Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
39th NOAA Annual Climate Diagnostics and Prediction Workshop
St. Louis, MO, 20-23 October 2014

CPC's New Week-2 Probabilistic Hazards Forecast and Extremes Tool

Melissa Ou, Ken Pelman, Mike Charles, and Jon Gottschalck

Climate Prediction Center, NCEP/NWS/NOAA

1. Introduction

For the last 13 years, the Climate Prediction Center (CPC) has been issuing days 3-14 hazards forecasts daily. This is a manually produced forecast in deterministic format highlighting areas that have elevated risk of various hazards, such as wildfires, flooding, much above/below normal temperatures, heavy precipitation, etc. Model improvements in daily weather forecasts have enabled weather centers (such as the Weather Prediction Center, Storm Prediction Center, *etc*.) to issue national forecasts with longer lead times, mainly up to 7 days. Therefore, CPC is now focusing on improving the national hazards forecasts beyond week-1, from days 8-14, which will be referred to as week-2 in this article.

Beyond week-1, it is more useful and robust to present forecasts in a probabilistic format, rather than deterministic due to the greater uncertainty associated with models at longer lead times. CPC is now issuing experimental manual forecasts daily of hazards presented in a probabilistic format, similar to the majority of CPC's outlooks. The underlying probabilistic threshold used to issue a deterministic hazard is intrinsically high. Shifting the paradigm to a probabilistic format allows CPC's forecasters to highlight more areas with lower probabilities, thus providing users with more information about potential hazards during week-2. The probabilistic hazards outlook has three risk levels slight, moderate, and high corresponding to a 20%, 40%,

Fig. 1 Sample screenshot of the CPC probabilistic extremes tool for September 16, 2014. Probabilities of daily maximum temperatures on day 8 exceeding the 15th percentile are shown. Various features of the tool are highlighted in purple boxes.

Correspondence to: Melissa Ou, Climate Prediction Center, NCEP/NWS/NOAA, 5830 University Research Court, College Park, MD 20740-3818; Email: Melissa.Ou@noaa.gov

and 60% chance of occurring. This additional information about the probability of an event occurring may significantly add value to the forecast for decision makers.

The main guidance used to produce the new probabilistic hazards forecast is a newly developed extremes tool at CPC. This tool has probabilities retrieved from the CPC reforecast tool, which involves post-processing ensemble model guidance using historical reforecasts to produce calibrated week-2 daily probabilistic forecasts of daily maximum and minimum temperatures at varying climatological percentiles and threshold values, including those deemed as extreme. This tool will also be made available publicly in the future.

A main motivation for this project was the need by CPC hazards forecasters for more guidance and tools in the week-2 period. Much of the available guidance is presented as deterministic model run output from both deterministic models and ensemble means. Additionally, there is a growing need and interest in extremes and hazards outlooks, especially in the subseasonal timescale beyond week-1. These extreme/hazardous events often have the most impact to life, property, and commerce. More interests want longer lead times for early detection and preparedness. The hope is that this new extremes tool and probabilistic hazards outlook will be beneficial to a wide variety of users, especially decision makers.

The main goal of this article is to introduce the new probabilistic hazards forecast product and extremes tool and discuss the transition of the hazards product. Both the extremes tool and probabilistic hazards outlook are currently being produced for daily forecasts of 2-m minimum and maximum temperatures (which will be referred to as Tmin, and Tmax, respectively, in this article). CPC is planning to use a phased approach to add more variables in the future, such as for winds and precipitation. Some initial verification results are also shown from the probabilistic extremes tool for week-2 daily forecasts of Tmin being less than the 15th

