A Review of PoP Reliability and PoP Bias Trends at WFO SLC for the Winter of 2007-08

Randy Graham
National Weather Service Salt Lake City, Utah

I. Introduction

Verification of National Weather Service (NWS) forecasts has long been a way to measure forecast skill as well as identify areas for possible improvement. Historically, National Weather Service Warning Forecast Office's (WFO's) have been limited to performing temperature and Probability of Precipitation (PoP) verification at a select number of sites within an office's County Warning Area (CWA). Numerous studies have examined the performance of NWS forecasts versus numerical weather predication guidance at locations across the country (Hughes 1979; Baars and Mass 2005). In addition to the more traditional verification techniques alternative approaches to temperature and PoP verification have also been considered, including statistical models (Murphy and Wilks 1998) and distribution-oriented approaches (Brooks and Doswell 1996). In recent years the National Weather Service has been issuing forecasts for 2.5 x 2.5 km or 5 x 5 km grid boxes covering their entire CWA as part of the National Digital Forecast Database (NDFD) (Glahn and Ruth 2003). With the advent of the NDFD vast quantities of forecast data is available opening the door for new verification measures and techniques that go beyond traditional site specific measures. The Salt Lake City WFO alone produces detailed forecasts for nearly thirty-one thousand 2.5 x 2.5 km grid boxes in its CWA. Traditional point verification techniques and measures are no longer adequate to assess the quality of this amount of data and new approaches are being examined.

In the winter of 2007-08' verification tools including BOIVerify (developed by Tim Barker) and a PoP/QPF Verification page (developed by NWS Western Region Headquarters) provided forecast offices the opportunity to examine their performance over all of the grids that are generated as part of the NDFD. The forecast PoP grids created at the WFO can be compared against a Quantitative Precipitation Estimate (QPE) analysis from the National Precipitation Verification Unit (NPVU) (McDonald et al 2000). These tools are affording offices the opportunity to examine much larger datasets than has been possible in the past. It should be noted that observations are not available for all of these forecast grids and therefore the forecasts are simply compared to the best available analysis such as the NPVU QPE for PoP and QPF forecasts.

With the ability to produce verification statistics for such large quantities of data one must consider what a 'good' forecast entails. Murphy (1993) described three basic categories of 'goodness' for forecasts. His three 'goodness' factors included, consistency, value (i.e., economic), and quality. This review will focus solely on the 'quality' of the of the Probability of Precipitation forecasts issued by the Salt Lake City WFO for the winter of 2007-08'. The goal of the analysis is to identify forecast trends and biases that may offer opportunities for improved forecasts or be worthy of further examination and study. The discussion of trends and biases, while based in the statistical methods available in the BOIverify software, will not be centered on the statistical significance of the identified trends and biases, but rather on a more subjective analysis of areas where the quality of NWS forecasts may be improved.

II. Methodology

The Salt Lake City CWA was sub-divided into five subsections for analysis of PoP reliability and PoP bias for the winter of 07-08'. The sections were demarcated based on regions of similar climatological characteristics utilizing National Weather Service zone forecast boundaries. The subsections (Fig 1) consisted of three valley groups (Northwest Valleys, Southwest Valleys, and Eastern Valleys) and two mountain groups (Northern Mountains and Southern Mountains).

Utilizing NCDC Reanalysis data (Kalnay et al 1996) each day from December 1st, 2007 through February 29th, 2008 was subjectively categorized based on the large scale pattern across the western United States. The days were divided into five categories: Southwest Flow (15 days), Northwest/North Flow (16 days; hereafter referred to as Northwest Flow), Trough (21 days), Ridge (33 days), or Split-flow/Cut-off (6 days). The Ridge days will not be reviewed in this analysis as they are generally associated with low probabilities of precipitation across Utah. Split-flow/Cut-off days will not be analyzed as the number of days in this category (six) precludes a meaningful analysis. Therefore, the review will focus on days that were categorized as Southwest Flow (Fig 2), Northwest Flow (Fig 3), or Trough (Fig 4).

