (CLN)

CLN is located at an elevation of 2,945 meters and sits in the upper portion of Little Cottonwood Canyon, UT. The Central Wasatch Mountains are a North to South oriented mountain ranges while Little Cottonwood Canyon is a glacier formed canyon that is oriented on an east-west direction. This orientation significantly influences the location of the highest precipitation amounts within the Central Wasatch. CLN is most favored by a West through North West flow direction, while allows precipitation to fall in the upper portions of Little Cottonwood Canyon.

On average, CLN annually receives approximately 53" of precipitation and over 500" of snowfall. Little Cottonwood Canyon is well known for its low-density snow that is commonly between 5 and 9% snow water equivalent.
3. METHODOLOGY

A two-week forecast period was chosen, 1/24/16 to 2/6/16, during which four human forecasters rotated in producing a precipitation forecast for the three sites. The forecasters entered daily forecasts by 9:00am MST (1600 UTC) for six different forecasting periods as followed;

- Period 1: 18z Day 1 – 00z Day 2
- Period 2: 00z Day 2 – 12z Day 2
- Period 3: 12z Day 2 – 00z Day 3
- Period 4: 00z Day 3 – 00z Day 4
- Period 5: 00z Day 4 – 00z Day 5
- Period 6: 00z Day 5 – 00z Day 6

For each forecast period the forecaster would produce a precipitation forecast for each site that included a deterministic QPF amount (inches) and a range representing the middle 50% of the probability distribution (also in inches).

After the two-week forecasting period all observational data was collected for the entire six forecast periods. Using the observational data, statistical validation was performed for the Human Forecaster, NAEFS, and NCAR calculated precipitation forecasts. Absolute error ($\varepsilon_{\text{abs}}$) was calculated for each of the deterministic forecasts that were produced using the following equation;

$$\varepsilon_{\text{abs}} = |f_{\text{QPF}} - \sigma_{\text{QPF}}|$$

Where $f$ is the forecast QPF amount and $\sigma$ is the observed QPF amount. The absolute error was calculated for each of the six daily forecast periods for the Forecaster QPF, NAEFS Mean, NAEFS Median, NCAR Mean, and NCAR Median.

The second statistical value that was calculated was the overall Bias for each of the three deterministic forecasts. This Bias was calculated using the following formula;

$$\text{Bias} = \left( \frac{\bar{f}_{\text{QPF}}^2}{\bar{\sigma}_{\text{QPF}}^2} \right)$$

Where $\bar{f}$ is the mean of all the QPF forecast values for a specific forecast period and $\bar{\sigma}$ is the mean of all the observed QPF values for the same forecast period. The Bias was calculated for each of the six forecast periods for the forecaster QPF, NAEFS Mean, NAEFS Median, NCAR Mean, and NCAR Median.

The third value calculated was the Hit and Miss Rate of the Forecaster 50%, NAEFS 50%, and NCAR 50%. This was performed by counting the amount of occurrences that the observed QPF amount fell inside the forecasted 50% range. A separate Hit and Miss count was calculated for each location. A value is then calculated from the sum of the hits and misses from all six daily forecast periods over the entire two-week study period.

The fourth value calculated was the false alarm rate for the Forecaster, NAEFS, and NCAR. This was calculated by the simple formula;

$$\text{FAR} = \frac{\text{False Alarms}}{\text{hits + false alarms}}$$

Where false alarm is the number of occurrences where a precipitation event was forecasted but no precipitation was observed.
The Heidke skill score was calculated for each location. This skill score was calculated using the following formula.

\[
\frac{(hits + correct negatives) - (expected correct)_{random}}{N - (expected correct)_{random}}
\]

where

\[
(\text{expected correct})_{random} = \frac{1}{N} [(\text{hits} + \text{misses})(\text{hits} + \text{false alarms}) + (\text{correct negatives} + \text{misses})(\text{correct negatives} + \text{false alarms})]
\]

A positive Heidke skill score indicates that the forecast was more accurate as opposed to a forecast with a negative Heidke skill score.

Additionally, the four possible outcomes from the above equations are used to determine if a forecast is a hit or miss. Table A provides a visual representation of the four possible forecast outcomes.

<table>
<thead>
<tr>
<th>Outcome Description</th>
<th>Forecast &gt;.25 (in)</th>
<th>Forecast &lt;.25 (in)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit</td>
<td>False Alarm</td>
<td></td>
</tr>
<tr>
<td>Miss</td>
<td>Correct Negative</td>
<td></td>
</tr>
</tbody>
</table>

4. RESULTS

Due to the large spatial variability between forecast locations, each site was analyzed individually to help account for the unique climatology of each location. At the end of the two week forecast period, precipitation data was collected for the three forecast locations. Statistical analysis was completed using MATLAB software.

4.1 CLN

Statistically, the CLN site resulted to be the most difficult location to produce an accurate QPF for. The decreased skill scores for the forecast periods one through three is shown in Figure 1b.
The spread of forecast bias was very large for CLN. Additionally, the forecast bias from the mean increases dramatically at days 4 and 5 of the forecast period. To see the real improvement and value to forecasters at CLN, focus on the inner quartile range for forecast periods one through three. Table 2 shows these values, with a 50% value being the best forecast, indicating that the ensemble has a good handle on all of the possible outcomes.

Table 2. The hit rate percentage for the IQR for periods 1-3

<table>
<thead>
<tr>
<th>Forecaster</th>
<th>NAEFS</th>
<th>NCAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.905</td>
<td>61.905</td>
<td>52.778</td>
</tr>
</tbody>
</table>

In fact, the inner quartile range (IQR) actually got better for the later forecast periods, as show in Table 3.

