

Forecast Performance of the New Local 3 Month Temperature Outlook (L3MTO)

¹Andrea Bair, ²Marina Timofeyeva,
³Annette Hollingshead, ⁴Ellen Lay, ⁴Holly Hartmann,
¹David Unger, and ¹Robert Livezey

¹NWS/NOAA, ²UCAR, ³RSIS, ⁴University of Arizona

Outline

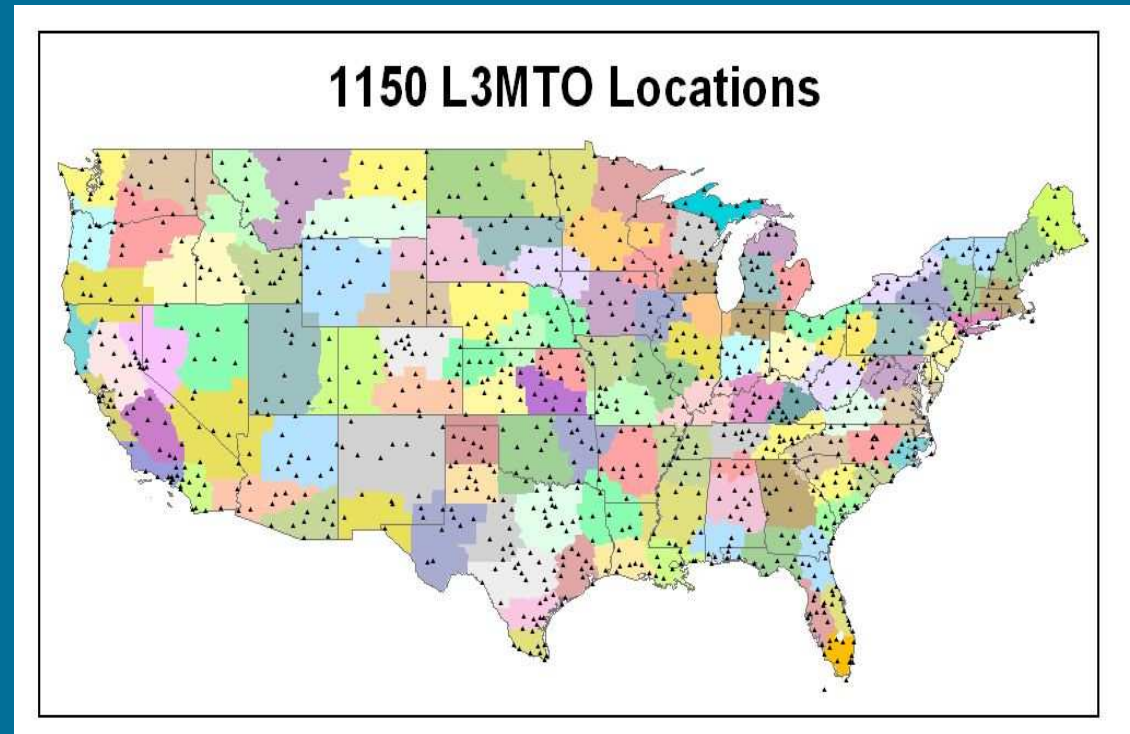
L3MTO Verification

- L3MTO Introduction
- Motivation
- Data
 - Creating a retrospective forecast archive for the stations
- Procedure
 - Methods used
 - Summary of current results
- Filter
 - Lessons learned

L3MTO Introduction

L3MTO:

- **First NWS operational local climate forecast product**
- **Available for 1150 locations, including AK and HI**
- **A new approach to local climate products**



<http://www.weather.gov/climate/l3mto.php>

Verification Motivation

- Two p

External (cu

- Guide cus
understan
performan
- Help cust
the foreca
used in th
applicatio

Three-Month Temperature Outlook (Issued: February 2007)

National Outlook Local Outlook Background Information Questions and Feedback

Seattle Tacoma, WA
King County, Coop ID: 457473
Elevation: 400 ft.
Latitude: 47° 27' N Longitude: 122° 17' W

NOAA Online Weather Data Help

Three Category Temperature Outlook: Nov-Dec-Jan 2007/8

Category	Probability
Above Normal (Higher than 42.1 deg F)	44%
Near Normal (40.3 - 42.1 deg F)	33%
Below Normal (Lower than 40.3 deg F)	23%

There is a 44.0% chance for the average temperature during this 3-month period to be higher than 42.1°.