After subjectively binning each day into one of the five aforementioned pattern types, PoP reliability statistics were calculated for each of the five subsections of the CWA for each pattern type. For the purposes of this review, the phrase 'PoP reliability' refers to the correspondence of forecast PoPs at gridpoints in the SLC CWA versus the frequency of analyzed precipitation at those same gridpoints in the NPVU QPE analysis. A perfect PoP reliability score occurs when the frequency of analyzed precipitation matches the PoP in question (i.e., a 40% PoP corresponds to analyzed QPE 40% of the time). The PoP reliability statistics were calculated for forecasts from the first period through Day 7 (168 hours) verifying on the day that the specific pattern type was observed. Although the data was analyzed for all seven days the majority of the discussion will focus on Pop reliability through Day 5. The PoP reliability statistics were then examined to determine if the WFO exhibited identifiable biases and trends, by both section of the CWA and pattern type, which could then be utilized to improve future forecasts. In addition to the PoP reliability statistics, PoP bias grids were created and the geographic distribution of the PoP biases will be briefly discussed. Throughout this review the terms 'slight chance', 'chance', 'likely', and 'categorical' will be utilized. These terms are associated with the NWS traditional PoP breakdown where 'slight chance' refers to PoPs of 15 to 24%, 'chance' equals PoPs of 25 to 54%, 'likely' refers to PoPs of 55 to 74% and 'categorical' refers to PoPs of 75% or greater.

III. Results

The PoP reliability and PoP bias were examined for each region stratified by the observed pattern type. However, not all pattern types or regions will be discussed in this review. Instead the focus of this analysis will be on identifiable trends and significant biases that may be of some use for forecasters or which may be worthy of additional exploration. While some notable trends and biases will be discussed, it should be understood that the analysis of PoP reliability and PoP bias revealed that the WFO provided excellent distribution of PoPs in most of the sub-region and pattern type combinations.

A. Northwest Valleys

Several identifiable, and potentially useful, trends were noted for Utah's northwest valleys. During Southwest Flow events there was tendency for the WFO to overuse slight chance and chance PoPs and to underutilize likely or greater PoPs in the short term portion of the forecast. As an example, for gridpoints where the WFO forecast PoPs were 20 to 40%, the NPVU QPE analysis had measurable precipitation for these same grid points only 10 to 15% of the time for forecasts with a 36 hour lead time (Fig 5). This is an indication that the office had a tendency to overcast chance PoPs on southwest flow across the valleys of northwest Utah. Conversely, when the WFO forecast PoPs of 50 to 70% the frequency of analyzed precipitation ran 15 to 25% above the forecast PoP values. These trends were consistent for forecast with lead times of 12 to 48 hours.

For Trough events the PoP reliability of WFO forecasts was very good through the first 48 hours. Beginning around the 60 hour forecast there was a tendency for the WFO to underutilize PoPs of 50% or greater (Fig 6). This became more apparent the further the forecast extended into the medium range and it was not uncommon to see the forecasts 'underforecast' PoPs by about 20% from 60 hours on out. This trend is also evident in the Day 4 PoP bias across the northwest valleys (Fig 7). While the PoP reliability indicates trends at specific PoP values and time ranges, the PoP bias highlights potentially consistent trends across the range of PoPs. The PoP bias for Trough events at Day 4 indicated that the WFO had a tendency to underdo PoPs across Utah's northwest valleys, particularly in the Cache Valley, Tooele Valley, and the central and southern Wasatch Front. The trends identified in both the PoP reliability and PoP bias for Trough events across Northwest Valleys indicate that forecasters could be more a bit more aggressive with PoPs, particularly along the Wasatch Front, in the medium range given a sufficient level of confidence in the forecast.

The SLC WFO is traditionally more aggressive with PoPs along the Wasatch Front in northern Utah for events that occur in northwest flow. Many locations along the Wasatch Front receive a favorable orographic contribution with events that occur in northwest flow and the tendency to be aggressive in these events was born out in the PoP reliability numbers. A look at the PoP reliability graphic for 12 hour forecasts (Fig 8) indicates that the WFO had a tendency to overuse chance and likely PoPs (by 10 to 20%). This trend was evident for forecasts valid from 12 through 48 hours. A closer examination of several Northwest Flow events indicated that on occasion this was actually a result of holding onto the chance PoPs a little too long at the end of events. Additional insight can be gleaned from an examination of the 12 hour PoP bias grid (Fig 9). The PoP bias grid indicates that the WFO had a high PoP bias across much of the northwest deserts, Cache Valley, and the northern Wasatch Front. The distribution of this positive PoP bias was consistent through 48 hours.

