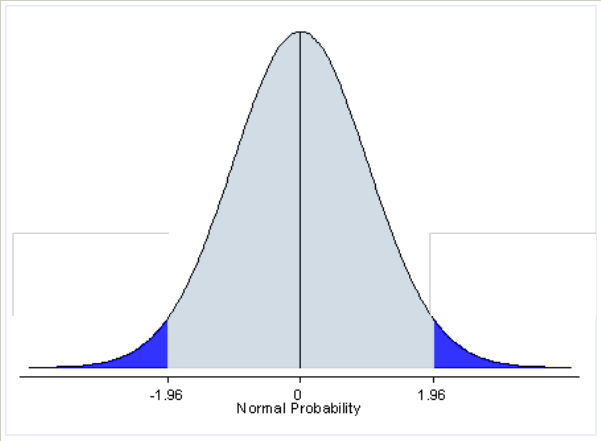
CPC's New Week-2 Probabilistic Hazards Forecast and Extremes Tool

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

Acknowledgements:
Dave Unger, Stephen Baxter, and Dan Collins



Overview

- Goals
- Background / Motivation
- Mission
- Probabilistic extremes Tool
- Verification results
- Probabilistic week-2 hazards forecast
  - Example case
- Summary
- Ongoing/Future Work

Background / Motivation

- CPC currently produces categorical (deterministic) daily hazards outlooks for days 3-14
- CPC is now focusing on week-2 probabilistic hazards
  - Probabilistic format more appropriate beyond week-1
- Hazards forecasters needed more guidance/tools for the 8-14 day period
- Growing need/interest in extremes and hazards forecasts
  - Longer lead times

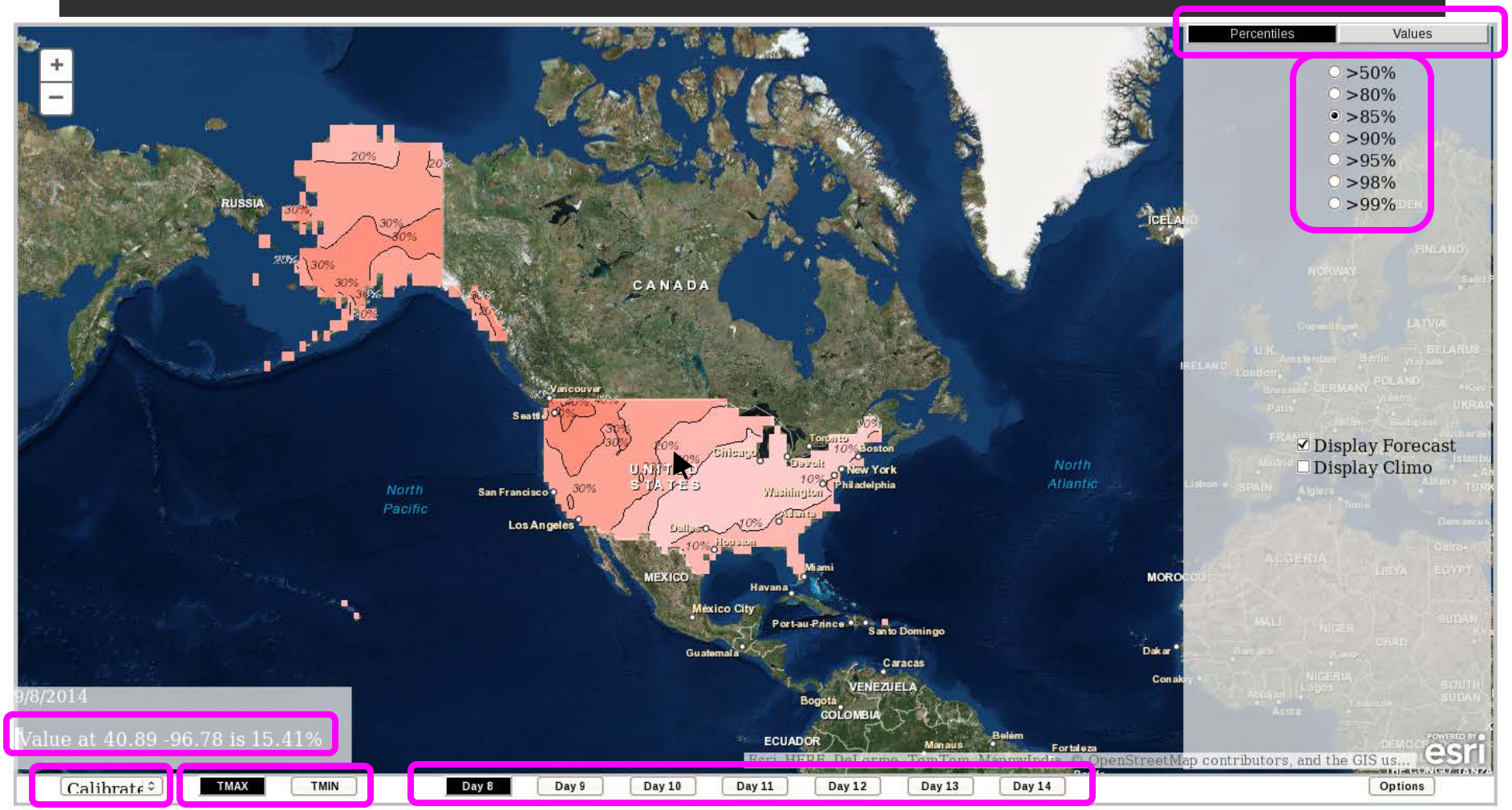
Mission

- Develop an objective probabilistic outlook tool of daily extremes for week-2
- Issue week-2 daily probabilistic hazards forecasts
  - more hazards on map
- Focused on creating a tool that would be user-friendly as well as informative (especially for decision makers)
- Engage users and get feedback prior to product development (ie. FEMA) via mockups

Probabilistic Extremes Tool

- Forecast probabilities extracted from the GEFS Reforecast Tool (POEs)
- Currently producing Tmin and Tmax
- Daily forecasts, days 8-14
- Forecasts in percentiles and actual values
- Has been used since this July in the manual forecast process
- Will be publicly available in future, potentially global domain

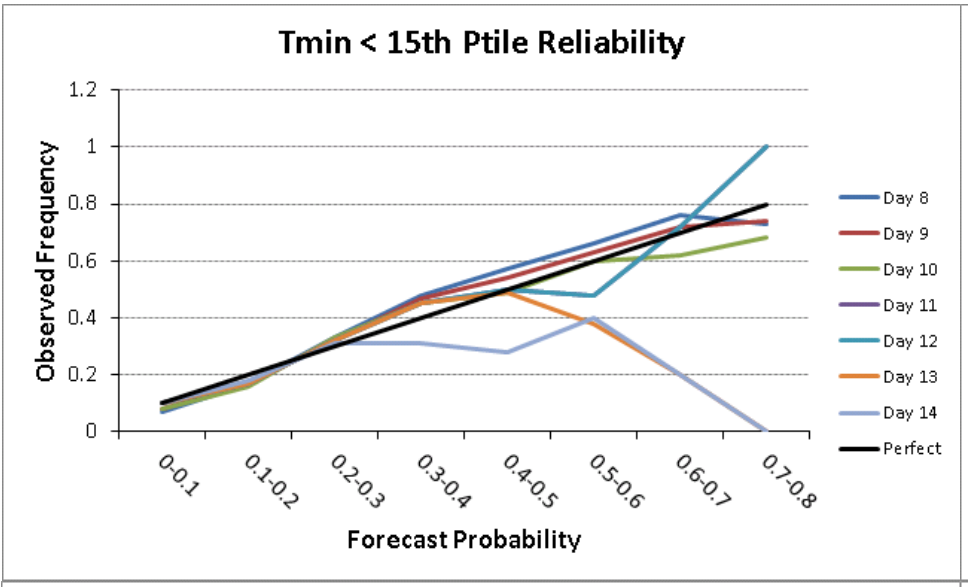
Example tool display (Tmax) Valid 9/16/2014 (Day 8)

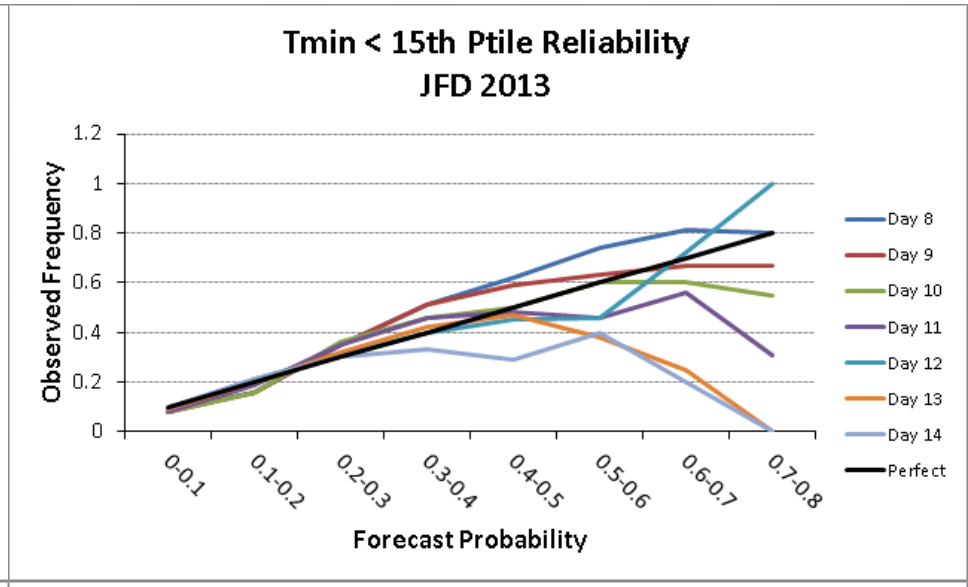


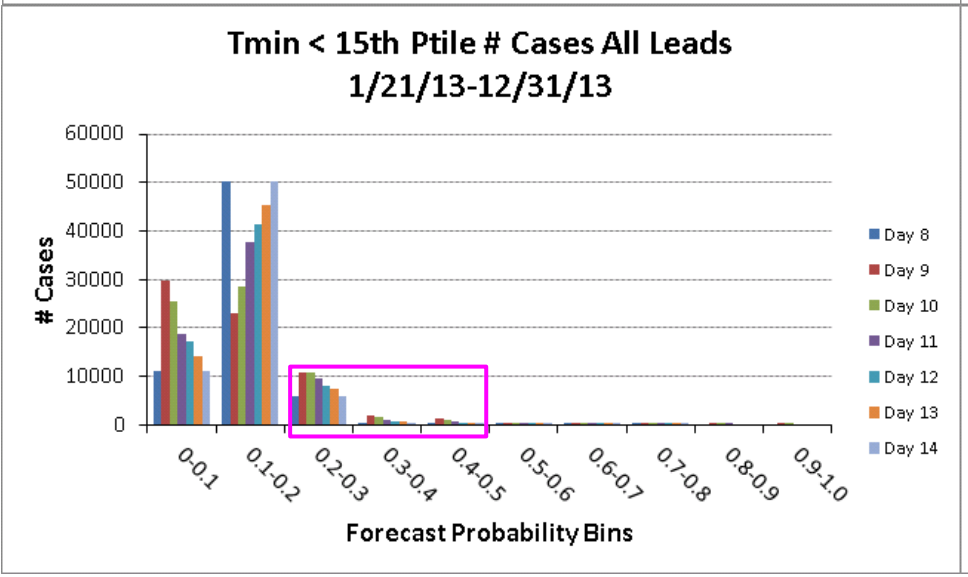
Verification

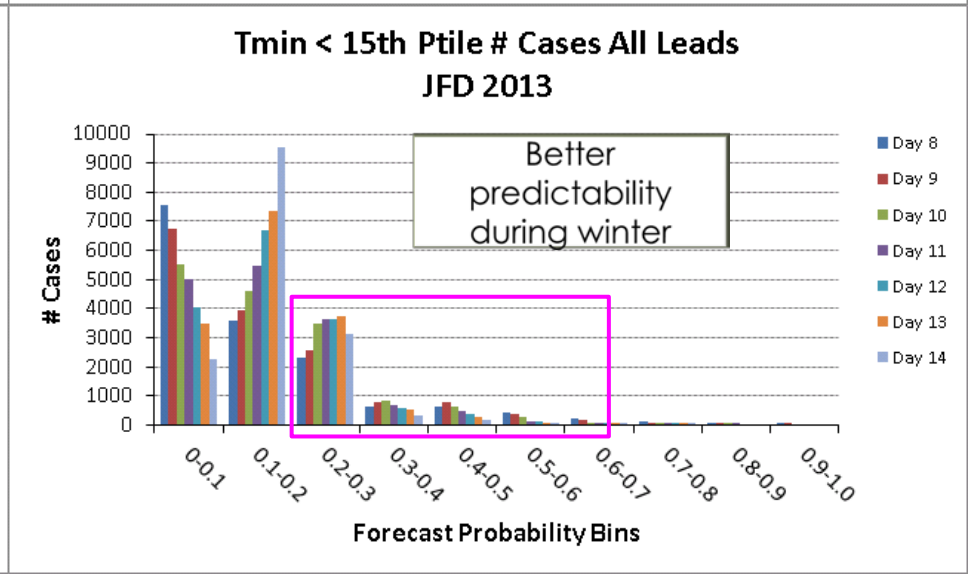
- Results obtained using CPC verification web tool
- RPSS and Reliability results presented
- Evaluate Tmin < 15th ptile
- CONUS Domain
- Tmin verified for 1/21/13-12/31/13
  - Evaluation over this entire period and winter months of 2013 (Jan, Feb, Dec 2013)
- Data:
  - Forecasts - GEFS Reforecast Tool,1 deg
  - Observations - 1 deg. Merged : Land uses 1/6 deg grid comprised of GTS ~600 US stations. Water uses GEFS analyses
  - Climatology - Based on merged obs 1985 - 2010