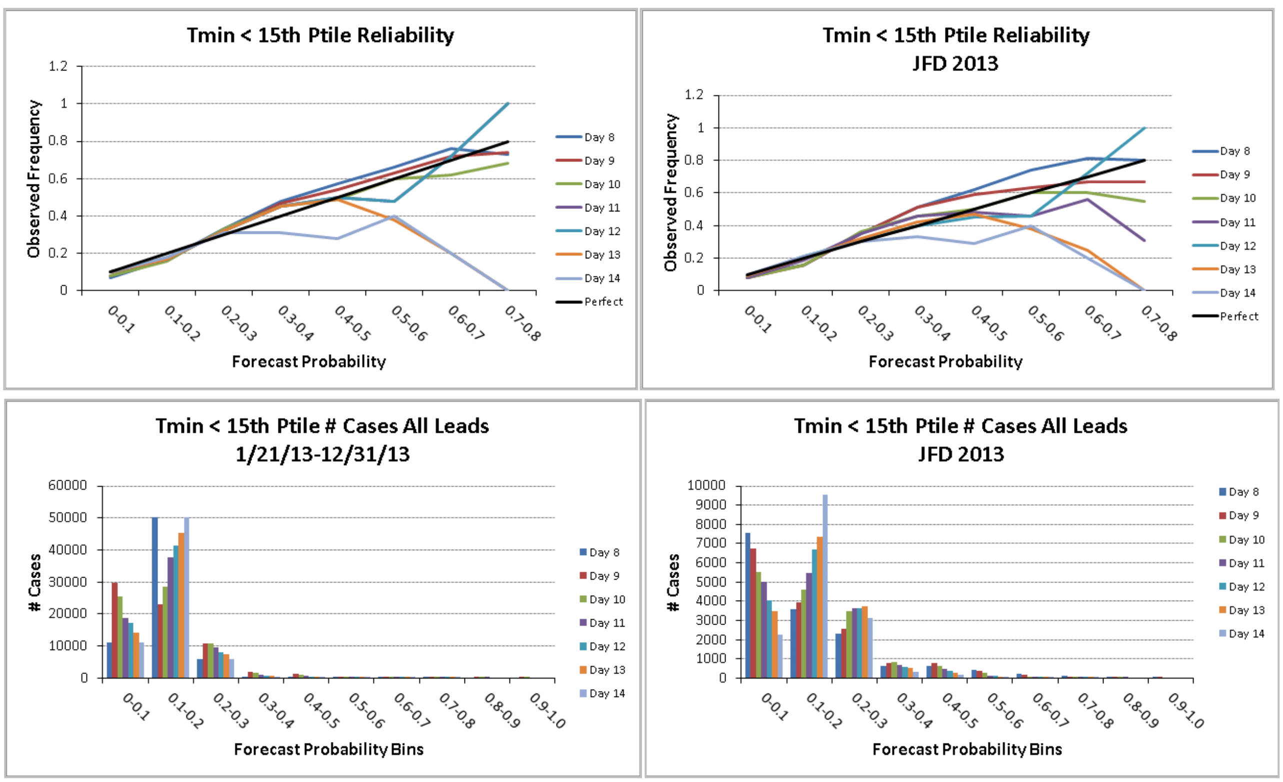


Fig. 2 Reliability diagrams are shown on the top row for daily minimum temperature forecasts produced by the extremes tool for probabilities less than the 15th ptile for all day leads including all forecasts from 2013 (left) and just the winter months (Jan, Feb, and Dec) of 2013 (right). The black line indicates what would be considered a perfectly reliable forecast. The bottom row shows histograms of the number of cases for all the probability bins for the associated reliability plots above. The '# Cases' on the y-axis refers to the fact that this shows the number of forecasts in each probability bin over time and space.

percentile, over the CONUS domain for a year (January 2013 to December 2013). We focused on Tmin because we wanted to get an idea of how skillful the tool was since CPC would be using it as guidance to produce probabilistic hazards forecasts for the upcoming winter season. The 15th percentile was evaluated because CPC typically uses this as the general threshold for deeming an event as hazardous.

2. Data

2.1 Training and forecast data

The underlying data for the extremes tool is from the CPC Reforecast Tool for daily minimum and maximum 2-m temperatures. These are formatted as the probabilities of a location being greater than a specified range of percentiles. The extremes tool software performs further postprocessing to convert this data for being less than or greater than varying thresholds for temperature values as well as percentiles. The reforecast tool calibrates realtime ensemble forecasts from the Global Ensemble Forecast System (GEFS), with physics operational during 2012, using the ensemble regression method (Unger 2009). Longterm statistics are derived using the 25-year GEFS reforecast dataset (past forecasts produced by the frozen 2012 GEFS model), provided by the Earth System Research Laboratory (ESRL), and the associated observations (the "day zero" analysis from GEFS). The reforecast dataset includes 11 ensemble members (including a control run) for each day of the 25 years, and the realtime GEFS has 21 members, daily. Further details regarding these datasets can be found in the 38th CDPW Digest summary article, "Sensitivity Study of the Skill of the CPC Week-2 Reforecast Tool to Reforecast Sampling" (Ou *et al*. 2014).

2.2 Verification data

The observational data used to produce verification results is 1 degree global, gridded, daily minimum temperature values that merges land and ocean observations from 2 different datasets from 1985 to 2010. Land data uses 1/6 degree gridded data that is derived from the Global Telecommunication System (GTS), which contains observations from about 600 US stations; data over oceans comes from the GEFS analyses, taken as the mean from 4 update cycles of the GEFS day zero analyses (forecast hours F006, F012, F018, and F024). The observation and forecast percentiles were determined using a climatology dataset derived from this merged observation dataset using all 25 years. CPC typically uses 30year climatologies, but due to the fact that reforecast data only goes back to 1985, a 25year dataset was created and used.

3. About the Week-2 CPC Probabilistic Extremes Tool

A week-2 probabilistic extremes outlook tool was developed to provide guidance to forecasters for producing probabilistic hazards forecasts. This tool was also developed with the intention for the public to eventually be able to access and utilize it across various interests via a web interface, hopefully aiding decision makers in planning. As a change in the typical paradigm to product development at CPC, feedback was sought prior to the development of the tool, in the hopes that the end product would be more userfriendly and that the content would be found useful to a wide variety of interests. Mockup designs of the user interface were presented to a number of users, including the Federal Emergency Management Agency (FEMA) and at the 11th Climate Prediction Applications Science Workshop (CPASW). This feedback was the driving factor for the design, features, and content, such as the set of percentile and temperature thresholds.

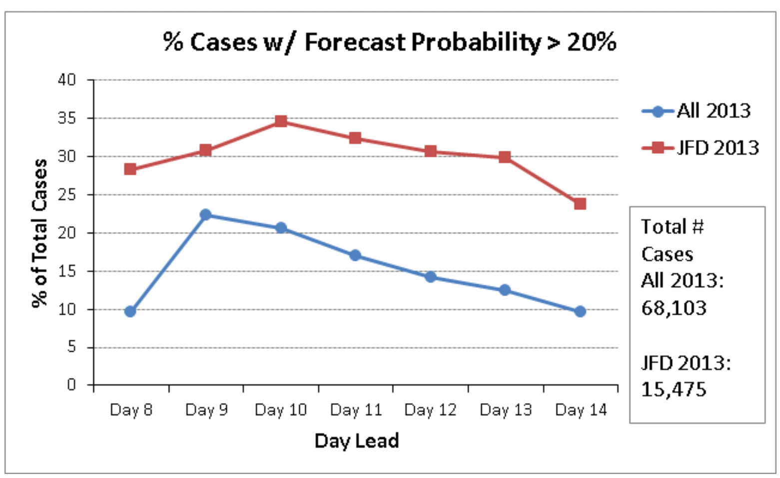


Fig. 3 Percent of total cases with forecast probabilities greater than 20% for all leads, including all dates in 2013 (blue line) and only the winter months (red line).