B. Northern Mountains

For Trough events WFO SLC had a tendency to overuse chance PoPs (by 10 to 20%) in the first 48 hours for the Northern Mountains (Fig 10). There are several ways that this could be interpreted, but investigation of individual events indicated that this may have been partially a result of beginning the chance PoPs in the northern mountains a little too soon. If the office was not confident enough to go with at least likely PoPs there was a tendency to overuse the chance PoPs. An examination of the PoP bias grids (not shown) revealed that there was a positive bias across the Uinta Mountains and the far northern Wasatch Mountains. This bias was persistent through the first 48 hours and likely played a role in the PoP reliability statistics. It should be noted that the office had excellent PoP reliability for likely or greater PoPs in the first 48 hours of the forecast.

Another significant trend that was noted in the northern mountains was that the WFO tended to under utilize likely or greater PoPs (by 10 to 20%) beyond 60 hours in both Trough (not shown) and Northwest Flow events (Fig 11). Correspondingly, a negative PoP bias is evident across much of the northern mountain region in Day 4 (Trough example, Fig 7). The WFO PoP reliability was very good in the northern mountains in the first 48 hours. The fact that the PoP reliability was very good in the first two days of the forecast with a low PoP bias for 60 hours and after for both Trough and Northwest Flow events may indicate the opportunity for a more aggressive approach in the medium range period of the forecast under certain conditions. It is certainly easier to go with high PoPs in the short term portion of the forecast where confidence is traditionally higher. However, in some events, when model ensemble spread is low for example and several model runs have persisted with the same solution, forecasters may be more aggressive in the Day 3 through Day 5 time frame. For WFO SLC this may particularly true for the northern mountains in significant Trough and Northwest Flow events.

C. Southwest Valleys

The strongest pattern type signal for the southwest valleys occurred during Southwest Flow events. There was a high PoP bias in the first 48 hours for slight chance and chance PoP values across the southwest valleys. Over the first 48 hours the slight chance and chance PoPs tended to be over utilized by 5 to 20%. Figure 12 is an example of this high PoP bias and is valid for forecasts with a lead time of 36 hours. This trend was not evident in either the Trough or Northwest Flow events (where PoP reliability was very good), but was consistent for Southwest Flow events. The fact that this was isolated to Southwest Flow events indicates that it may be a more systematic bias related to a single pattern type.

D. Southern Mountains

Several important trends were noted for forecasts for Utah's southern mountains. The WFO had a tendency to over forecast PoPs for nearly all PoP values (generally, by 10 to 20%) for the southern mountains through the first 48 hours for Southwest Flow events. Figure 13, valid for forecasts with a lead time of 24 hours, provides an example of this bias which was present through the first 48 hours of the forecast. An examination of the PoP bias for 24 hour forecasts in Southwest Flow events indicated that there was a strong high PoP bias across the northern half of the southern mountain sector (Fig 14). This trend mimicked an over forecast bias on Southwest Flow events for Utah's southwest valleys. There are limited precipitation measuring sites across southwest Utah and radar coverage is poor, so perhaps there is an issue with the analysis in this region. Conversely, this could be a result of the WFO being a little more aggressive than needed in Southwest Flow events across the southern portion of the state. It is interesting to note that the PoP reliability for the southern mountains in both Trough and Northwest Flow events was very good in the first 60 hours of the forecast. As an example, Figure 15 displays the PoP reliability for 48 hour forecasts in the southern mountains for Northwest Flow events. This discrepancy in the trends amongst the different pattern types could be an indication that the trend noted for the Southwest Flow events is not associated with analysis issues and may instead be a systematic bias that can be examined in further detail.