Verification - Reliability (All 2013 and JFD 2013)





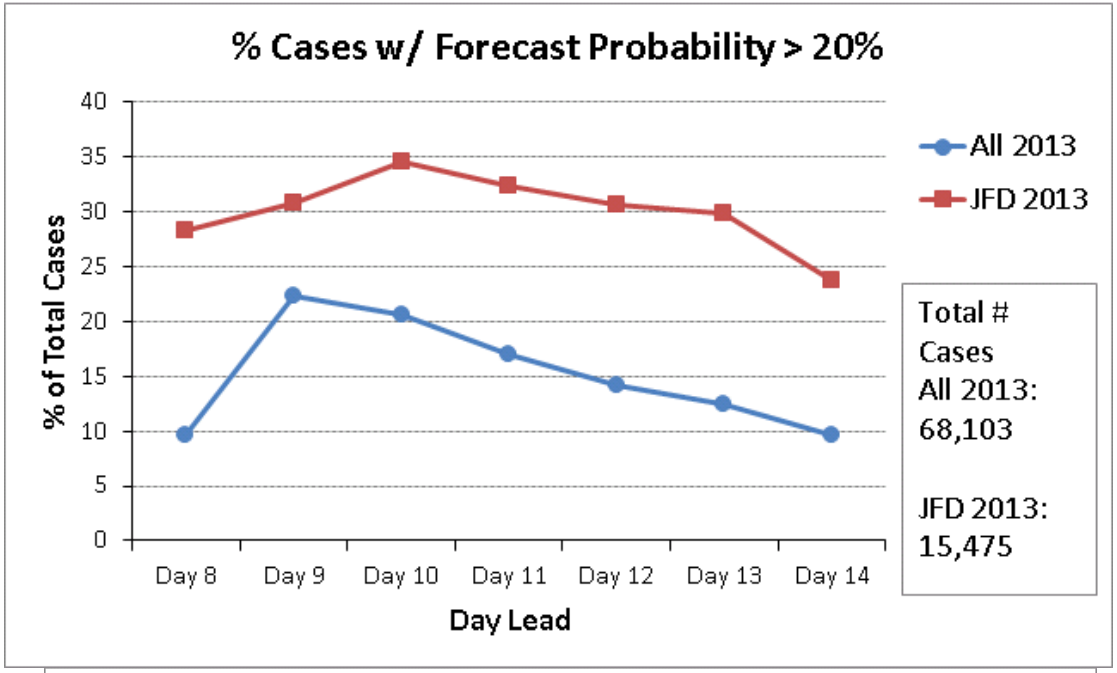


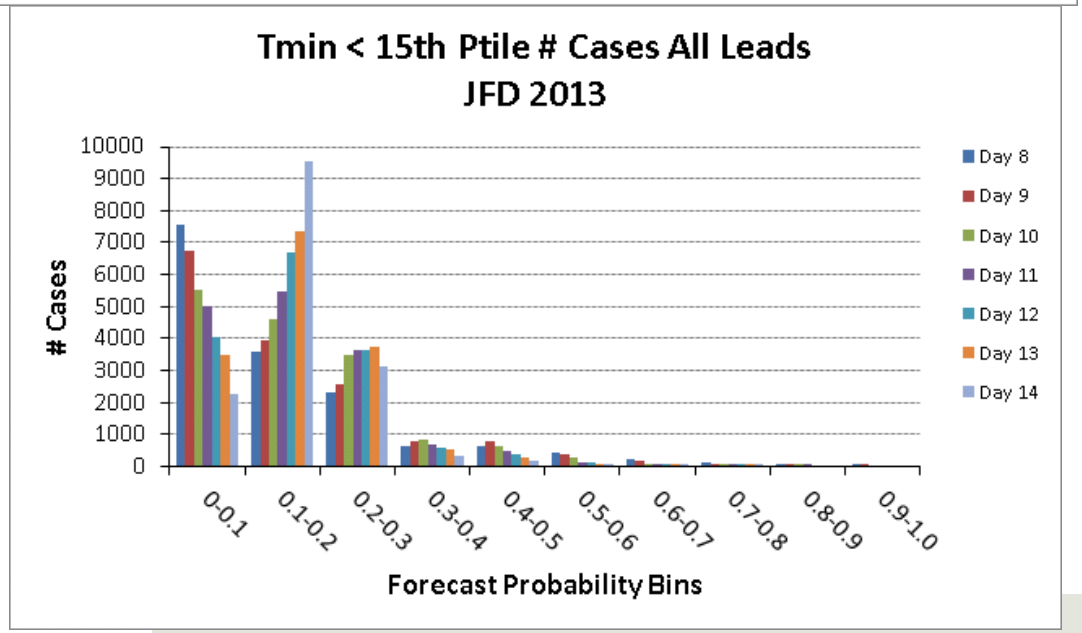


All 2013: ~ 68,000 cases in total

JFD 2013: ~ 15,500 cases in total

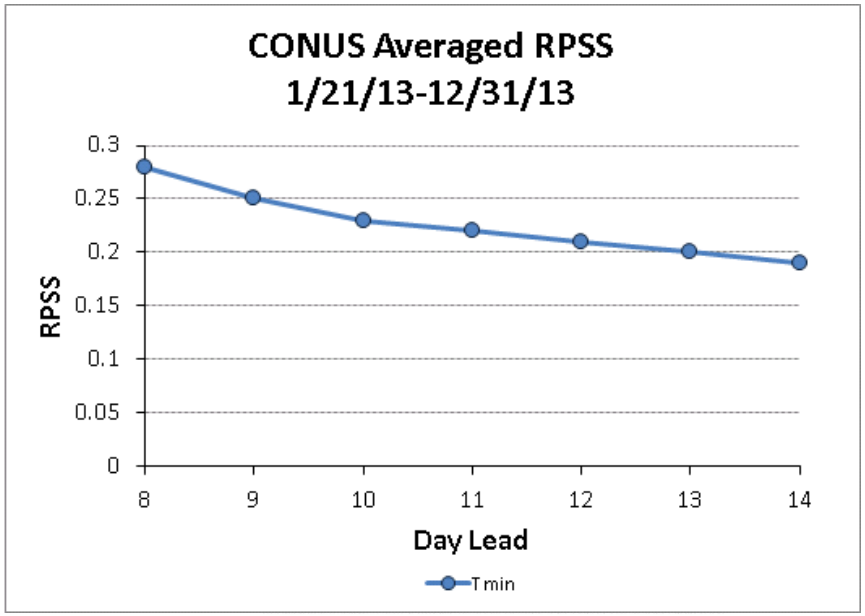
Frequency of forecasts with > 20% probability

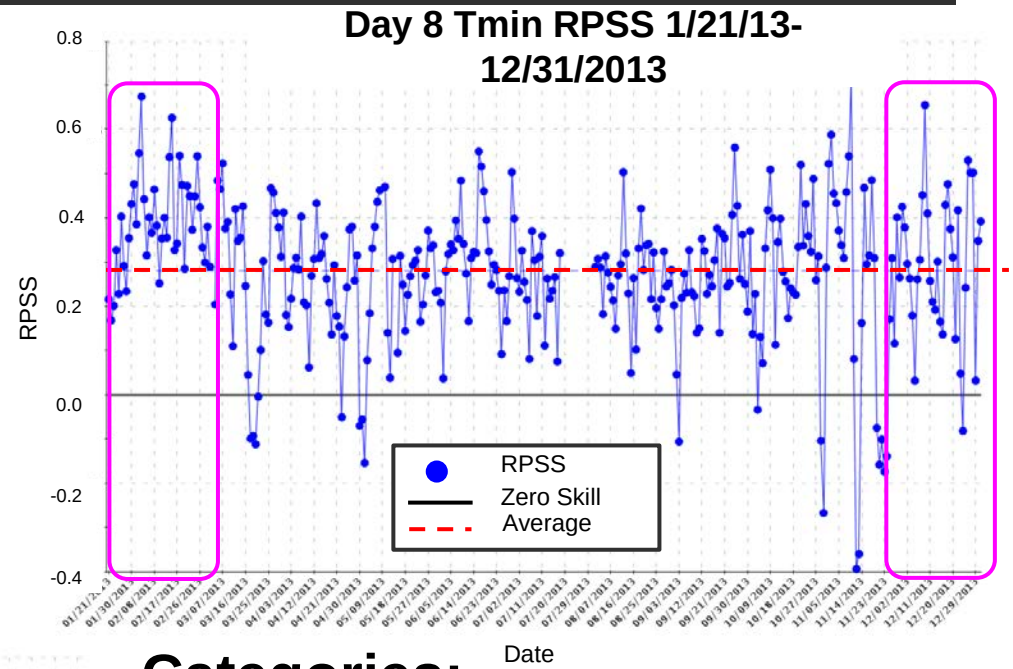


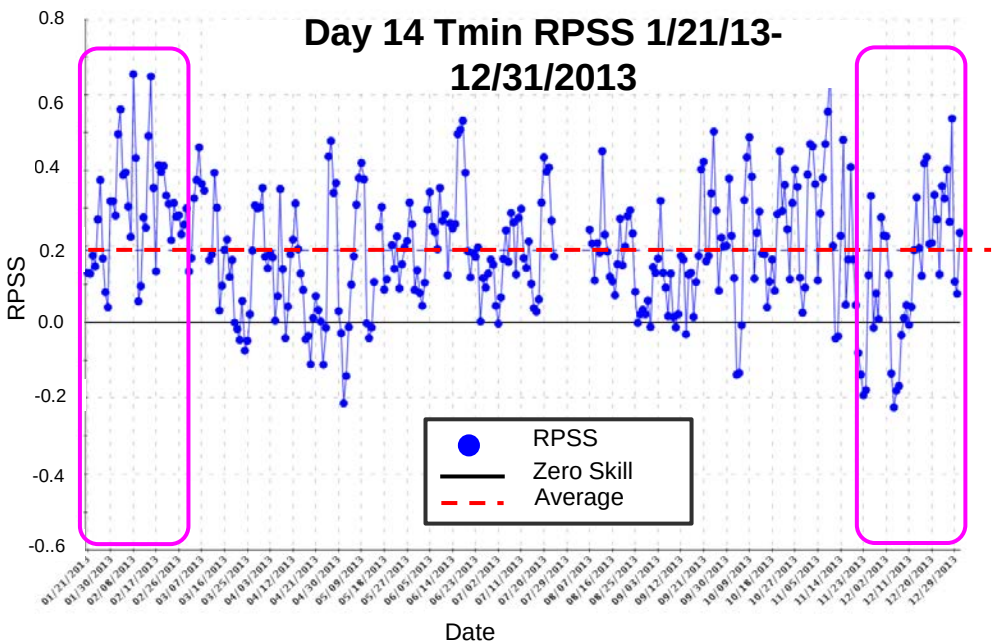


- One concern was that there would be a low frequency of forecasts with sufficient probabilities at the week-2 timescale
- Focus on > 20% probabilities (threshold for issuing slight risk)
- 24 - 35% of winter cases had probabilities > 20%

Verification - RPSS using 3 categories







Categories:

Tmin< 15th, 15-85th, > 85th ptile

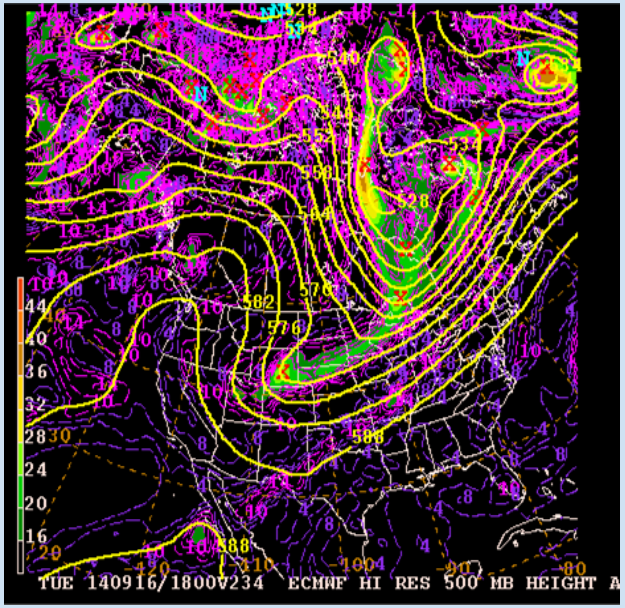
- CONUS averaged RPSS is positive at all leads at around 0.2 or greater
- Much skill is retained even until day 14
- Better scores during winter months

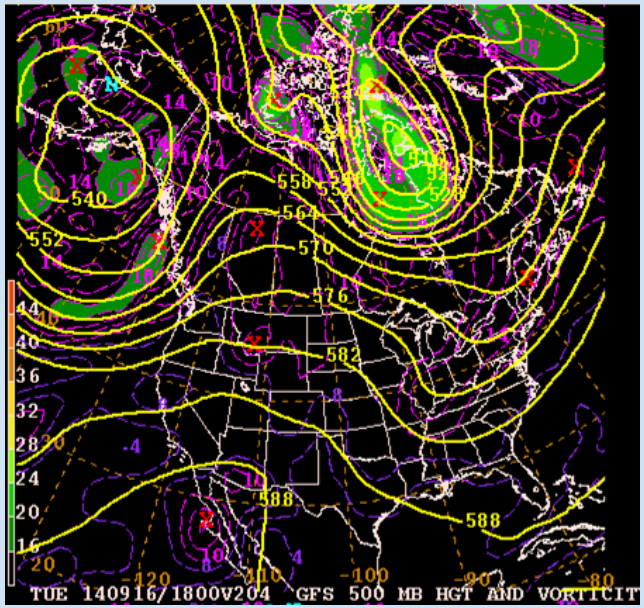
Probabilistic Hazards Forecast

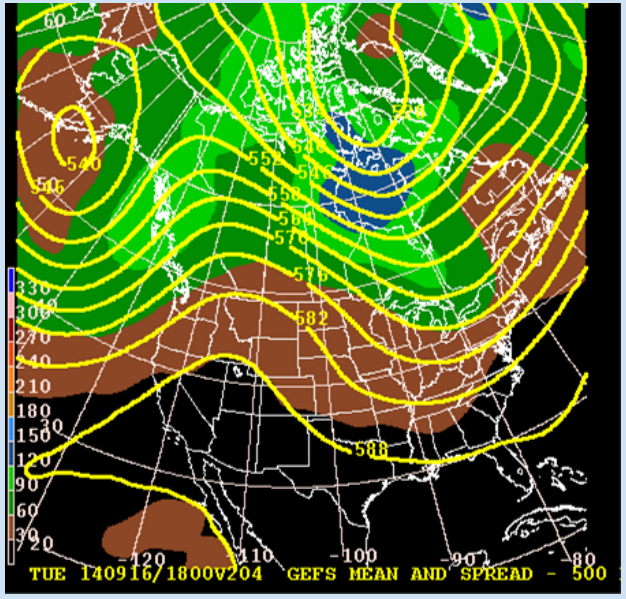
- Typically use upper and lower 15th percentiles, combined with a subjective assessment of temperatures (regional/temporal)
- Forecasters use this tool in combination with analyzing synoptic patterns and fields from model guidance
- Much below normal Tmin/ Much above normal Tmax
- 3 probability levels
  - Slight Risk (20%)
  - Moderate Risk (40%)
  - High Risk (60%)
- Probabilities >= 40% are included on the deterministic map

Sample Case Valid 9/16/14 (D8)

Model Runs of 500MB Heights Valid 9/16/14 18Z







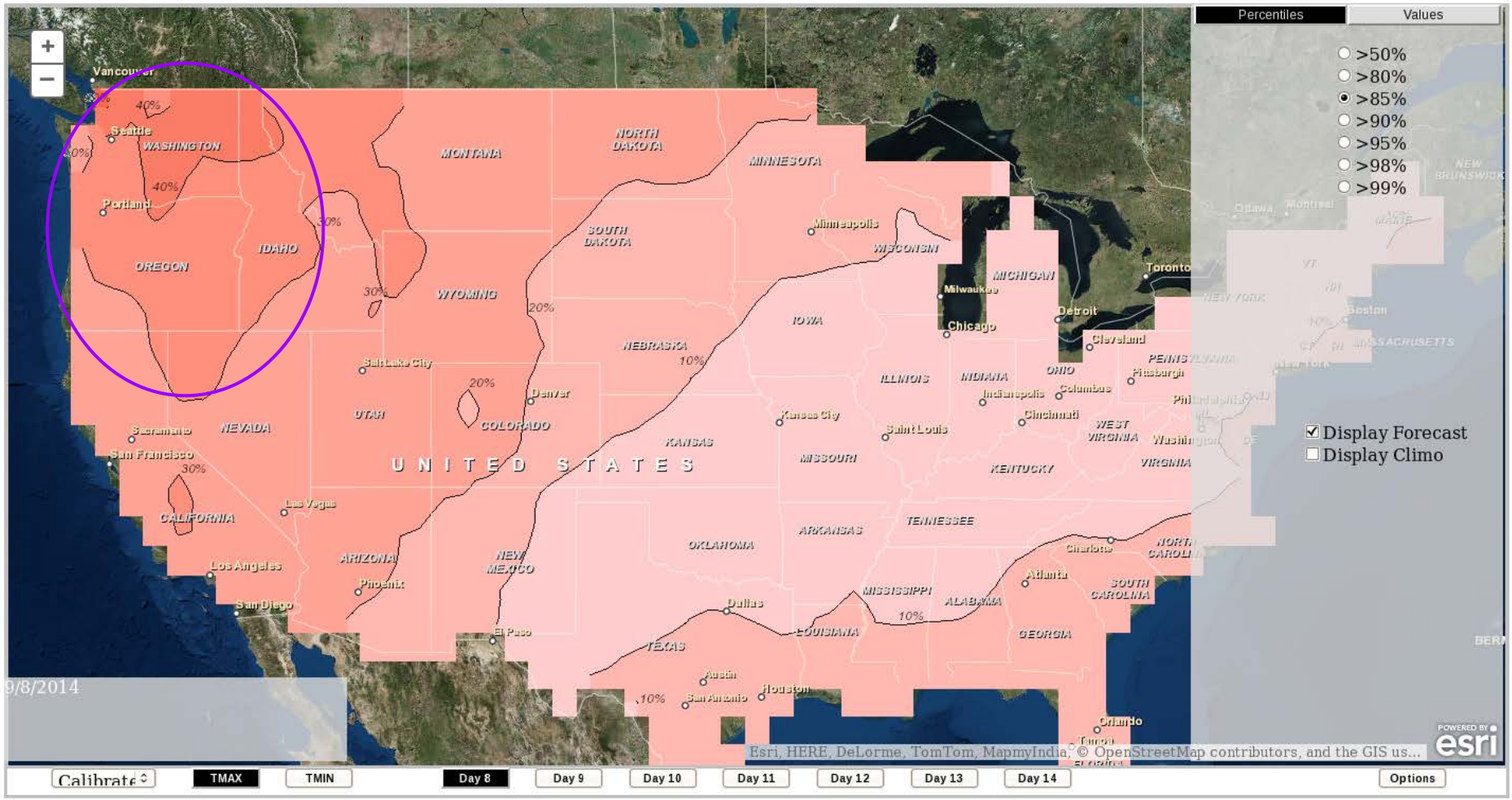
Det 0Z ECMWF Heights and Vorticity

Det 6Z GFS Heights and vorticity

6Z GEFS Mean and spread

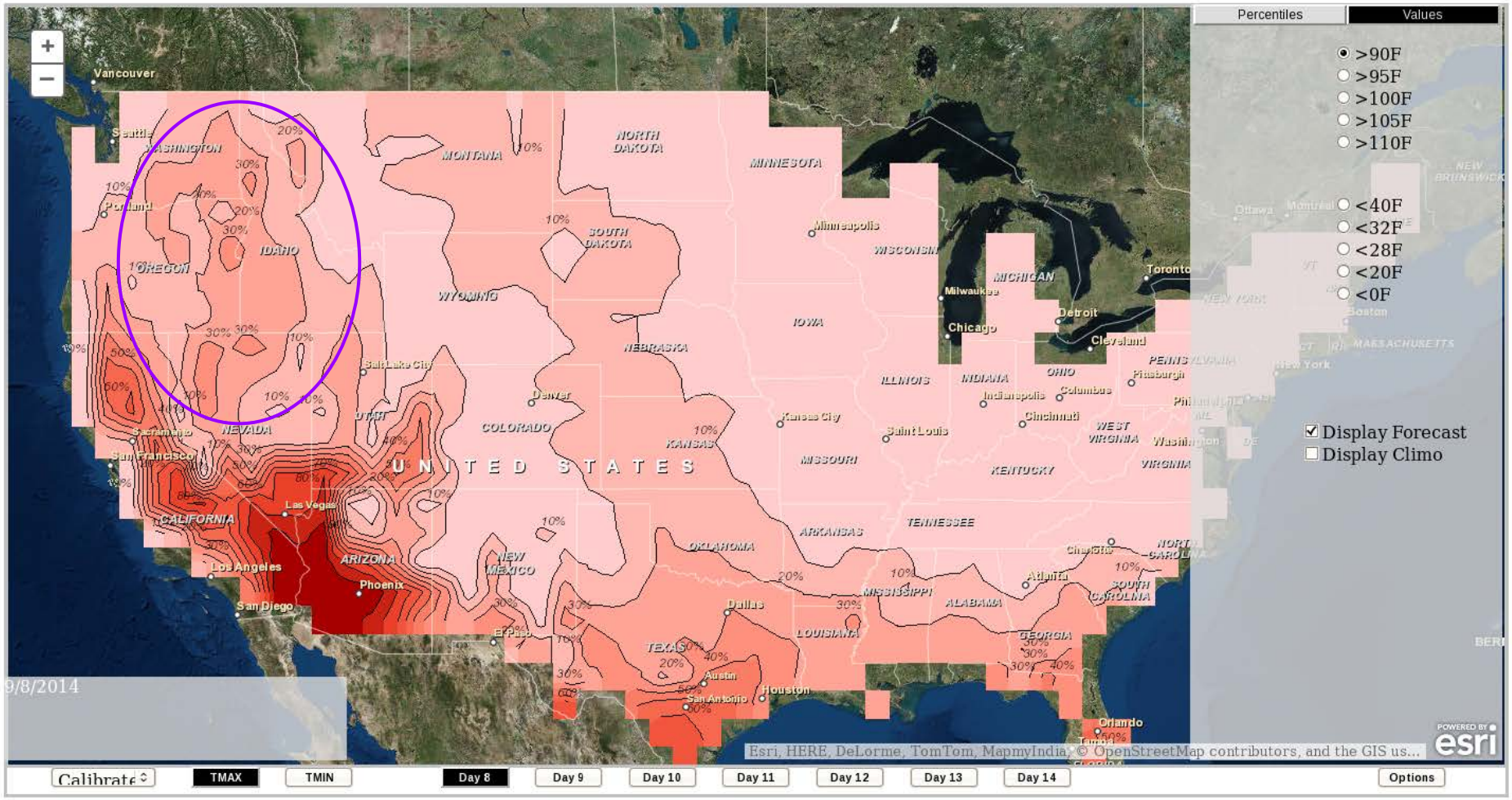
Sample Case Valid 9/16/14 (D8)

Tmax percentiles > 85th percentile



Sample Case Valid 9/16/14 (D8)