The extremes web tool presents the forecast as the probabilities of minimum or maximum temperatures being less than or greater than a set of thresholds. Thresholds are available as both percentiles and temperature values. Figure 1 shows a sample screenshot of the extremes tool. Below is a list of the current features and options in the web tool (highlighted in purple in Figure 1):

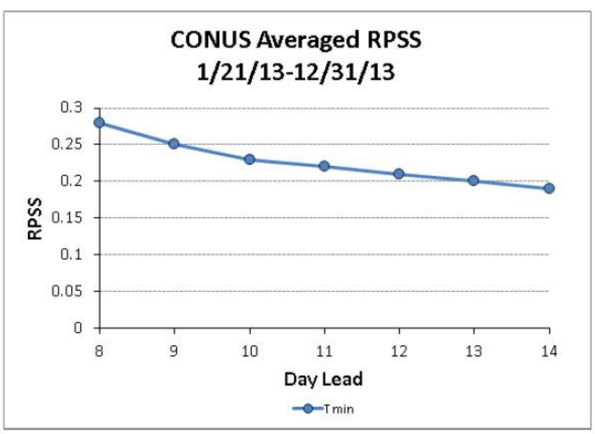
- Minimum and maximum temperatures
- Thresholds in percentiles and temperature values
- Daily outlooks for days 8 to 14
- Probabilities based on the calibrated GEFS reforecast tool and uncalibrated GEFS forecasts
- Ability to click on any location and retrieve the latitude, longitude, and probability associated with that location
- Zoom-in and zoom-out functionality

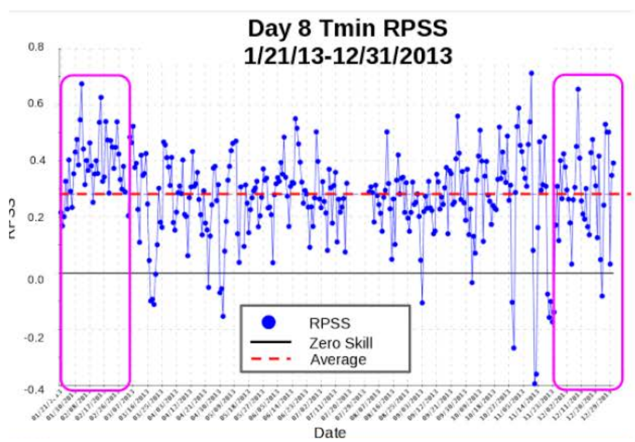
This tool is currently available for the U.S. domain, including Alaska and Hawaii. Based on further evaluation results and quality control tests, the global domain of forecasts may be made available publicly as well.

4. Verification results

The skill of daily week-2 probabilistic forecasts of minimum temperatures being less than the 15th percentile is evaluated using the ranked probability skill score (RPSS) and reliability. These skill metrics were chosen because they have been shown to be appropriate scores for assessing the skill of probabilistic forecasts. Results are presented for using all dates from January 21, 2013 to December 31, 2013 as well as only the winter months (Jan, Feb, and Dec) for 2013.

Overall, reliability (Fig. 2) was pretty good until day 13, for including all dates in the assessment period and only the winter months. There was poorer reliability at higher probabilities (*i.e.* greater than 60%), which may be due to the fact that there are less forecasts in this probability range and that it is typically inherently





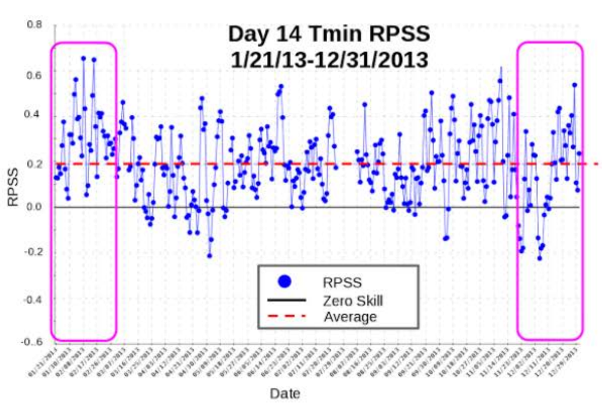


Fig. 4 CONUS averaged RPSS of daily week-2 Tmin forecasts for each lead time (top left) assessed over 2013. Timeseries of the RPSS of minimum temperature forecasts over 2013 for days 8 (top right) and 14 (bottom left). Zero skill (black horizontal line) and the average score over time (dashed red) are shown. Purple boxes surrounding values highlights the winter months. All scores are based on forecasts valid 1/21/13-12/31/13.

harder to have accurate reliability at higher probabilities.

For a forecaster, what would be useful is knowing that at days 8 and 9, forecast probabilities tend to be underconfident, potentially indicating that they should slightly bump up probabilities at i.e. the 20-50% range. Beyond day 9, forecasts tend to be overconfident at probabilities of 40% or greater, so they should be cautious of putting higher probabilities. One possible explanation for the underconfidence of the days 8 and 9 forecasts is that the ensemble regression method (used to produce the forecasts) does not widen the distribution about the members at the tails relative to near the ensemble mean. Therefore, the probabilities at the tails of the distribution are not increased sufficiently. From our experience it would appear that the probabilities do increase in the calibrated versus uncalibrated GEFS forecasts, but perhaps not enough. The overconfidence of forecasts at longer lead times likely stems from the fact that the spread of the ensemble is higher at those leads, resulting in probabilities that are more confident in the tails of the distribution, but not with more skill.

The associated histograms (Fig. 2) showing the number of forecasts (referred to as cases in the plots) in each probability bin indicates that the winter season of 2013 had a greater portion of occurrences of higher probabilities than when including all dates from all seasons, where cases are over time and space. This is most likely due to better predictability during the winter season because it is typically harder to predict summer convective events.

One concern that we had when we first started discussing a probabilistic hazards product and extremes tool was that the tool may produce a very low frequency of forecasts for the tails of the distribution, especially with probabilities that we deem high enough to consider drawing an associated probabilistic hazards contour on the forecast map. For our purposes, we chose 20% as the minimum forecast probability to consider as a potential hazard.

Our results show that a significant number of cases predicted forecasts with probabilities greater than 20 percent. Figure 3 shows the percent of forecasts (over time and space) in 2013 that had probabilities of greater than 20 percent, for both all dates and winter months only. This plot indicates that 24 to 35 percent of the winter forecasts had probabilities greater than 20%. Day 10 had the greatest percent of > 20% probabilities (35%), and day 14 had the least (24%). Day 8 and 9 fell somewhere in the middle. When considering all dates in 2013, the percent of >20% probabilities was much lower than only winter months, ranging from 10% (days 8 and 14) to 23% (day 9). This is again, likely due to the inclusion of seasons (such as summer) with less predictability, although the day 8 value for all 2013 looks suspiciously low compared to the other longer leads.

To evaluate the RPSS and reliability, we created 3 categories to represent the distribution of percentiles associated with the 15th percentile (RPSS requires multiple categories to be used). The 3 categories are Tmin < 15th percentile, 15-85th percentile, > 85th percentile.