IV. Conclusions

The availability of new verification tools such as BOIverify and the Western Region PoP/QPF verification Page have provided WFOs the unprecedented capability to examine temperature, PoP and QPF verification data over their entire CWA. Utilizing these new capabilities, an analysis of PoP reliability and PoP bias was undertaken for five subsections of the SLC CWA for the winter of 2007-08'. The verification results were stratified by pattern type in an attempt to assess potential biases and office trends that could offer opportunities for future improvement.

Some of the more significant biases and trends that were identified include:

- Tendency to under utilize likely or greater PoPs from 60 hours on out in both the northwest valleys and northern mountains in Trough events.
- Overuse of slight chance and chance PoPs in the northwest valleys for Northwest Flow events. This corresponded with a high PoP bias across the northwest deserts in these same events.
- Overuse of slight chance and chance PoPs across the southwest valleys in Southwest Flow events.
- Tendency to have PoPs that were too high across the southern mountain sector in Southwest Flow events. A high PoP bias was particularly evident across the northern half of the southern mountain sub-region.

V. References

Baars, J.A., and C.F. Mass, 2005: Performance of National Weather Service Forecasts Compared to Operational, Consensus, and Weighted Model Output Statistics. *Wea. Forecasting*, **20**, 1034–1047.

Brooks, H. E., and C. A. Doswell III, 1996: A comparison of measures-oriented and distributions-oriented approaches to forecast verification. *Wea. Forecasting,* **11,** 288–303.

Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195-01.

Hughes, L.A., 1979: Precipitation Probability Forecasts–Problems Seen via a Comprehensive Verification. *Mon. Wea. Rev.*, **107**, 129–139.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

McDonald, B. E., T. M. Graziano, and C. K. Kluepfel, 2000: The NWS National QPF Verification Program. Preprints, 15th Conference on Hydrology, Long Beach, CA, January 9-14, American Meteorological Society, p. 247-250.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting,* **8**, 281-293.

_______, and D.S. Wilks, 1998: A Case Study of the Use of Statistical Models in Forecast Verification: Precipitation Probability Forecasts. *Wea. Forecasting*, **13**, 795–810.

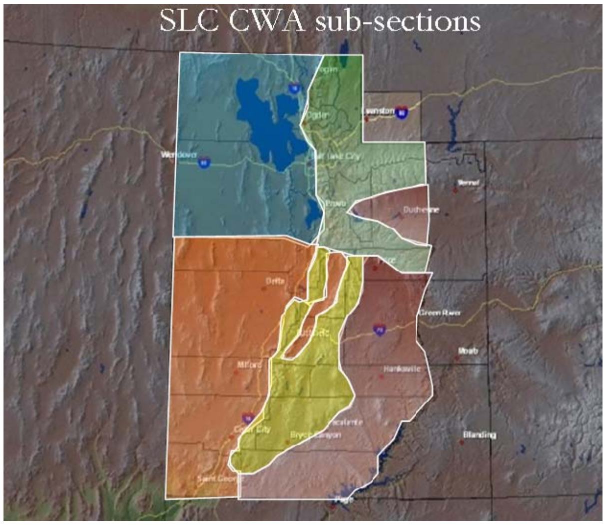


Figure 1 – Boundaries of the five subsections of the SLC CWA utilized in the review: Northwest valleys (blue), Southwest Valleys (orange), Northern Mountains (green), Southern Mountains (yellow), Eastern Valleys (red).

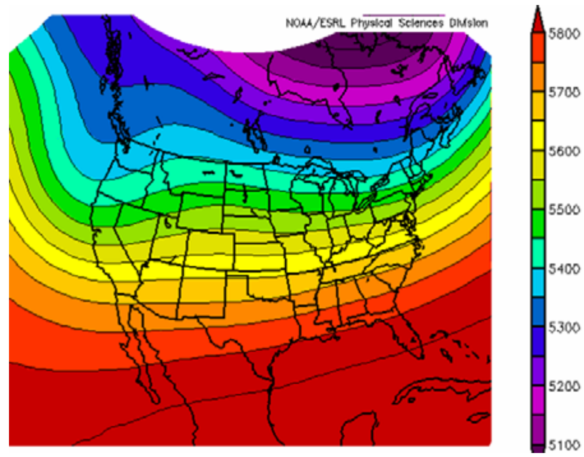


Figure 2 – Composite 500 hPa pattern for the fifteen Southwest Flow days included in analysis.