There is a 33.0% chance for the average temperature during this 3-month period to be between 40.3° and 42.1°.

There is a 23.0% chance for the average temperature during this 3-month period to be lower than 40.3°.

Table Text Only

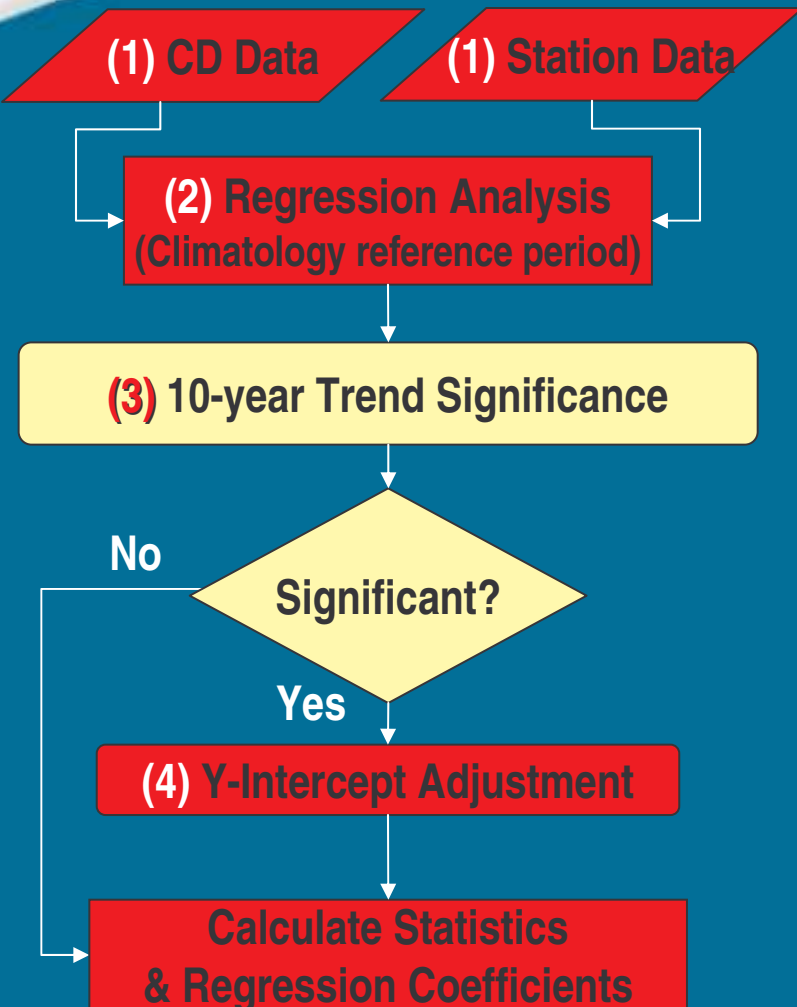
Select Location: Seattle Tacoma Select Time Period: Nov-Dec-Jan 2007, Select Product: Three Category Ou

Outlook Evaluation

ters with
the forecast

ters
are
actices need

Creating the Stations' Retrospective Forecast



1. CD and Station data used for each verification year
2. Regression analysis estimates the slope, intercept and correlation coefficient
3. Significance of the last 10-year trend is assessed by 2 tail sample t-test at the 95% confidence level. The trend is in the difference between station and CD temperature. The test compares the difference in the last 10 years and in the climatology years
4. Adjustment is done by adding the difference between station and CD mean, weighted by the last 10-year data, to the y-intercept, estimated in step 2

Verification Data

The archived station forecast created using:

- CPC forecasts from Dec. 1994-present (2005)
- Regression equation coefficients from:

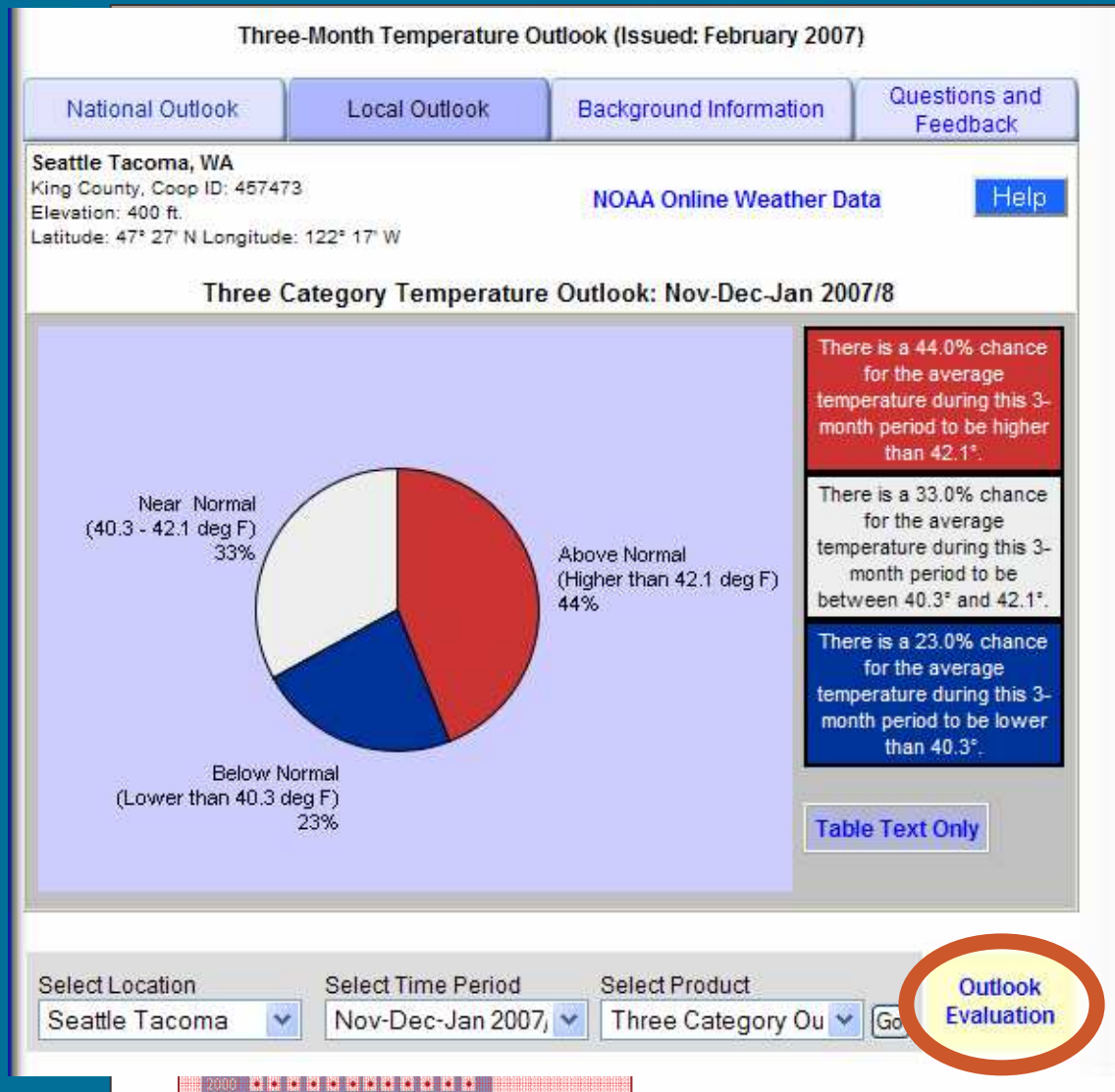
Verification Year	Climatology Reference Years	Current Trend Years
1995	1961-1990	1985-1994
1996	1961-1990	1986-1995
...
2001	1971-2000	1991-2000
...
2006	1971-2000	1996-2005

Verification Scores

(Available from all L3MTO Webpages)

Univ. of AZ verification tool expanded to include:

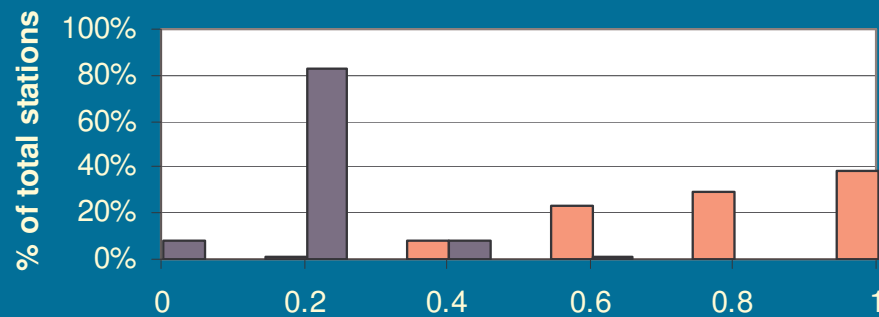
- L3MTO forecast archive
- Additional scores to coordinate with NWS verification practices
- Extended selection choices for years and leads