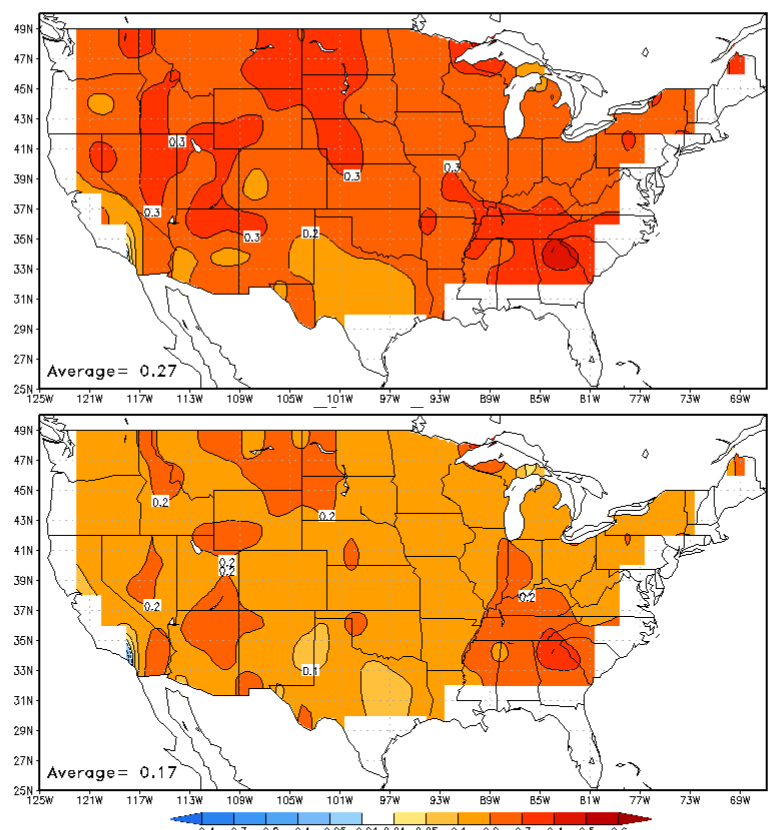


Fig. 5 RPSS over the CONUS for Tmin forecasts for days 8 (upper) and 14 (lower) for 1/21/13-12/31/13.

The RPSS averaged over the CONUS (Fig. 4) for 2013 yielded pretty good results, with the highest skill for forecasts on day 8 at about 0.28, and just below 0.2 by day 14. RPSS was positive for all leads, meaning the forecasts performed better than climatology for all leads. These results showed minimal loss in skill, even out to day 14, which is promising. For comparison, CPC's official 8-14 day tercile, probabilistic forecasts for the 7day mean temperature, had an RPSS of 0.12 for the same period. The higher skill for the extremes compared to the mean temperature may be attributed to the fact that there is more reward in RPSS for the tails of the percentile distribution. The timeseries of RPSS for days 8 and 14 shows that almost all of the forecasts had positive RPSS, with the highest scores during fall and winter.

Spatial evaluation of the RPSS of Tmin forecasts reveals positive skill across the entire CONUS for days 8 and 14 (other leads in the middle not shown), excluding a small area in southern California (which may be due to a bad data issue associated with the land border). This shows that there is an improvement upon climatology for all locations of the CONUS. The spatial distribution of skill was relatively uniform, with the least skill across the Southern Plains for both leads, and the most skill across the Southeast. The majority of the CONUS had skill of 0.2 or greater at day 8, and 0.1 or greater at day 14.

Reliability diagrams (Fig 6) reveal pretty good reliability for all three categories at day 8. The lower category (< 15th percentile) was underconfident for all probabilities, whereas the middle category (15-85th percentile) and above category (> 85th percentile) have better reliability, although somewhat overconfident at higher probabilities. At day 14, reliability is significantly worse than day 8, as expected. The lower category at day 14 is overconfident, which is the opposite of the day 8 results. This may be in part due to the fact that the model often produces higher probabilities at a longer leads, thinking it has more skill than would really