Table 3. The hit rate percentage for the IQR for periods 4-6.

As described in the climatology section, CLN has an abundance of light density snowfall during the winter season. Trying to accurately predict not only the density of the snow and the associated amount of QPF is a difficult task, especially for the forecaster. With all that said, the NCAR produced much more accurate forecast relative to the other NWP and human forecasters during the study period.

4.2 CSSL

CSSL was a very consistent location, with model skill scores being within +/- .03. The bias by period was very consistent with deterministic NWP performance by increasing with time, as shown in Figure 2.
Figure 2. Bias by period and forecast for CSSL

The take away from Figure 2 is that all of the included models were consistently under forecasting the QPF. This also shows in the IQR forecast Table 4 shows the IQR for all the forecasts. Table 4 shows the hit rate is under 50% for forecast models and the forecasters, indicating that the QPF was consistently lower than the observed precipitation value.

<table>
<thead>
<tr>
<th>Forecaster</th>
<th>NAEFS Mean</th>
<th>NAEFS Median</th>
<th>NCAR Mean</th>
<th>NCAR Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periods 1-3</td>
<td>35.714</td>
<td>47.619</td>
<td>44.444</td>
<td></td>
</tr>
<tr>
<td>Periods 4-6</td>
<td>38.095</td>
<td>58.333</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Hit rate percentage for all IQR forecasts.

Another interesting result is shown in Tables 5 and 6. The ensemble models consistently under forecasted in periods 4-6 and as well as.

<table>
<thead>
<tr>
<th>Forecaster</th>
<th>NAEFS Mean</th>
<th>NAEFS Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLN</td>
<td>0.3253</td>
<td>0.37975</td>
</tr>
<tr>
<td>SNO30</td>
<td>0.64924</td>
<td>0.66329</td>
</tr>
<tr>
<td>CSSL</td>
<td>1.0</td>
<td>0.86916</td>
</tr>
</tbody>
</table>

Table 5. Heidke skill score for periods 4-6
The forecaster had a near perfect skill score in the long-range periods. This is likely due to the common mistake of under forecasting precipitation amounts by human forecasters.

4.3 SNO30

In terms of model skill score from Table 1, SNO30 was by far the most consistent location in the study. The skill score on varied by only +/- .03.
Figure 7. Average absolute error by forecast period for SNO30.

The average absolute errors were much higher than those of both CSSL and CLN. This means that while the deterministic forecasts were accurate at predicting when QPF would be over .25" they were however, not forecasting precipitation close to the observed amount. Table 6 is an example of the under forecasting that occurred for February 3. Each 24-hour forecast produced significantly less precipitation than was observed.

![Average Bias by Forecast Period: SNO30](image)

Table 6. An example of systematic under forecasting by the forecaster and NAEFS.

<table>
<thead>
<tr>
<th></th>
<th>Forecaster</th>
<th>NAEFS Median</th>
<th>NAEFS Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/30/16 Period 6</td>
<td>0.1</td>
<td>0.062</td>
<td>0.159</td>
</tr>
<tr>
<td>2/1/16 Period 5</td>
<td>0.295</td>
<td>0.06</td>
<td>0.162</td>
</tr>
<tr>
<td>2/2/16 Period 4</td>
<td>0.35</td>
<td>0.442</td>
<td>0.492</td>
</tr>
<tr>
<td>Observed</td>
<td>1.16999766</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7 demonstrates the large values of absolute error in both under and over forecasting for SNO30 in the extended forecast.

This correlates to what was seen in the extended forecast hit rate for the IQR, as shown in Table 7, which has values that are below 50%. That would indicate under forecasting or over forecasting are present for forecasters, NAEFS, and NCAR. The NCAR ensemble shows consistent under and over forecasting. Out of the 36 total forecasts, the model produced only 6 hits at SNO30.
5. CONCLUSIONS

Ensemble NWP models are becoming a very important tool for weather forecasters. Throughout the study, forecasters had the worst or second worst skill score for each location respectively. This is in spite of the fact that forecasters were looking at various numerical weather prediction models including the European Centre for Medium Range Forecasting Integrated Forecast System (ECMWF), High Resolution Rapid Refresh (HRR) and, the Canadian Model. The NCAR was a very skillful model at producing QPF with skill scores being either at the top or near the top for the median. As a whole, the NCAR median was a much better forecast than either the human forecaster or the NAEFS. The NAEFS produced a consistent forecast throughout the study period.

For CSSL, human forecasters could improve their forecast accuracy by using the bias values to correct their QPF accordingly. CLN produced very chaotic validation results, likely due to CLN being a difficult location to forecast for due to many different contributory factors. That said the NCAR clearly produced a more accurate QPF in the short-term periods than the other methods used in the study. SNO30 also proved to be a difficult location to produce an accurate QPF for. One compelling trend was the average of the long-term forecast settling on one value. The over and under forecasting didn’t show up in the bias because the two extremes averaged out to be close to zero.

Due to the relatively short time period that this study was carried out in, it is difficult to take the results of this study as truth. Increasing the study period as well as incorporating more detailed information about snow density, hourly precipitation rates, wind speed and direction could provide useful insights into the QPF validations.

Acknowledgments

We would like to thank Jim Steenburgh for his guidance and input throughout this project. WRCC and NCRS provided climatology data for the three forecast locations. The Desert Research Institute provided validation data for CSSL. Mesowest provided validation data for CLN and SNO30. NCEP and the Canadian Meteorological Centre provided the NAEFS data for the project. Craig Schwartz, and NCAR provided support with accessing the NCAR data. Lastly, we would like to thank The University of Utah Department of Atmospheric Science for support and funding of this project.

References