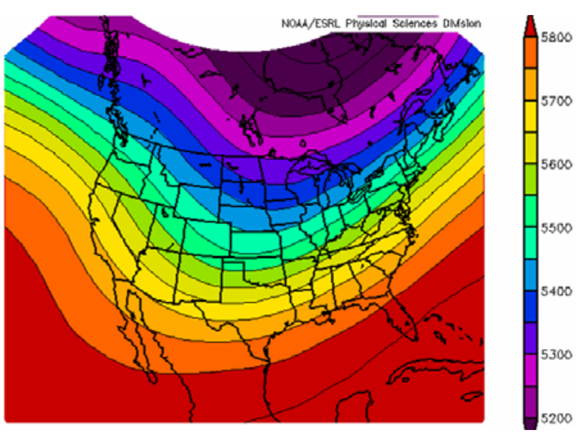


Figure 3 – Composite 500 hPa pattern for the sixteen Northwest Flow days included in the analysis.

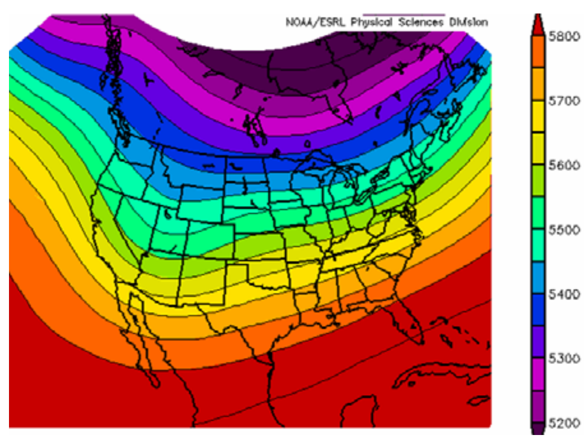
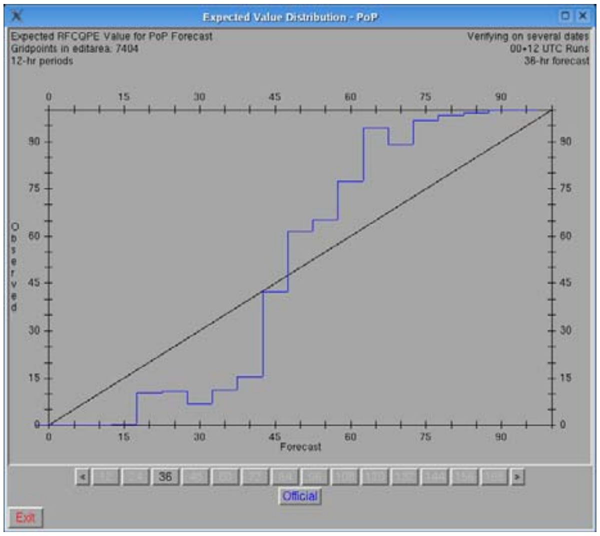


Figure 4 – Composite 500 hPa pattern for the twenty-one Trough days included in the analysis.