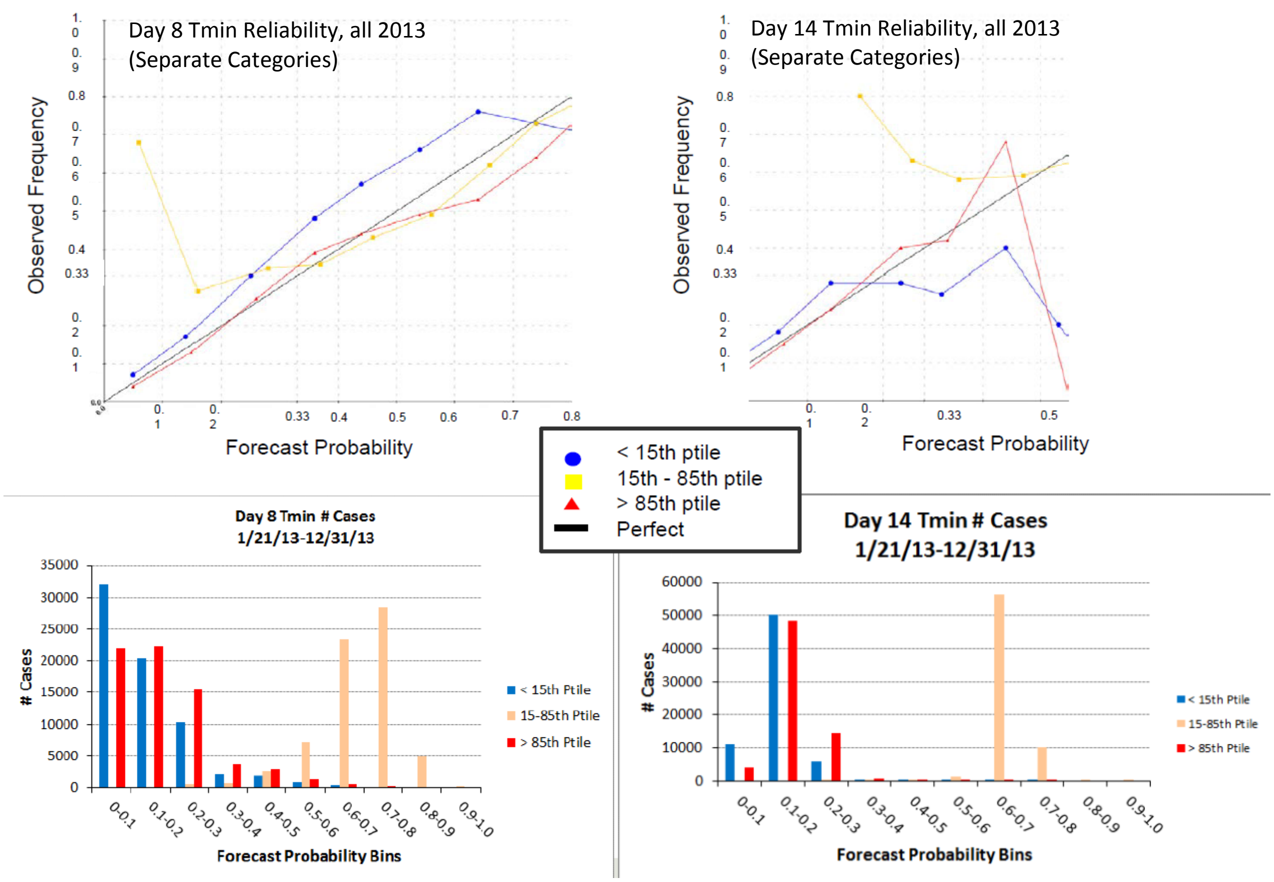


Fig. 6 The top row shows reliability diagrams for days 8 (top left) and 14 (top right) for 3 categories of percentile thresholds. The black diagonal line indicates what would be considered perfect reliability. The bottom row shows associated histograms (with the reliability diagrams above) with the number of forecasts associated with each of the probability bins for days 8 (bottom left) and 14 (bottom right) for the same categories.

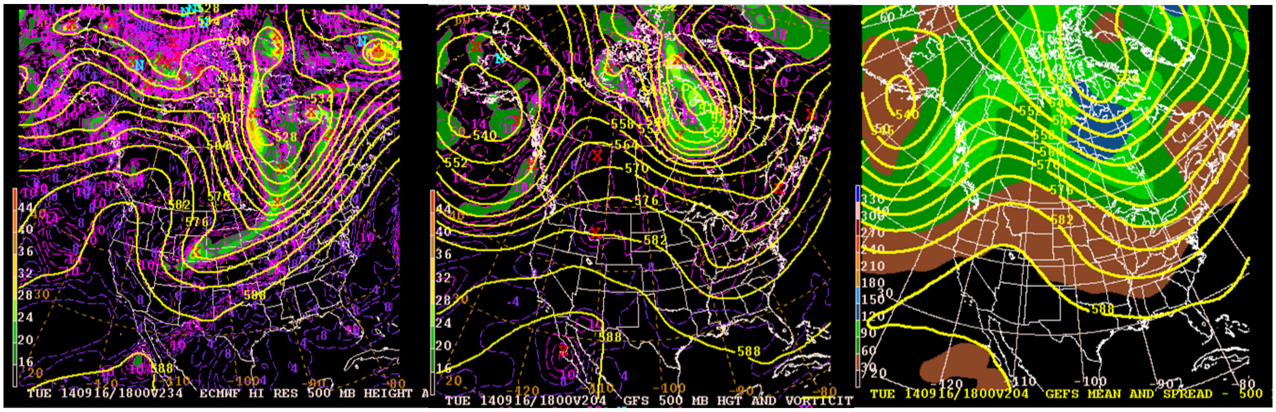


Fig. 7 Model run 500MB heights (yellow contours) and vorticity (shaded) from the deterministic 0Z ECMWF (left) and deterministic 6Z GFS (center). 500MB heights (yellow contours) and ensemble spread (shaded) are shown from the 6Z GEFS (right). All maps are valid 9/16/14, 18Z.

exist at that lead. There is likely less calibration performed at shorter leads than longer leads because skill is typically higher at shorter lead times. Therefore, the probabilities get scaled back more for shorter leads compared to the longer (calibration tends to lower the typically higher forecast probabilities from raw, uncalibrated ensemble forecasts). Results also indicate that at lead 8, there are more cases in the extreme categories (< 15th and >85th percentiles) with higher probabilities than day 14, as expected. There is better skill at shorter lead times, producing sharper probability distributions by the model. For both leads, the outlying reliability values for lower probabilities are likely due to the small sample sizes of forecasts in those probability bins. Curves in the reliability diagram are not shown in the day 14 diagram for probability bins greater than 0.6 because there were no forecasts of any category in those bins at that lead.

5. Week-2 Probabilistic Hazards Forecast

a. About the forecast

CPC has been issuing daily probabilistic week-2 (days 8-14) hazards forecasts over the CONUS and Alaska since July 2014, in addition to the week-2 deterministic map, which has been issued for a number of years. These are manually drawn by the forecaster and are publicly available on the CPC