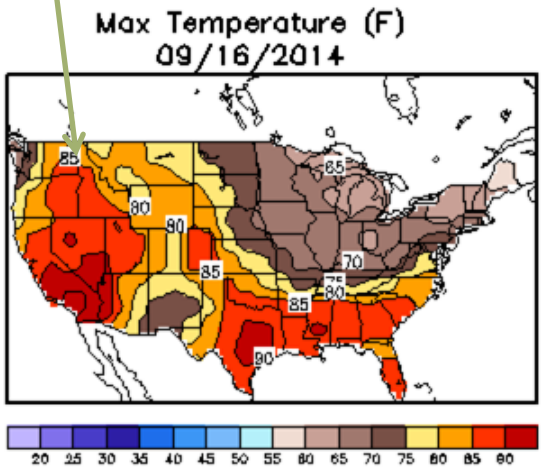
Tmax temperatures > 90 degrees F

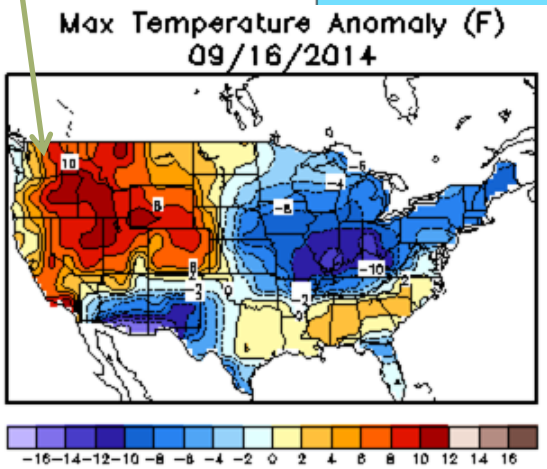


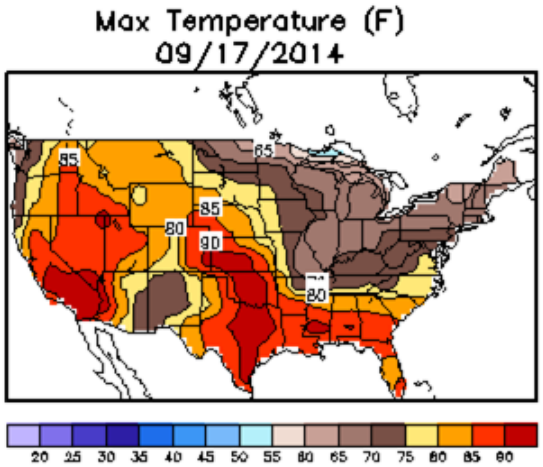
Verifying Obs Valid 9/16/14 (D8/D9)

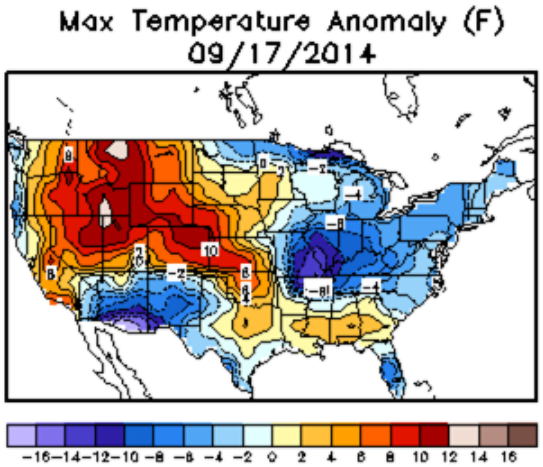
85-90 deg F

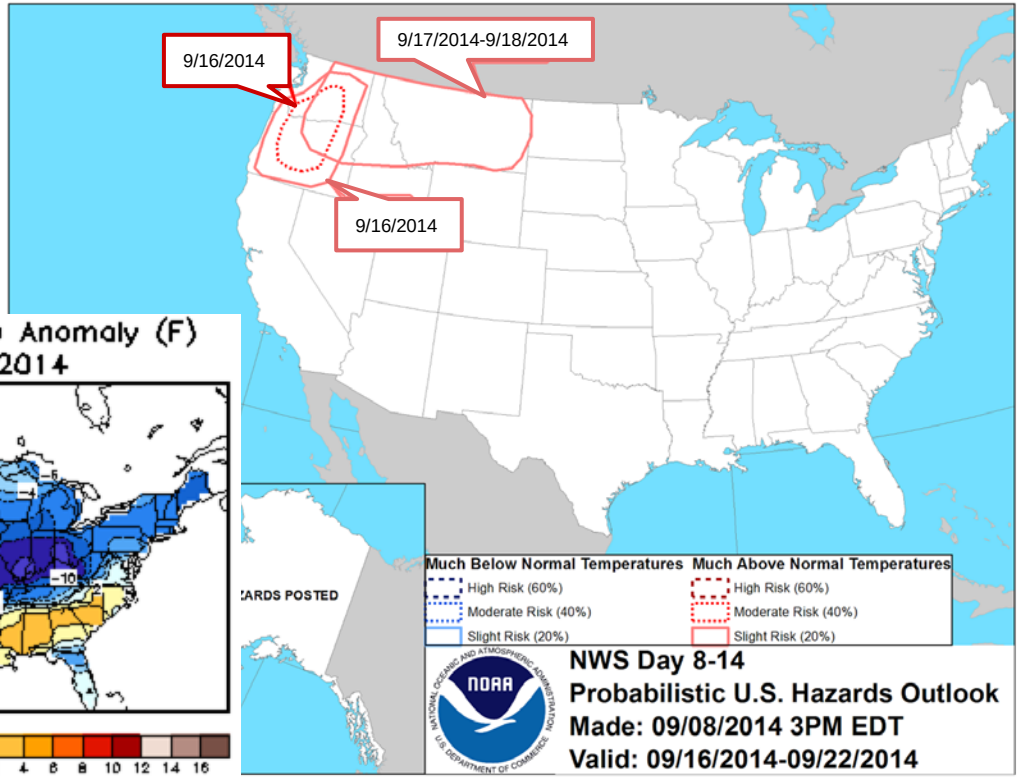
10-12 deg F











Summary

- CPC has a new week-2 probabilistic extremes forecast tool that is being used to produce daily manual week-2 hazards forecasts
- Probabilistic format enables more week-2 hazards to be put on the map
- Initial verification over 2013 for minimum temperatures show positive skill, good reliability
- Good frequency of Tmin probabilities in the extremes tool that reach our criteria for winter months

Ongoing/Future Work

- Phased approach for features (variables, thresholds)

- More models included in the tool, weighted consolidated forecast

- More tool evaluation
  - Tmax
  - Alaska verification
  - Spread vs. skill, calibrated vs. uncalibrated
  - Other percentiles

http://www.cpc.ncep.noaa.gov/products/predictions/threats/threats.php

National Weather Service
Climate Prediction Center

Home | Site Map

HOME > Expert Assessments > Hazards Outlook

Briefing Page

Download KML
Day 3-7 Precipitation
Day 3-7 Temperature
Day 3-7 Soils
Day 8-14 Precipitation
Day 8-14 Temperature
Day 8-14 Soils

Download Shapefiles
Day 3-7 Precipitation
Day 3-7 Temperature
Day 3-7 Soils
Day 8-14 Precipitation
Day 8-14 Temperature
Day 8-14 Soils

Search the CPC
Go

About Us
Our Mission
Who We Are

Contact Us
CPC Information
CPC Web Team

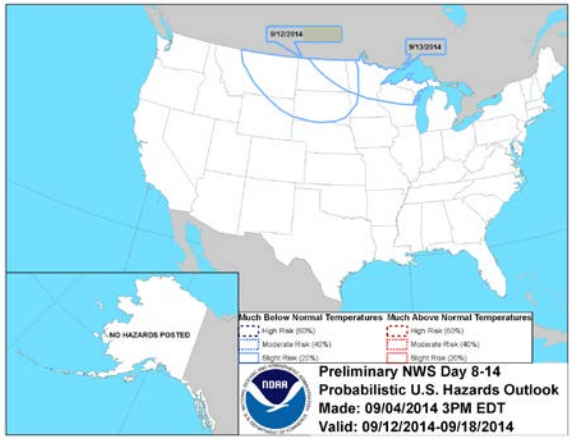
U.S. Hazards Outlook - Made October 02, 2014

	Days 3-7	Days 8-14	Prob. Days 8-14
Precipitation	☑	No Hazards	Not Available
Temperature	☑	No Hazards	No Hazards
Soils	☑	☑	Not Available

Legend

- High Winds
- Significant Waves
- Severe Weather
- Heavy Rain
- Freezing Rain
- Heavy Snow
- Heavy Precipitation
- Much Above Normal Temps
- Much Below Normal Temps
- Excessive Heat
- Flooding Likely
- Flooding
- Flooding Possible
- Enhanced Wildfire Risk
- Severe Drought

Categorical Outlooks	Day 3-7	Day 8-14
8-14 Day Probabilistic Outlooks	Temperature Hazards	Precipitation Hazards



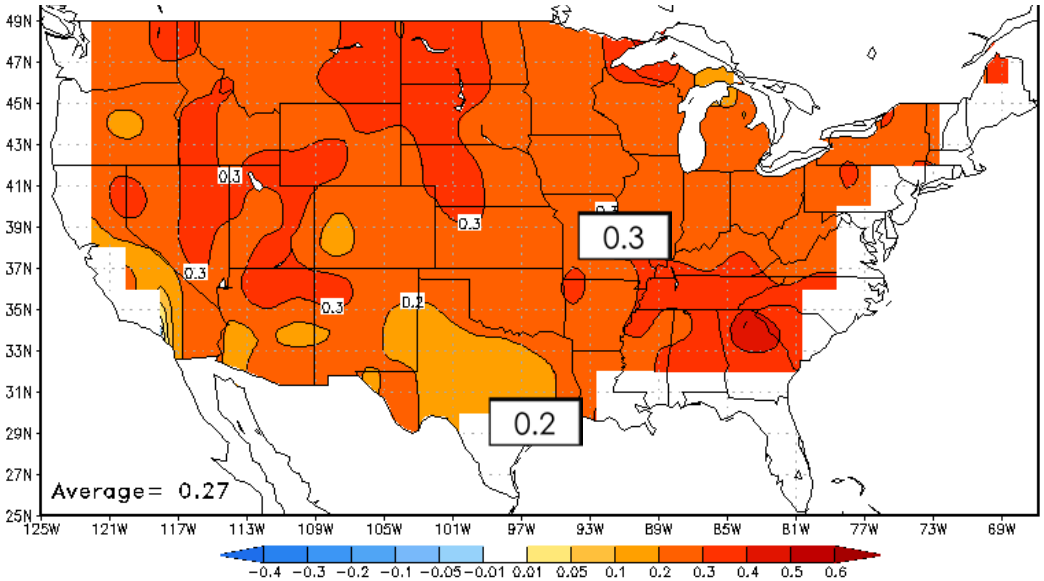
Thanks!