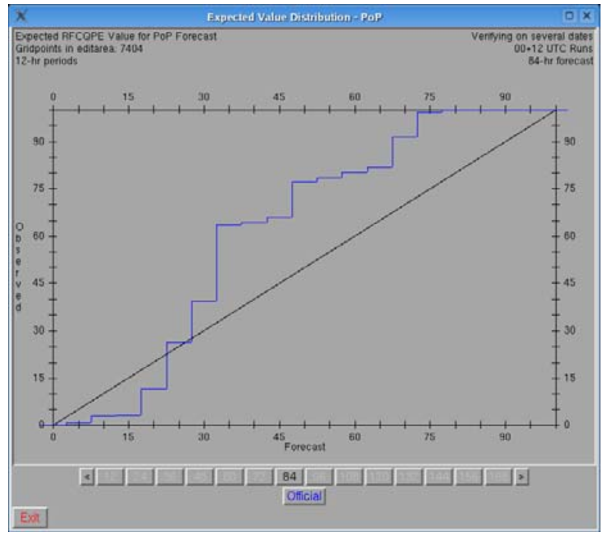
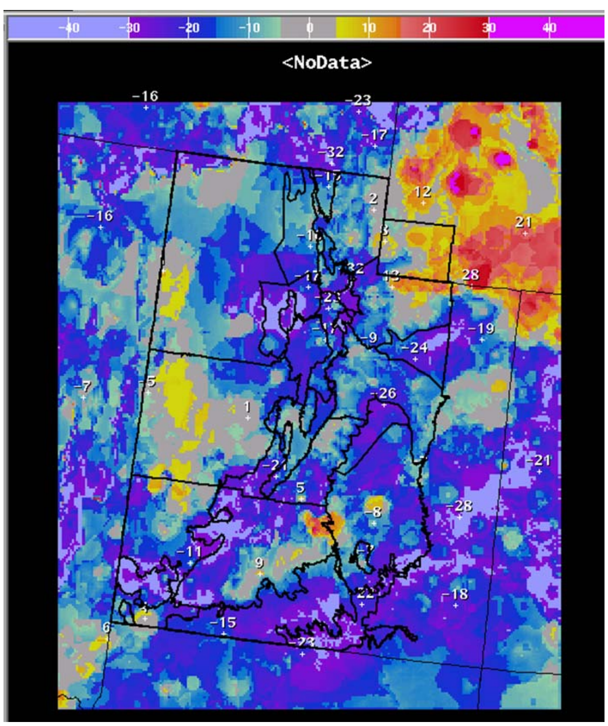


Fig 5 – PoP reliability table for 36 hour forecasts from WFO SLC for Southwest Flow events across Utah's northwest valleys.

Fig 6 – PoP reliability table for 84 hour forecasts from WFO SLC for Trough events across Utah's northwest valleys.



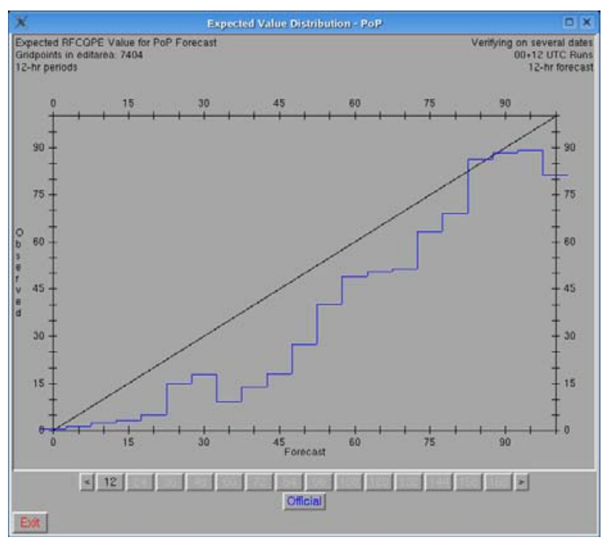


Fig 7 – PoP bias grid for Day 4 forecasts from WFO SLC for Trough events. Warm colors indicate positive PoP bias and cool colors indicate negative PoP bias.

Fig 8 – PoP reliability table for 12 hour forecasts from WFO SLC for Northwest Flow events across Utah's northwest valleys.

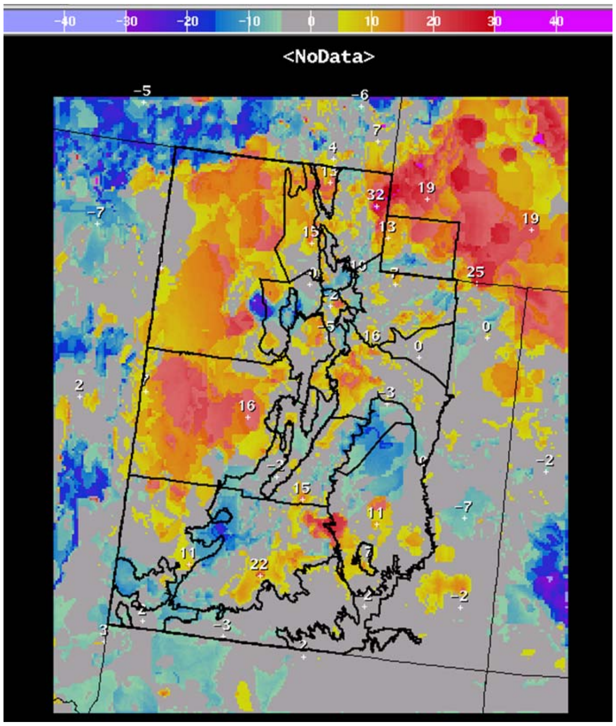


Fig 9 – Same as Figure 7 except for 12 hour forecasts in Northwest Flow events.