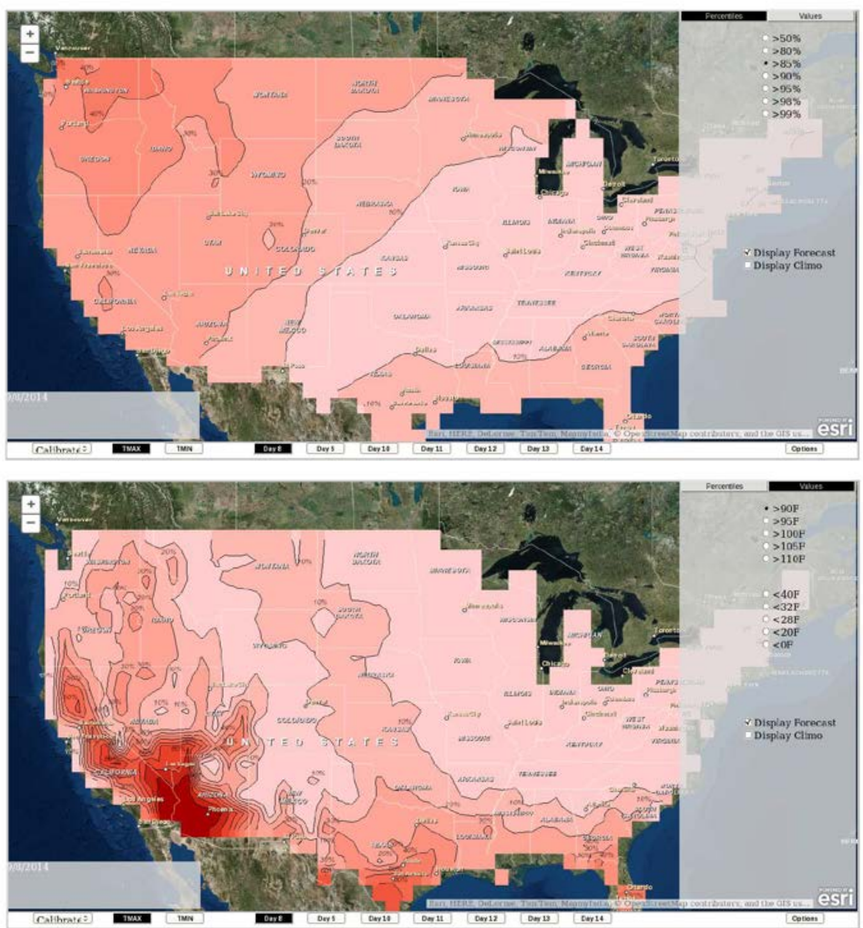


Fig. 8 Extremes tool forecast valid day 8, showing the probabilities of the daily maximum temperatures exceeding the 85th percentile (upper) and 90 degrees F (lower) based on CPC's GEFS reforecast tool.

website. The probabilistic week-2 maps represent the probabilities of a daily minimum or maximum temperature reaching what we deem as hazardous criteria for each of the days in week-2, which is by nature a subjective forecast. Forecasters use the previously discussed probabilistic extremes tool as the main guidance regarding the probabilities, looking at both the percentiles and temperature value thresholds. They additionally evaluate the daily synoptic pattern and temperatures from various models. The hazards highlighted on the map are based on the forecasters' previous experiences and base knowledge and agreement amongst forecasters of what are considered regional thresholds and conditions for hazardous events. This information is often obtained from discussions and criteria provided by external regional interests and forecasters, such as the local weather forecast offices (WFOs). As mentioned earlier, CPC typically uses the upper and lower 15th percentiles as guidance for possibly deeming an event as hazardous, in combination with the probabilities of a temperature being below or above a certain threshold, which varies based on the location and time regime. Every Tuesday, a briefing is hosted by the hazards forecaster in which various interests dial-in, providing feedback and suggestions regarding the forecast. This enables local and regional weather service offices to provide valuable information to the hazards forecaster regarding whether an event should be highlighted or not, in addition to ensuring forecast coordination on various National Weather Service levels.

There are 3 levels of probabilities of a hazardous event occurring that are included in the forecast. Each level is denoted on the map with a qualitative description which has underlying probabilities associated with them. Qualitative versions of describing the risk level were used because it may convey the concept of varying probabilities better to the public. These categories are similar to those used by the Storm Prediction Center (SPC). The risk levels and their associated forecast probabilities are slight risk (20%), moderate risk (40%), and high risk (60%). Any contours greater than or equal to 40% (moderate or high risk) are automatically included on the week-2 deterministic hazards map. The web page for this product is http://www.cpc.ncep.noaa.gov/products/predictions/threats/threats.php.

b. Example forecast case

An example case of a forecast valid September 16, 2014 is discussed here. The daily model runs from the deterministic 0Z ECMWF and 6Z GFS, and 6Z GEFS (18Z) (Fig. 7) all indicated the possibility of an amplified ridge developing across the northwestern CONUS, with some variation in the level of amplification and ridge axis orientation. The 0Z ECMWF had the most amplified pattern, with the ridge axis focused over the Pacific Northwest, while the GFS and GEFS had the ridge shifted slightly eastward and not as amplified. Overall, this pattern would indicate the potential for warmer temperatures across the northwestern CONUS.

The probabilistic extremes tool for this day (Fig. 8) showed a 30-40% chance of Tmax being greater than the 85th percentile for the Pacific Northwest, and 30% chance of this area being greater than 90 degrees F. This temperature would be considered unusually high for this time of year with respect to the region. Therefore, the forecaster decided to issue a slight risk for much above normal temperatures for parts of Washington and Oregon for this day, as well as a moderate area embedded within this slight risk contour, focused over the central portion of this region (Fig. 9).

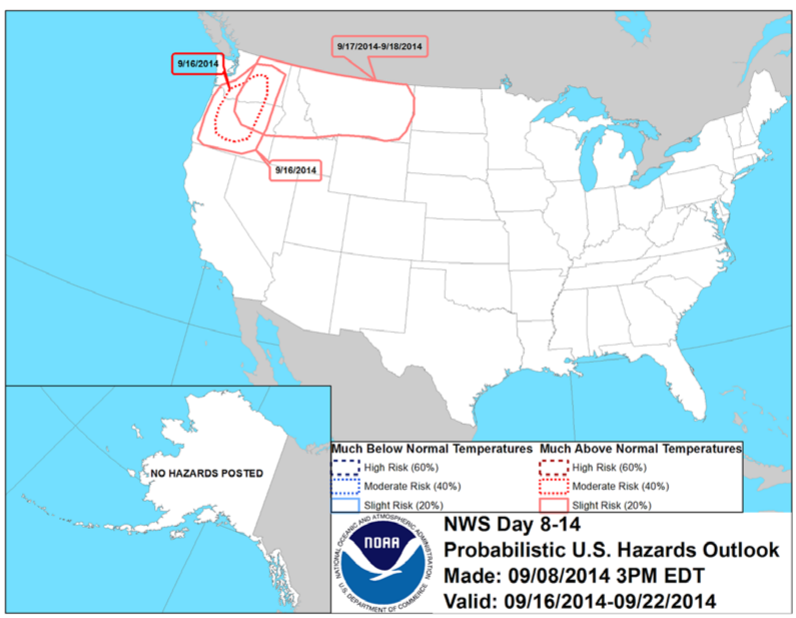


Fig. 9 Week-2 CPC probabilistic hazards forecast issued 9/8/2014, valid 9/16/2014-9/22/2014. Contours represent slight (light red solid line), moderate (medium red dotted line), and high risk (dark red dashed line) of much above normal temperatures.