Questions & Comments / Feedback

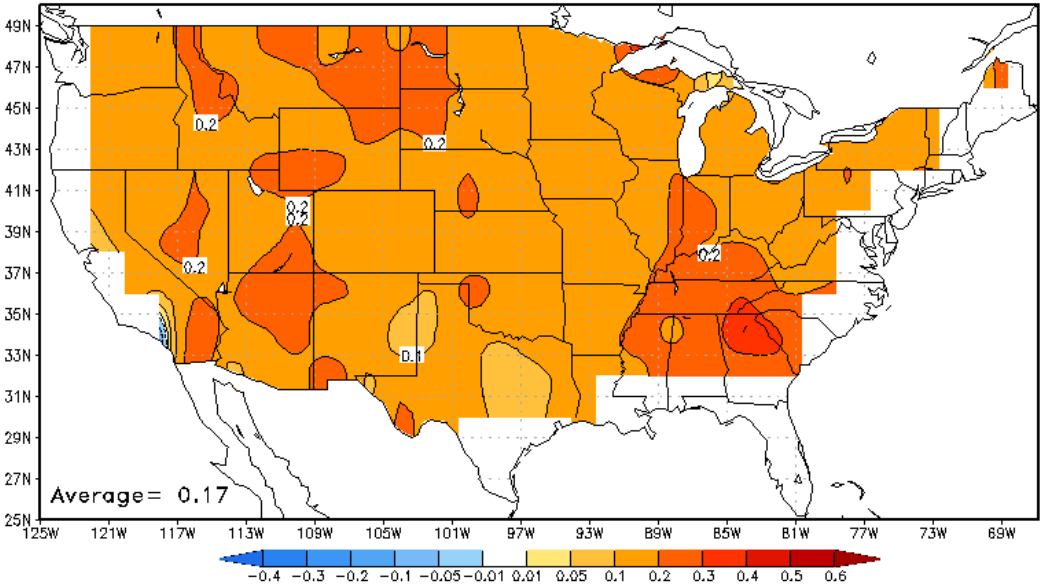
Questions / Comments

Verification - RPSS of combined categories

Day 8 Tmin RPSS, Combined Categories 1/21/13-12/31/13

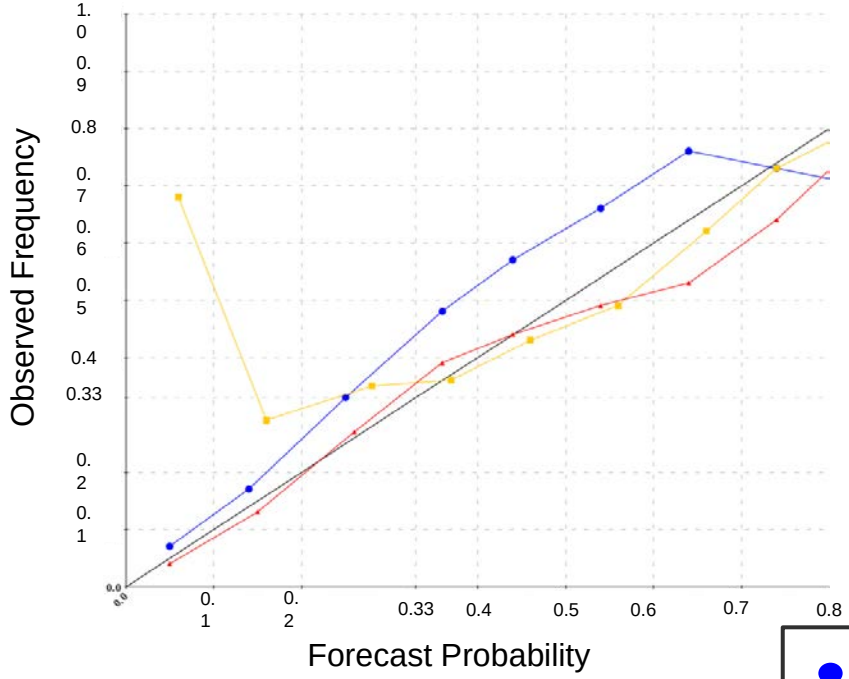


Day 14 Tmin RPSS, Combined Categories 1/21/13-12/31/13

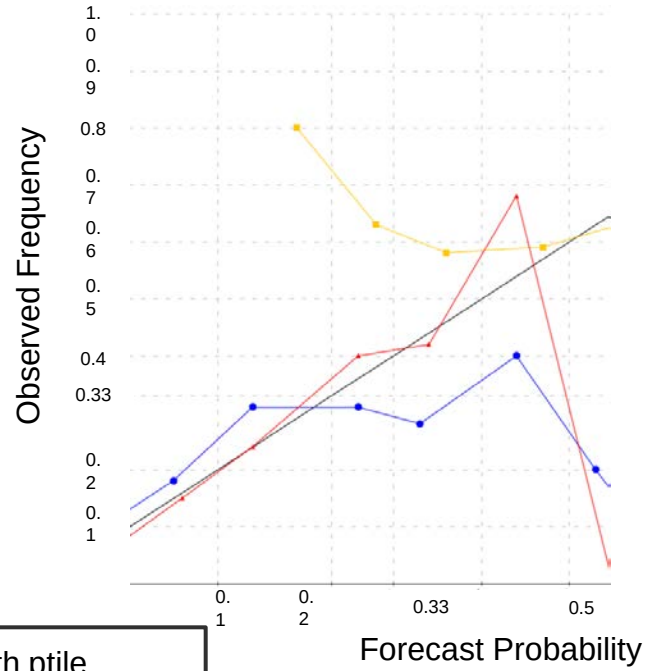


- Positive RPSS across entire CONUS for this period, relatively uniformly
- Less skill in Southern Plains for both leads, more skill across parts of the SE

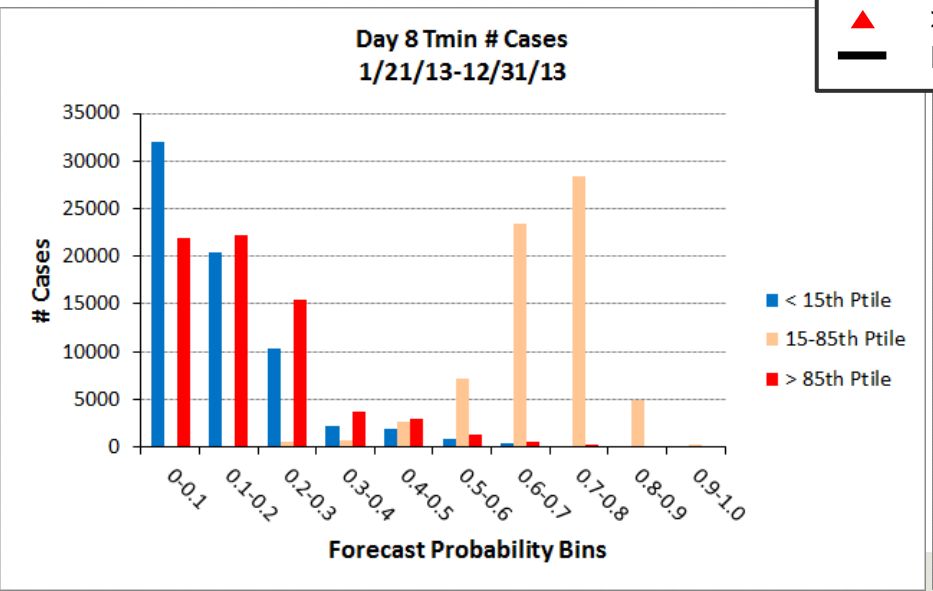
Day 8 Tmin 3-Category Reliability All 2013

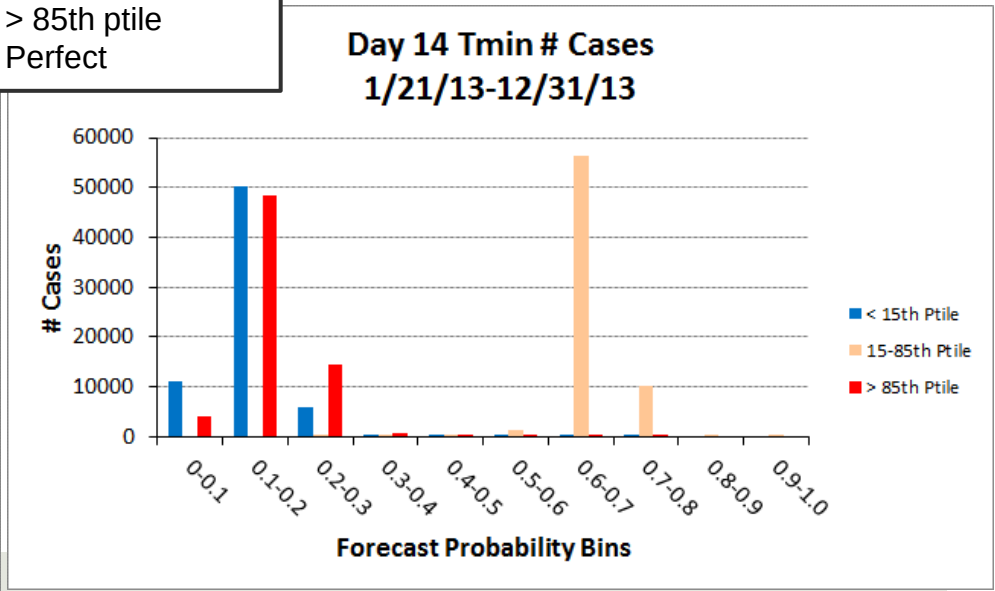


Day 14 Tmin 3-Category Reliability All 2013



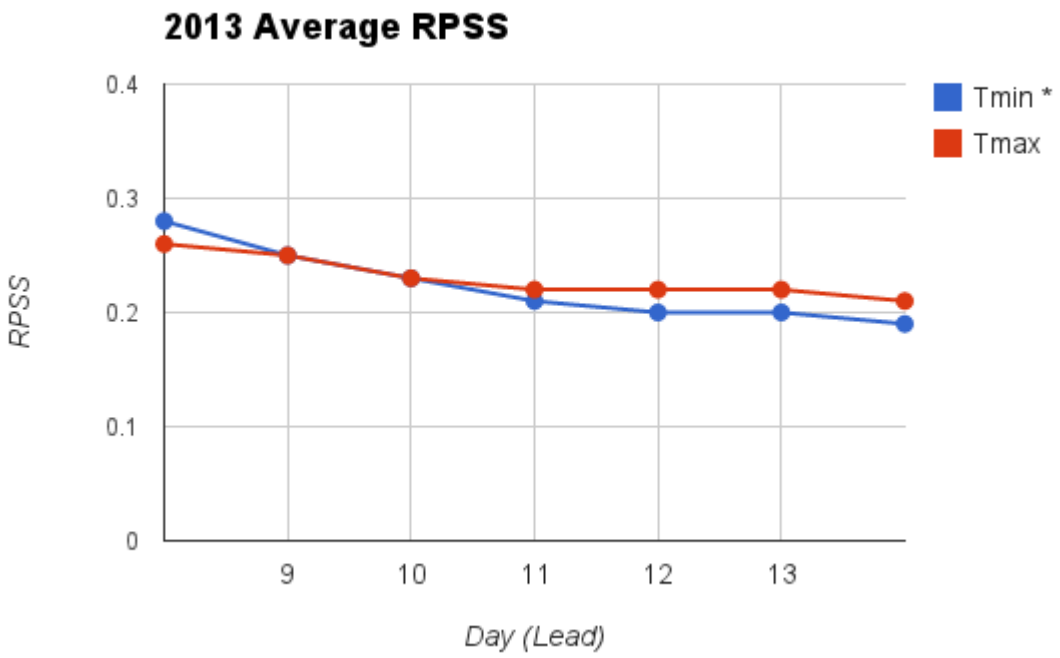
- < 15th ptile
- 15th - 85th ptile
- > 85th ptile
- Perfect