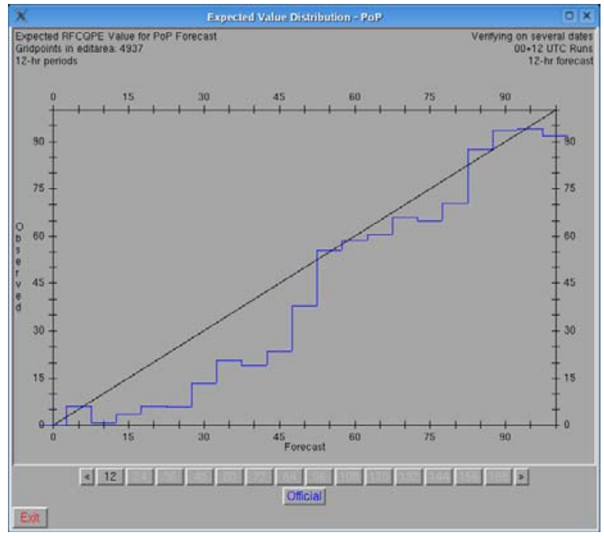


Fig10 – PoP reliability table for 12 hour forecasts from WFO SLC for Trough events across Utah's northern mountains.

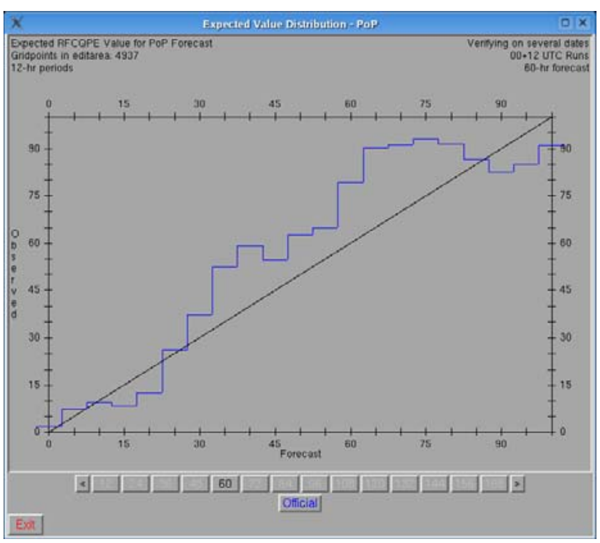


Fig 11 – PoP reliability table for 60 hour forecasts from WFO SLC for Northwest Flow events across Utah's northern mountains.

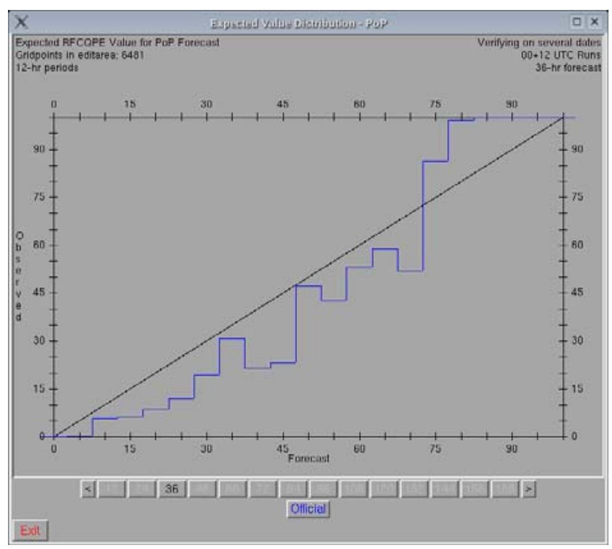
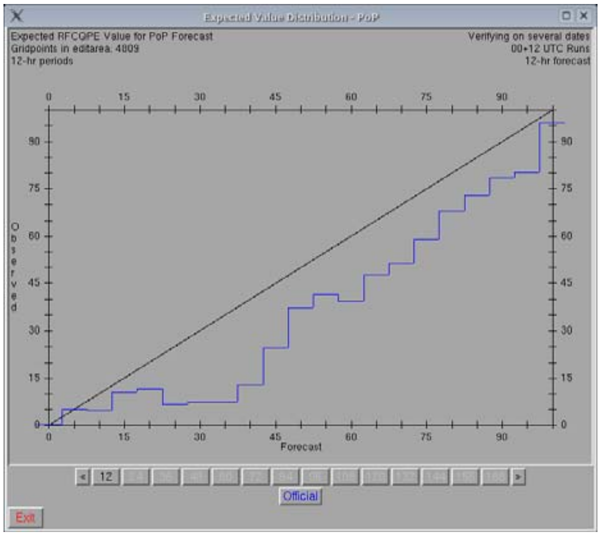