An additional area was highlighted across the Pacific Northwest eastward into Montana for days 9 and 10 (not shown).

The verifying observations (Fig. 10) show that on 9/16/2014 temperatures were above normal that day across the Pacific Northwest, with anomalies in the 10-12 degrees F range and temperatures reaching 85 to 90 degrees. Overall, this would be considered a pretty good forecast, in that it did highlight the main areas that these temperatures would be considered very anomalous. The forecast area probably could have been extended slightly eastward into Idaho as well.

For the following day, the highlighted area for slight risk was not too bad either. The greatest temperature anomalies occurred over western Montana, which was included in the contour, although temperatures were only around 80 to 85 degrees F.

An example screenshot of CPC's hazard webpage that includes the week-2 probabilistic hazards forecast is shown in figure 11. There are options on the page to view the days 3-7 and 8-14 hazards forecast as well as the days 8-14 probabilistic forecast.

6. Summary

CPC has a new week-2 probabilistic hazards forecast, issued daily, which highlights potential hazards in probabilistic format. Currently this is done for Tmin and Tmax, but additional variables are planned to be added in the future. There are 3 categories representing the levels of likelihood of an event occurring. The thinking is that as the lead time shortens for an event, the probabilities of the event occurring may increase

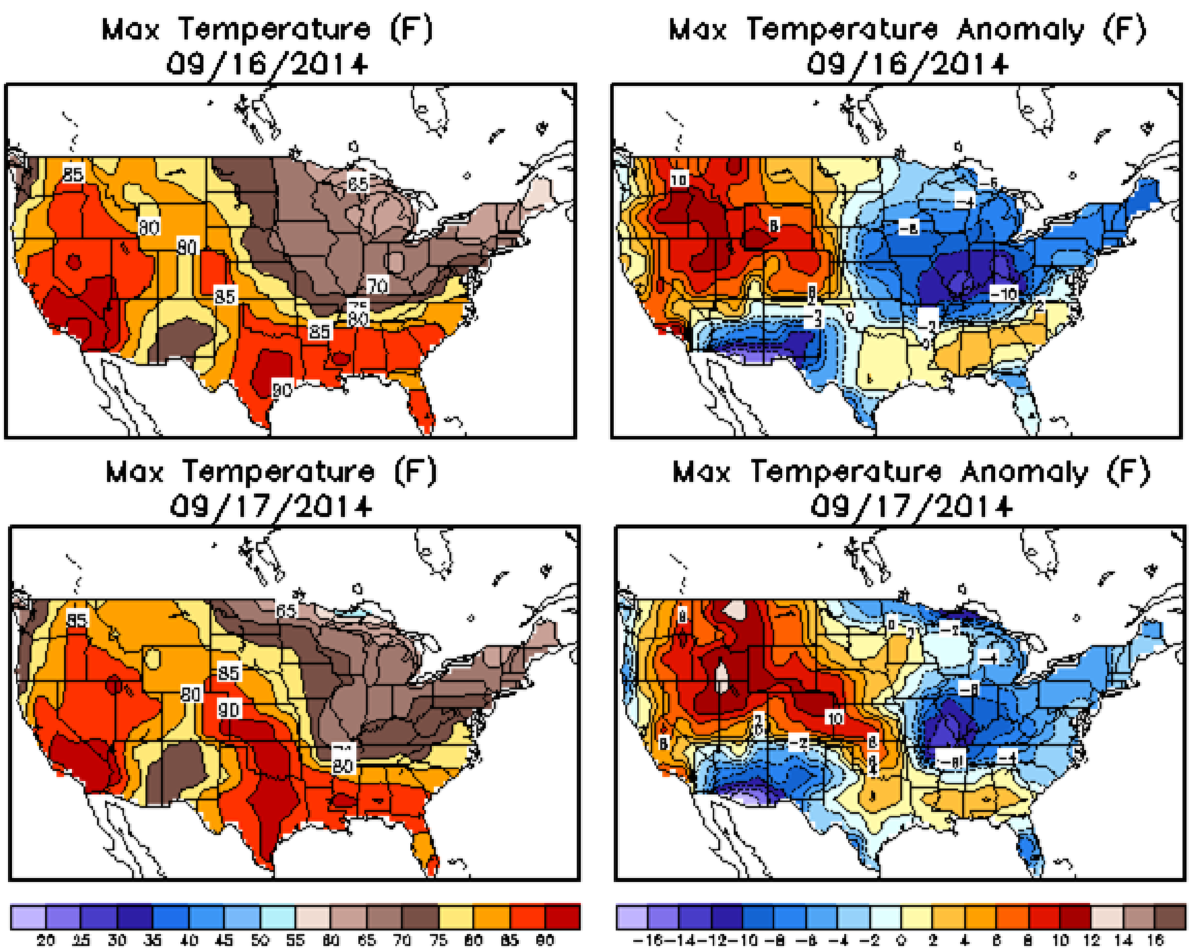


Fig. 10 Verifying observations of Tmax (left column) and Tmax anomalies (right column), for 9/16/2014 (top row) and 9/17/2014 (bottom row). All values are in degrees F.

due to higher certainty, finally enabling the hazard to be included on the deterministic map. The main guidance for this product is the newly developed probabilistic extremes tool at CPC, which provides an objective probabilistic outlook of extreme temperatures at varying percentile and temperature thresholds. The underlying data of the tool is currently from the reforecast tool at CPC, which calibrates GEFS model output using long term statistics from the 25year GEFS reforecast dataset.

The benefit of having the week-2 hazards in probabilistic format is that it enables more hazards to be put on the map. Previously, when CPC only issued the deterministic days 8-14 hazards, forecasters were often unable to put hazards on the map because of the inherent higher confidence/probabilities required to delineate a hazard deterministically would be very high. By expressing hazards as probabilities with varying tiers of likelihood, forecasters are able to include lower certainty/probability events at a longer lead. Extra information associated with these levels of likelihood will hopefully add value to the forecasts, especially for decision makers.