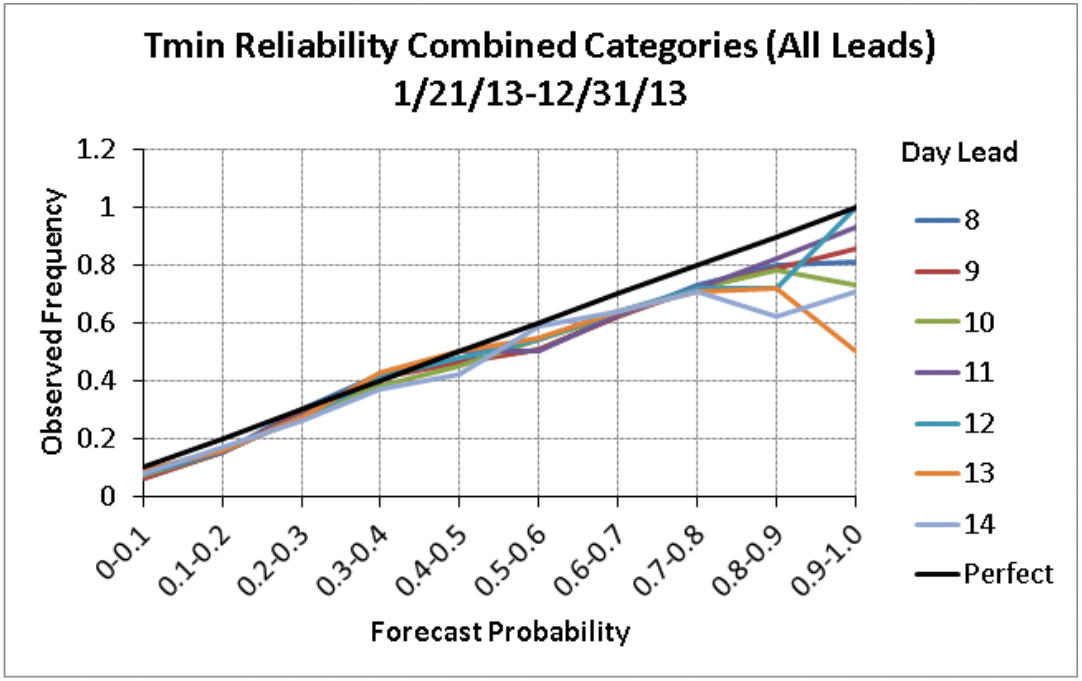


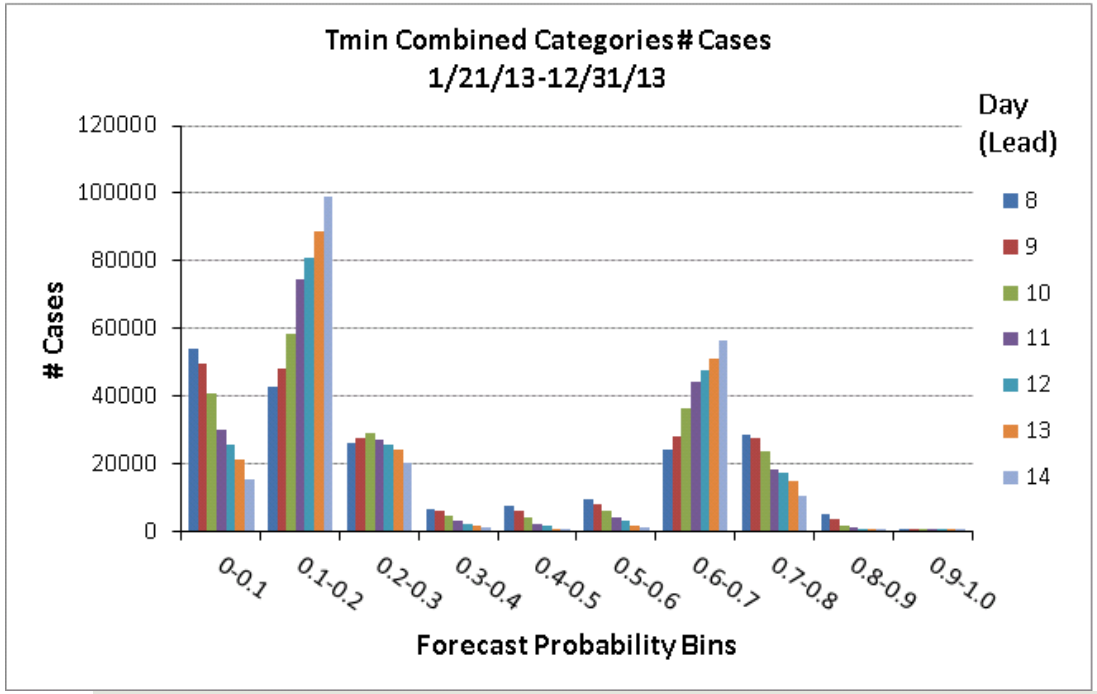
Verification

Table of RPSS averages for each lead

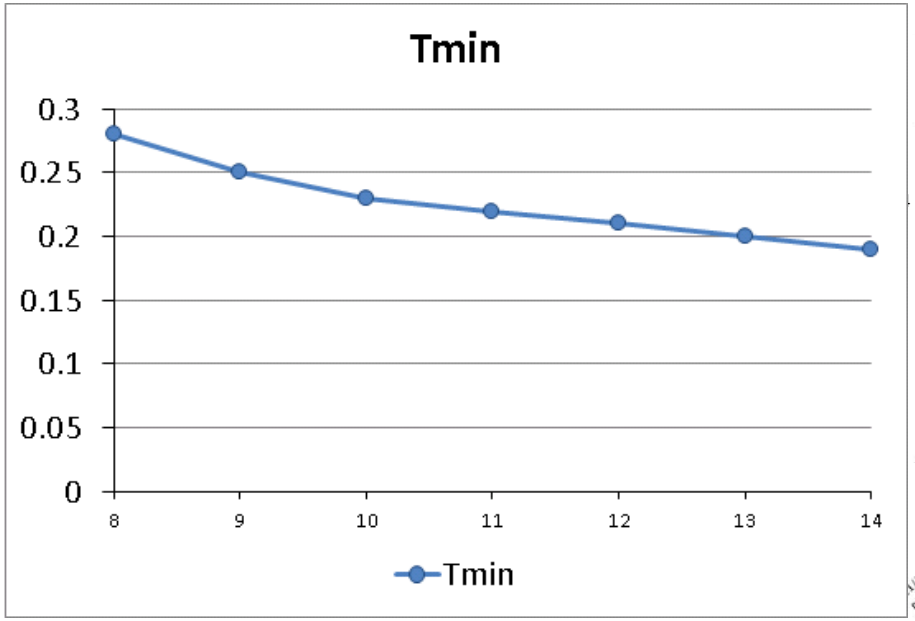


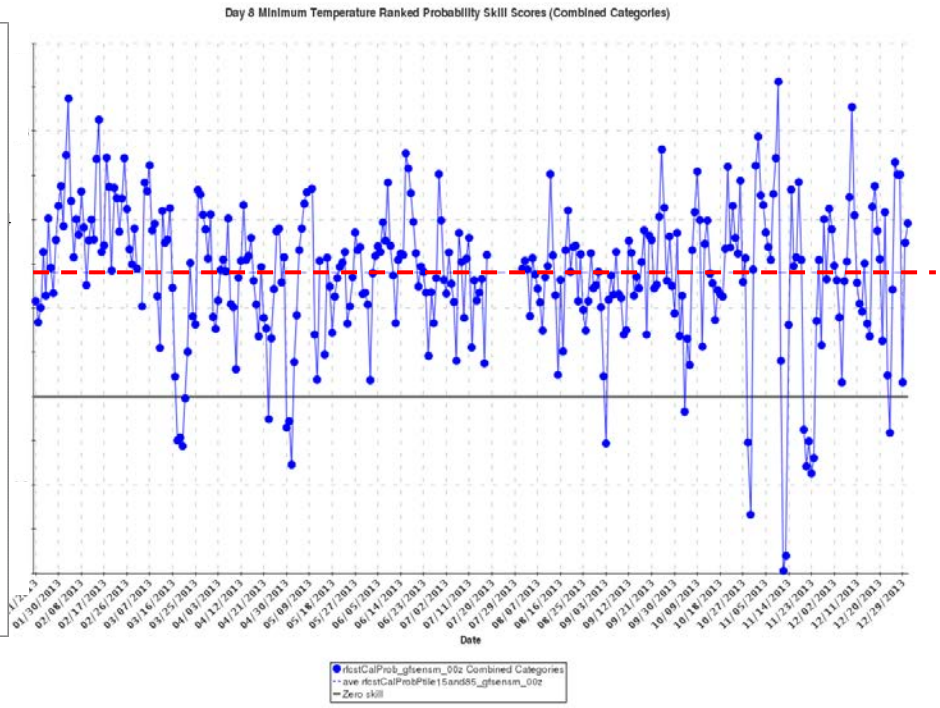
Reliability Combined Categories

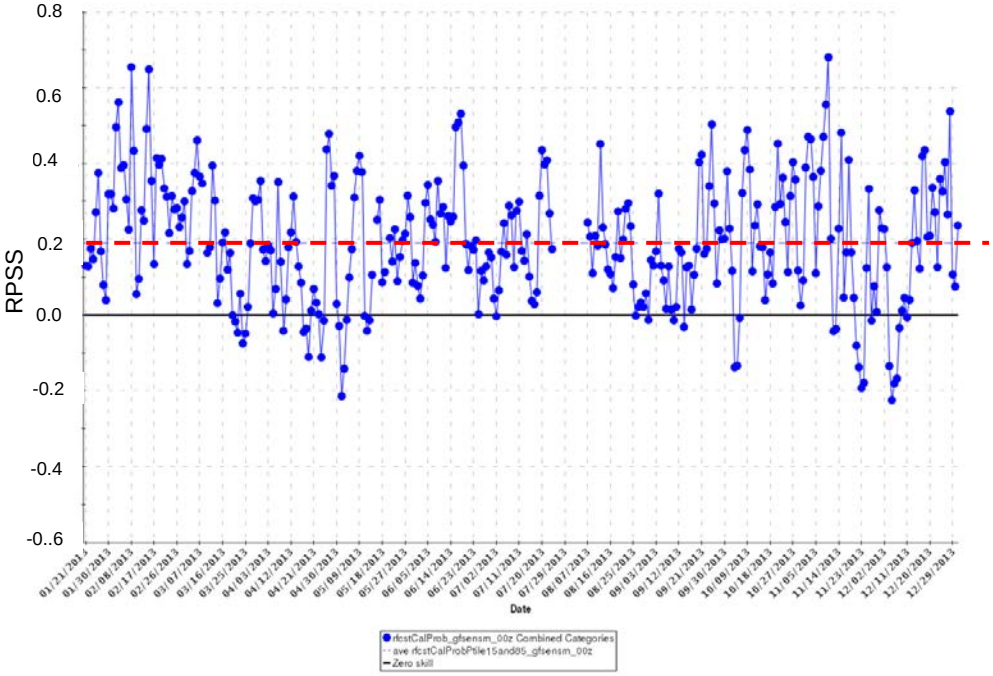




RPSS Combined Categories



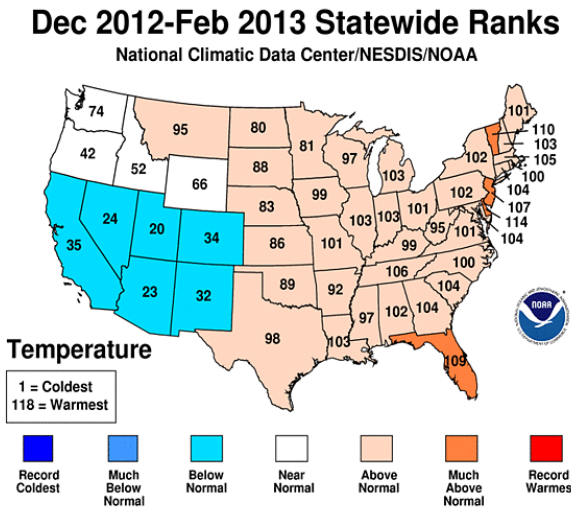


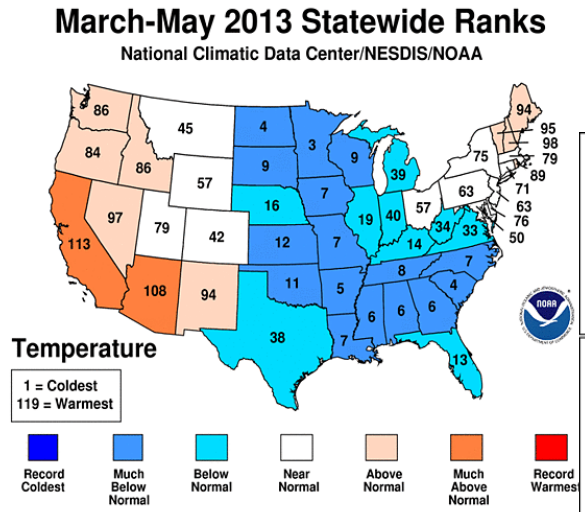


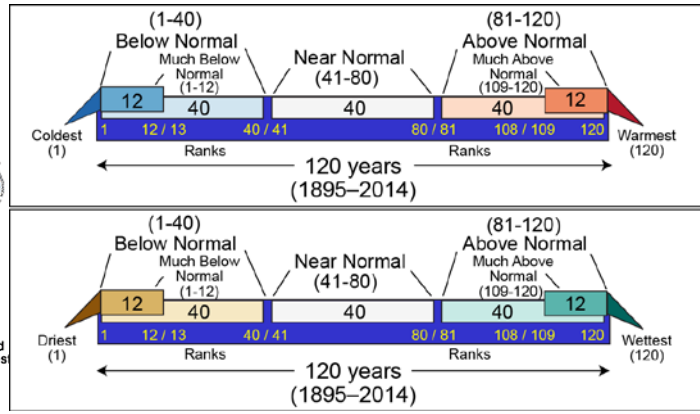
To Do:

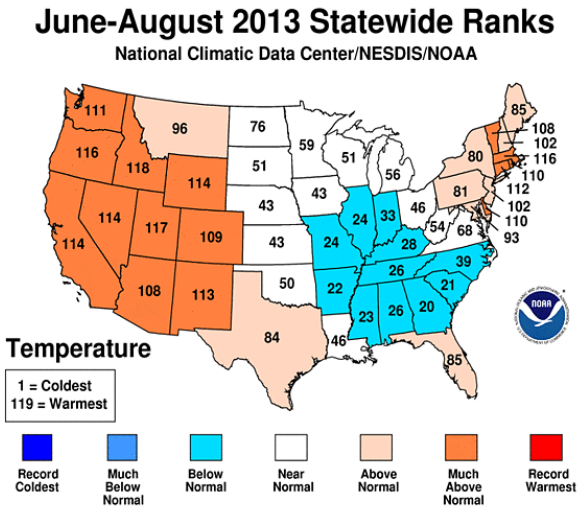
- look up RPS vs CRPS
- Choose dates of record mins/maxes to evaluate so Mike can make category plots of fcst vs obs
- CA tied 12th warmest year, below-average annual temps N. Plains, Cental Plains, Midwest, SE
- AK Extremes? not in vwt yet

calib vs uncalib plots and skill









http://www.ncdc.noaa.gov/monitoring-references/dyk/ranking-definition

Much above : 10th percentiles
Above/below: equal terciles (33.33rd percentile)

Question arose regarding why the reliability diagrams show that show that at longer leads (ie D14), the probs were overconfident but good skill, whereas D8-9 were underconfident, poor skill. Shouldnt calibration fix this at all leads?