Fig 12 – PoP reliability table for 12 hour forecasts from WFO SLC for Southwest Flow events across Utah's southwest valleys.



.
Fig 13 – PoP reliability table for 24 hour forecasts from WFO SLC for Southwest Flow events across Utah's southern mountains.

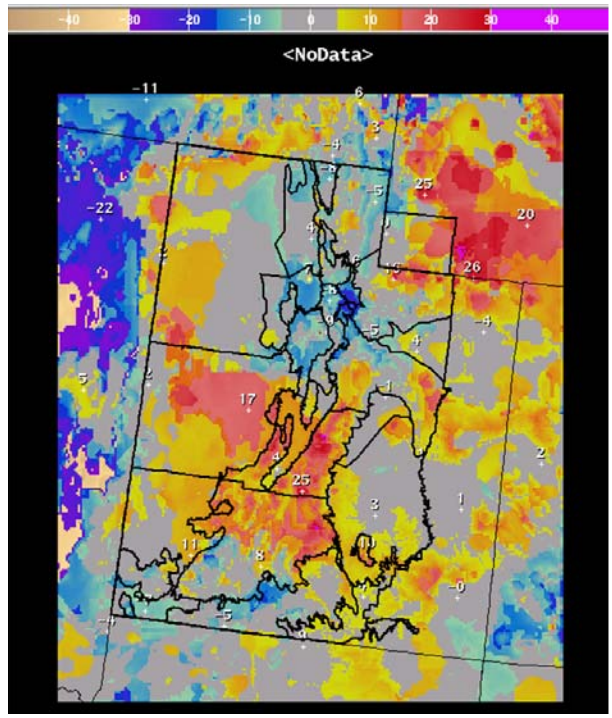


Fig 14 – Same as Figure 7 except for 24 hour forecast in Southwest Flow events.

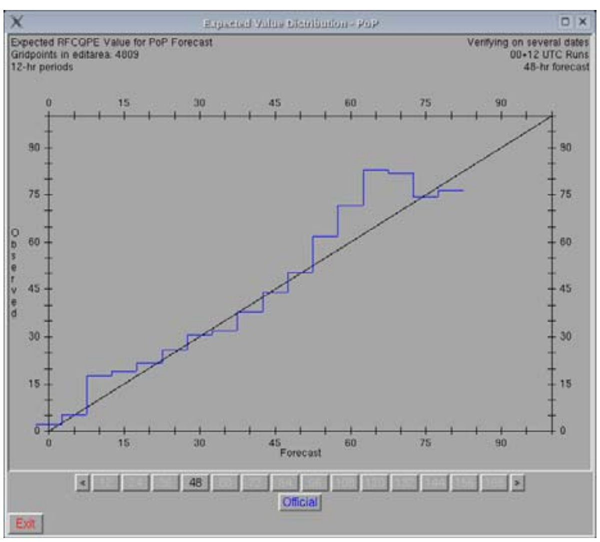


Fig 15 – PoP reliability table for 48 hour forecasts from WFO SLC for Northwest Flow events across Utah's southern mountains.