Overall, the initial verification of the probabilistic extremes tool shows positive skill (reliability and RPSS). There is decent improvement over climatology and shows the best skill for the winter season, which is expected since winter typically has better predictability. The tool produces significant probabilities (defined

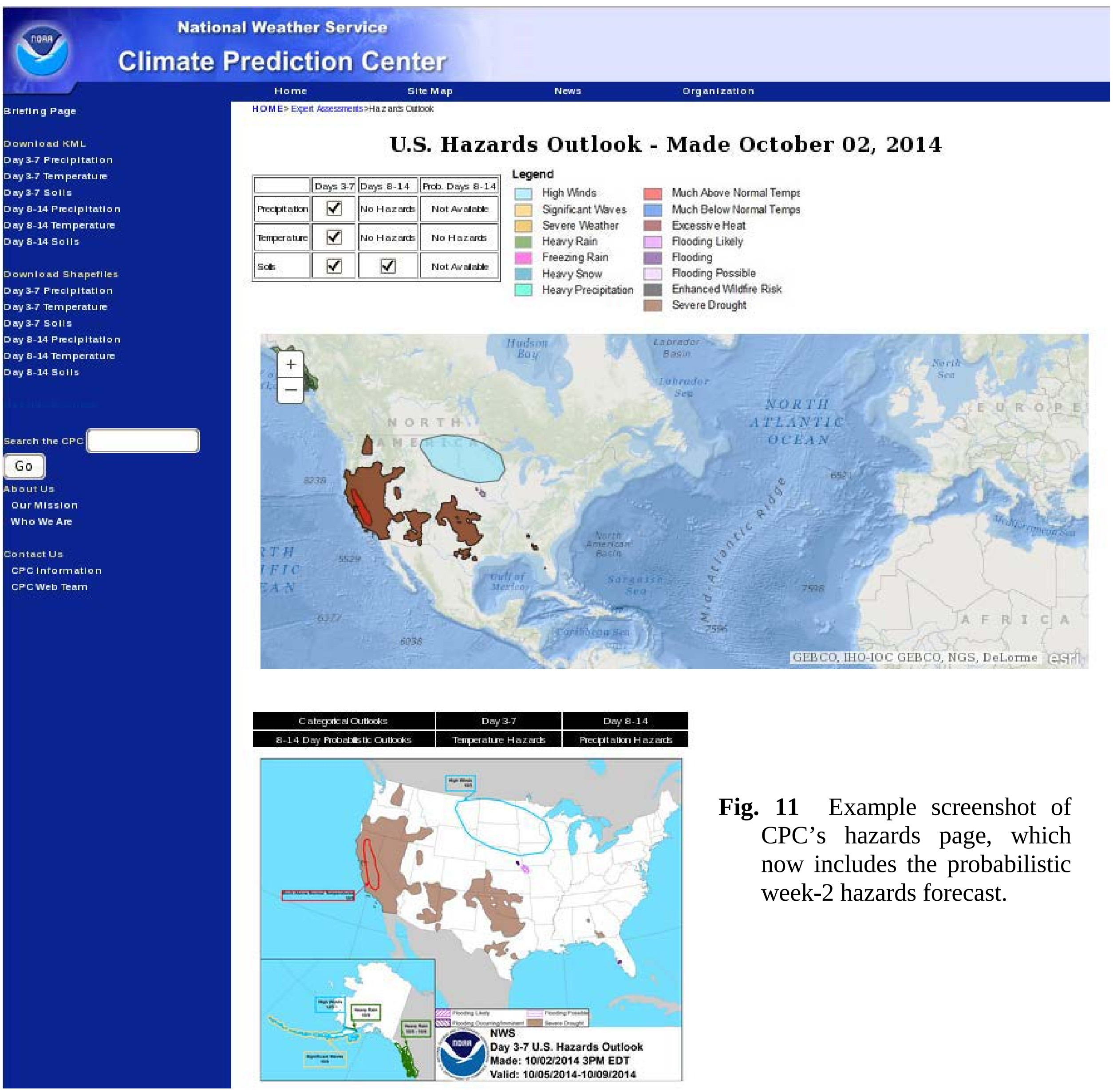


Fig. 11 Example screenshot of CPC's hazards page, which now includes the probabilistic week-2 hazards forecast.

as 20% or greater) of extremes at a pretty good frequency, enabling forecasters to potentially include an event on the probabilistic hazards forecast 10 to 20% of the time (based on a year's worth of evaluation over 2013). There are especially good opportunities for hazardous Tmin probabilities to be issued during the winter season.

7. Ongoing/future work

This project is planned to have a phased approach for adding new features, such as more variables (precipitation, wind, *etc.*), thresholds, and models. Another tool is currently being developed at CPC, which consolidates forecasts from different sources (dynamical and statistical models), by which the weights used for combination are determined by the past skill of the model/tool. CPC plans to utilize this consolidated forecast to derive a multimodel forecast for probabilistic extremes in the future. This would hopefully improve the skill of the probabilistic extremes tool, and thus the manually issued forecasts.

Additional skill evaluation of the extremes tool is planned to be extended for Tmax, Alaska, and other percentiles and temperature values. Verification of the tool will also likely be performed on forecasts produced by the calibrated versus uncalibrated GEFS reforecast tool and impacts associated with the relationship of spread to skill. Global or other regional verification may also be done by other members of CPC, such as to evaluate the ability to use this tool to support the global tropical hazards forecasts, as well as external users.

References

Ou, M., M. Charles, D. Collins, and E. Riddle, 2014: Sensitivity study of the skill of the CPC week-2 reforecast tool to reforecast sampling. *Climate Prediction S&T Digest*, 38th NOAA Annual Climate Diagnostics and Prediction Workshop, College Park, MD, National Oceanic and Atmospheric Administration, 36-43. [Available online at http://docs.lib.noaa.gov/noaa_documents/NWS/NCEP/CDPW/38th_2014.pdf]

Unger, D., A., and Co-authors, 2009: Ensemble regression. *Mon. Wea. Rev.*, **137**, 2365-2379.