Discussion followed with Dave, Dan, and Steve as to why this may be true :
We expect forecasts in general to be underconfident. At earlier time leads (d8/9), when the ens probs are increased relative to climo (To produce a skillful forecast hopefully), ens regr doesnt widen the distribution about the members at the edge relative to near ensemble mean (ensemble spread not taken into consideration), therefore the mean value may be better predicted (increased probs), but at the tails probs still too low since spread not adjusted to boost probs at tails.

There is likely less calibration performed at shorter lead times (ie. D8/9) than longer (D14) because there is higher skill at shorter lead times (ie the correleation values the ens regr uses to estimate how much correction to do?).

By week-2, lower model skill causes ens regression to “pay less attention” to individual members and rely more on wide kernels to get the distribution. So, ens regr applies more spread typically, widening kernels to estimate distribution which then happens to lead tomore model forecasts clustering at tails which is misleadingly overconfident (model not really forecasting it, just giving more spread to models making it more likely to overlap tails).

Also, this overconfidence at greater leads/under at shorter may be symptom of under-sampling. Model has more skill at shorter leads (d8) but theres not enough members to cover distribution accurately. Therefore, not enough samples to truly reflect potentially higher probabilities for the extreme forecasts when really the model would be more confident. At longer lead times, ie D14, individual members less important (as forecasts approach climos for skill), so wide kernel distribution is applied via ens regr to take care of undersampling (but skill not better due just to increasing the spread).

Calibration did increase probabilities for the extremes (at least for D8, check D14 calib. vs uncalib probs) correctly but not enough (since looking at reliability forecast probs at D8/9 still not high enough)

Talagrand diagram (rank histogram): Histogram of the ranking of ensemble members (lowest to highest forecast values). It is not necessarily a verification method, but rather a diagnostic tool to evaluate the **spread** of an ensemble.

You take a handful of runs of your mode for each timestep/gridpt, then calculate how the OBS rank relative to the model ensemble, then plot histogram of the ranks. For each specific forecast, the bins are determined by ranking the ensemble member foreasts from lowest to hightest. For each specific forecast, the bins are determined by ranking the ensemble member forecasts from lowest to highest. The interval between each pair of ranked values forms a bin. If there are N ensemble members, then there will be N+1 bins. The outer bins, lowest and highest – valued, are open-ended. Rank histograms are prepared by determining which of the ranked bins the observation falls into for each case, and plotting a histogram of the total occurrences in each bin, for the full verification sample. It is desirable to use a large sample of cases so that there is likely to be some occurrences in each of the bins. The example on the left below shows a single case where the observation (temperature) falls in the third bin. An example of a finished rank histogram for a 20 member ensemble (21 bins) is shown on the right.

If it is u-shaped, then variance is too low (obs are ranked high or low too often), if histogram looks kind of gaussian, variance is too high (obs rarely ranks high or low), if histogram if flat then variance is spot-on (obs have similar variance to ensemble).

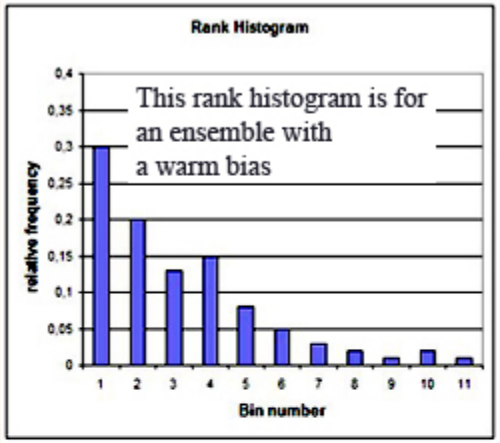
Undispersive ensemble - Rank histogram has u-shape, obs constantly verify at tails of ensemble distribution meaning the spread is not wide enough to capture the majority of events in the center of the distribution. Most common case for models.

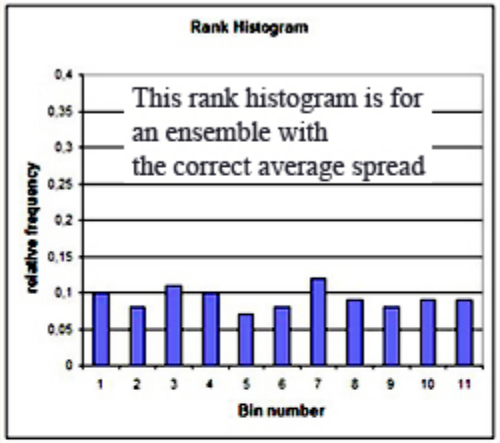
Overdispersive ensemble - Rank histogram has normal shape, obs constantly verify in the center and not frequently enough at the extremes. In practice this **doesn't frequently occur.**

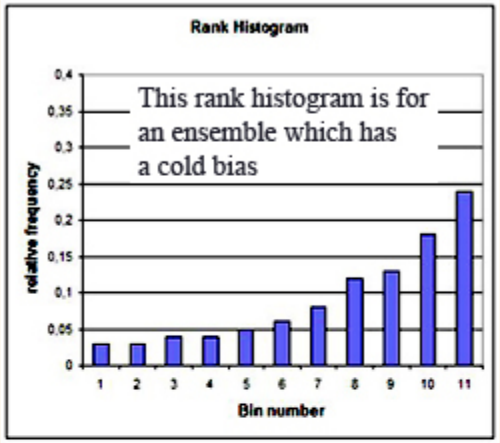
http://stats.stackexchange.com/questions/32370/what-pdf-should-be-fit-to-a-rank-histogram
http://www.eumetcal.org/resources/ukmeteocal/temp/msgcal/www/english/msg/ver_prob_forec/uos4b/uos4b_ko1.htm

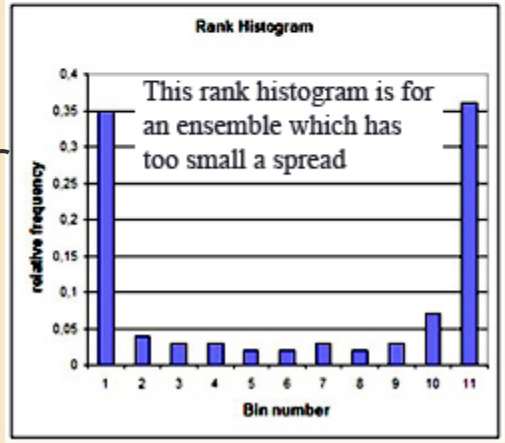
** Dave: Talagrands somewhat unfair at ends because the bins at ends include anything outside the bin range. It not be completely accurate because they "steal" the entire upper part of the pdf from the next-most extreme.

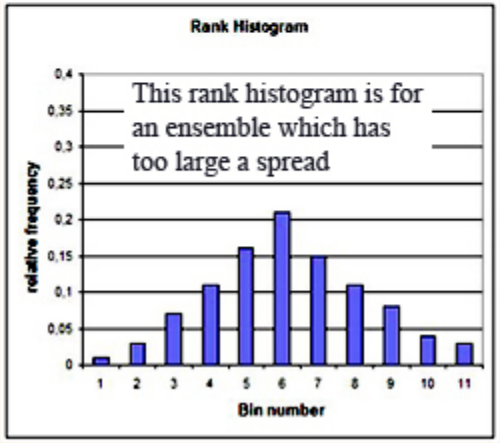
Correct. This is the most common form of the rank histogram for operational ensembles –they tend to be *underdispersive*.



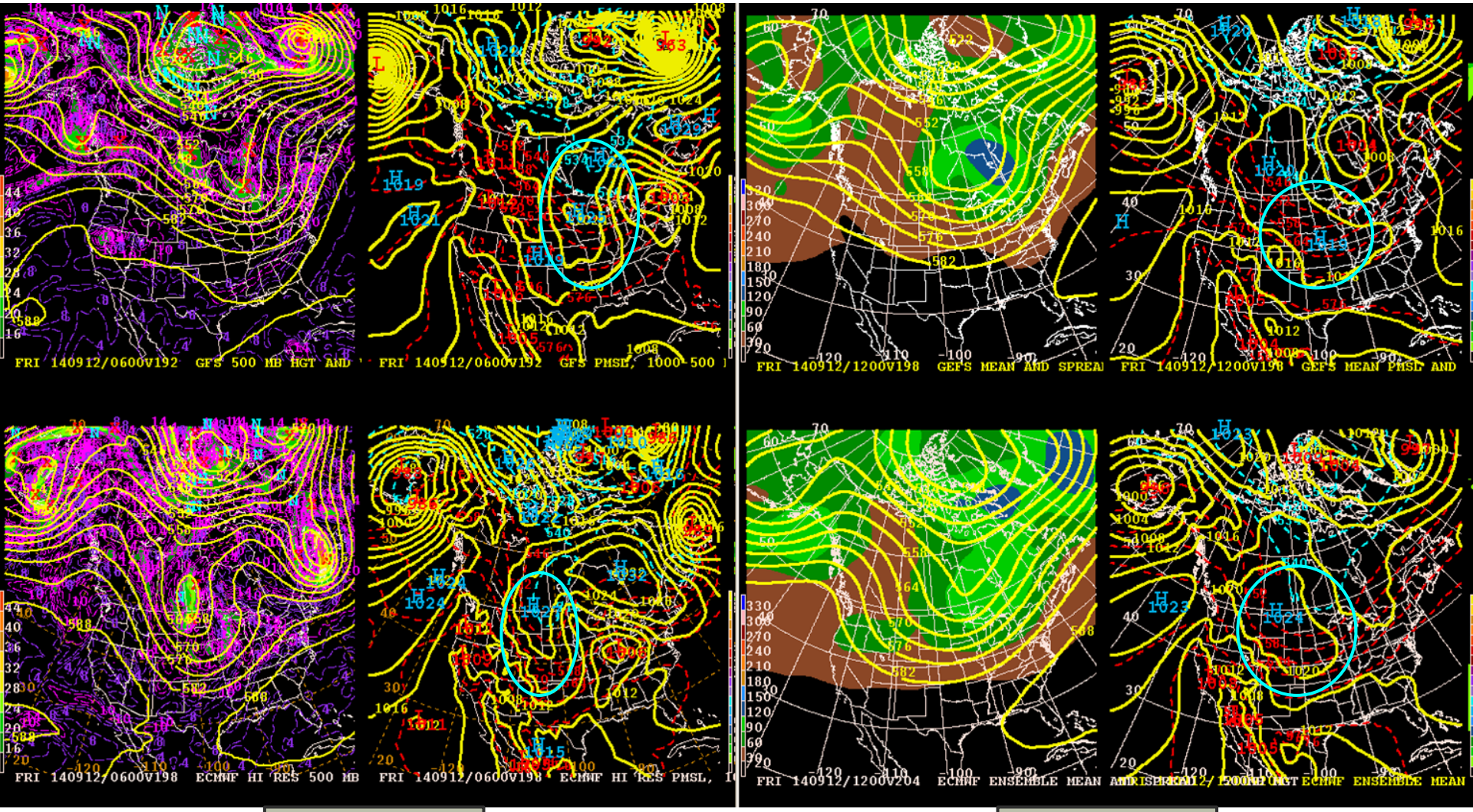








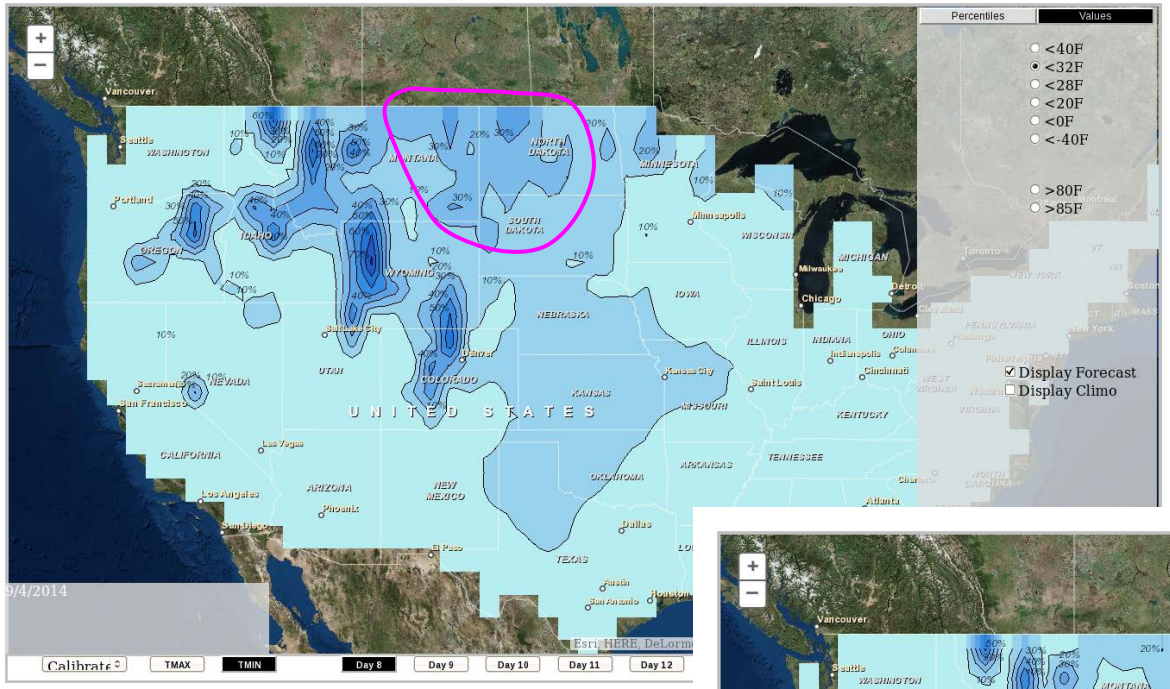
Sample Case Valid 9/12/14 (D8)

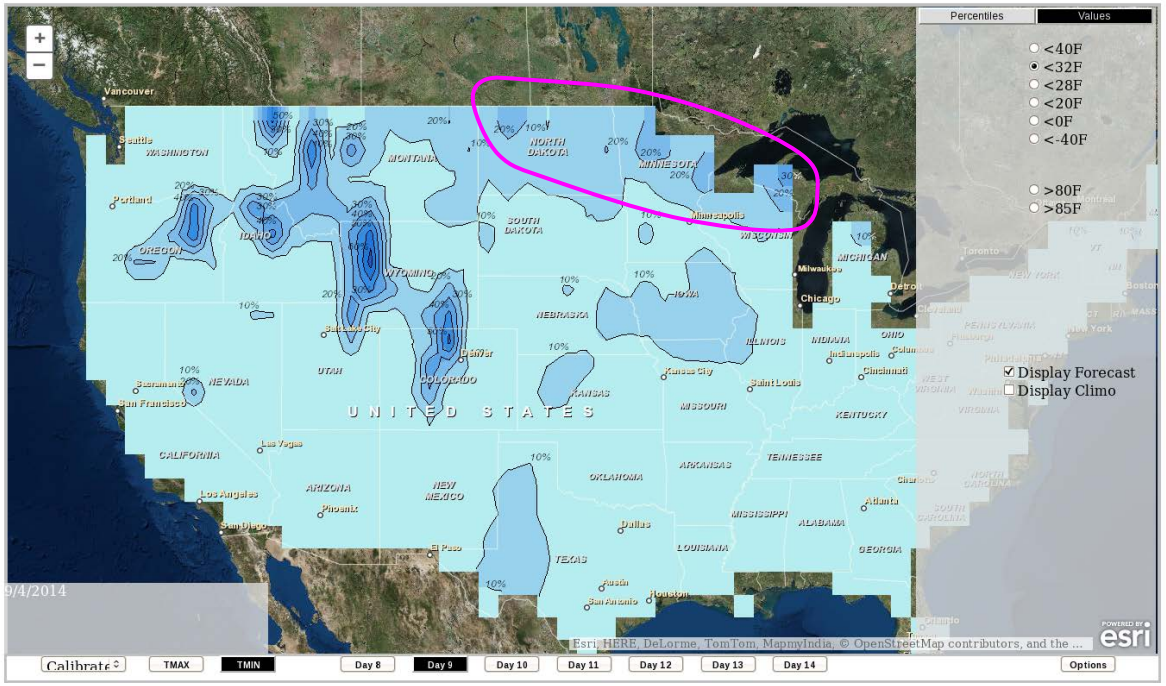


9/12 (D8) 06Z

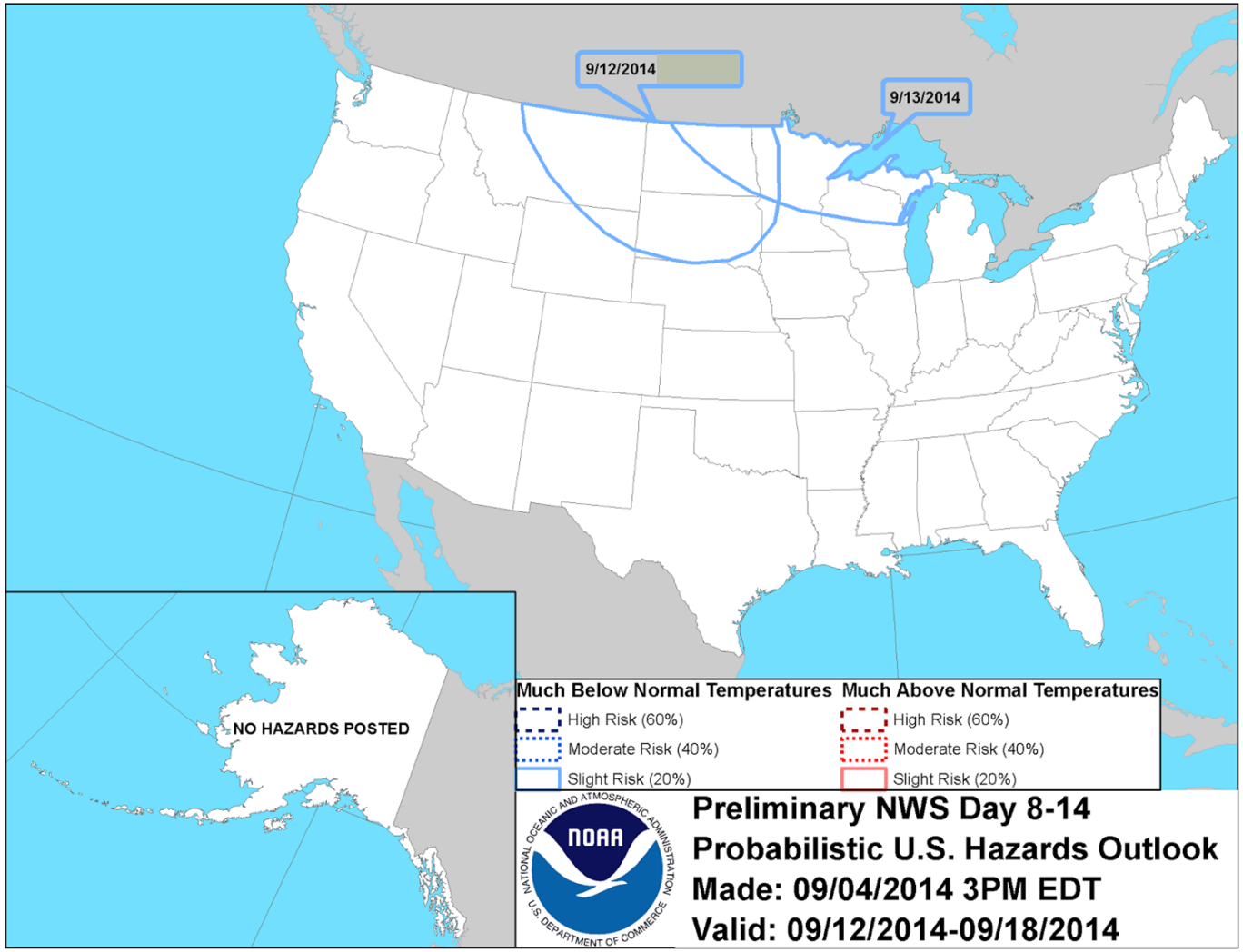
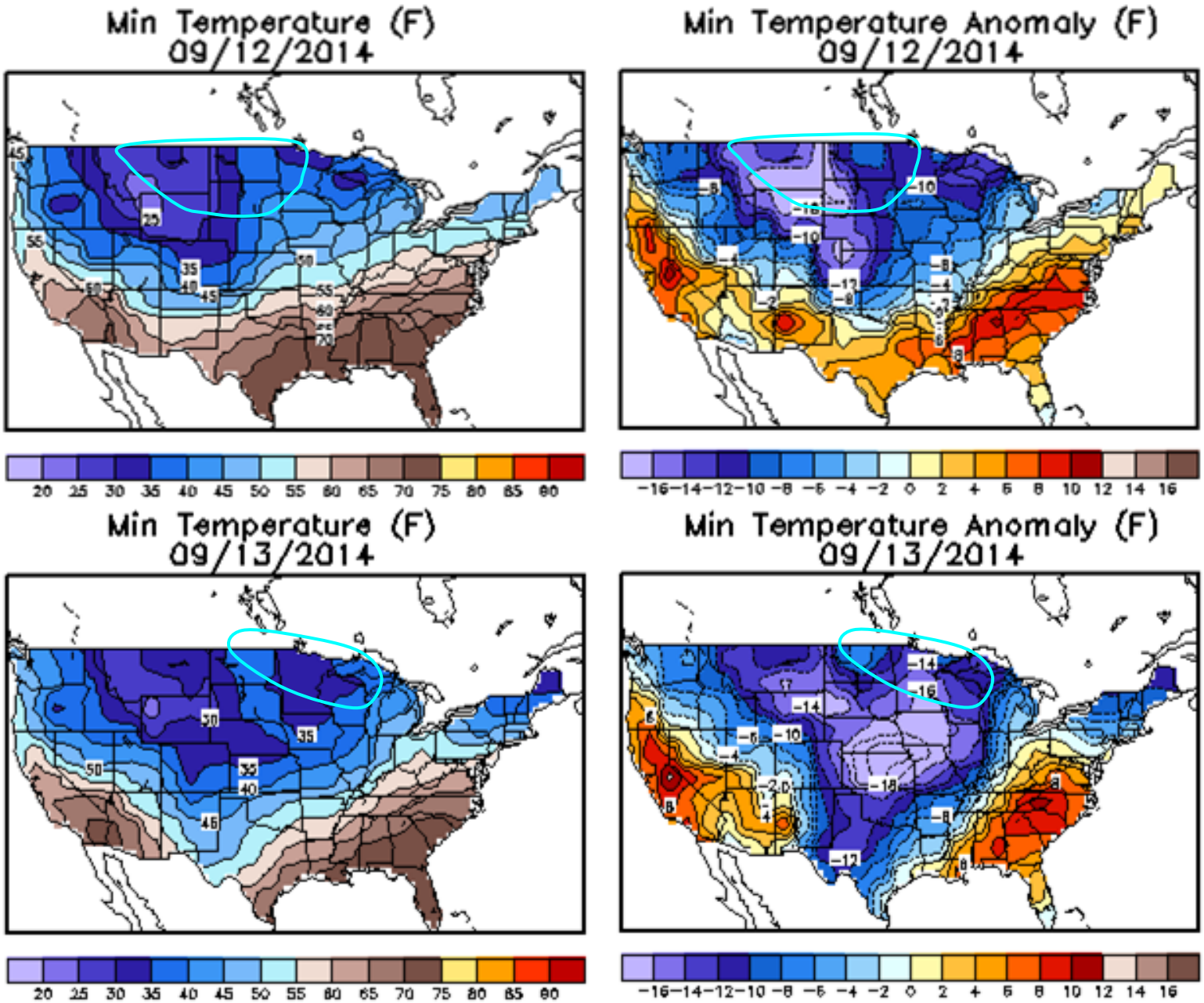
9/12 (D8) 12Z

Tool Valid 9/12/14-9/13/14





Issued 9/4/14


9/12/2014
9/13/2014
NO HAZARDS POSTED
Much Below Normal Temperatures
High Risk (60%)
Moderate Risk (40%)
Slight Risk (20%)
Much Above Normal Temperatures
High Risk (60%)
Moderate Risk (40%)
Slight Risk (20%)
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NOAA
U.S. DEPARTMENT OF COMMERCE
Preliminary NWS Day 8-14
Probabilistic U.S. Hazards Outlook
Made: 09/04/2014 3PM EDT
Valid: 09/12/2014-09/18/2014

Tmin observations 9/12/14-9/13/14