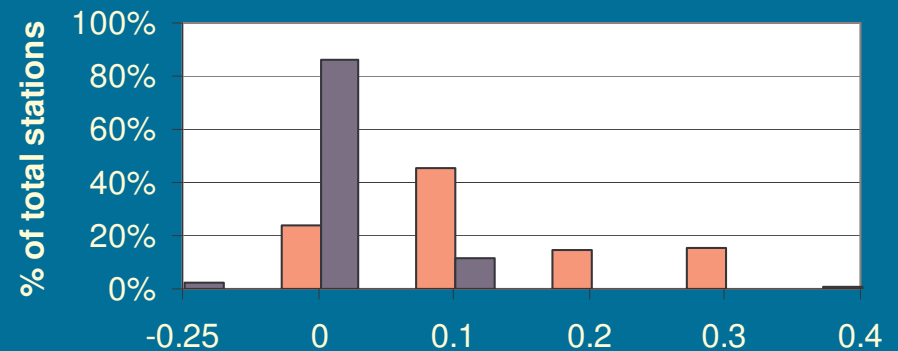
Verification Scores

(Available on all L3MTO webpages)

Probability of Detection



Brier Skill Score



■ Warm

■ Cool

■ Warm

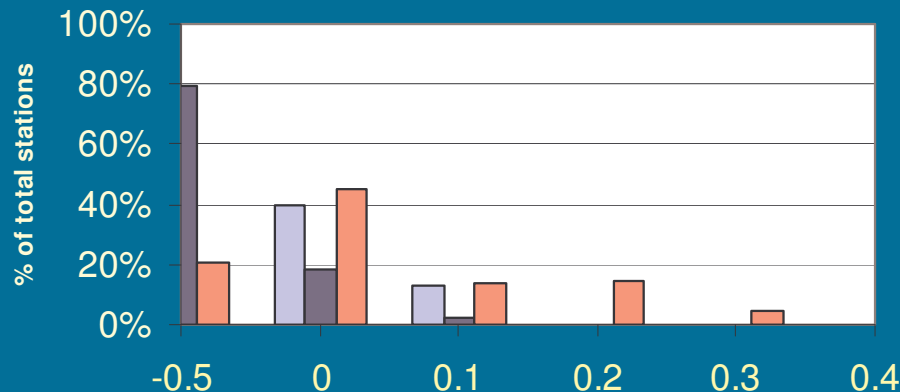
■ Cool

Forecast Evaluation of Categories shows the forecast for the Warm (Above Normal) category is performing better than the forecast for the Cool (Below Normal) category.

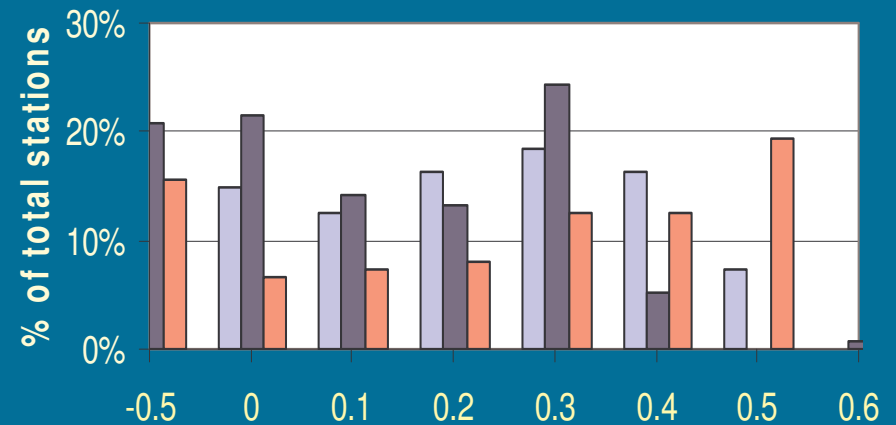
Verification Scores

(Available on all L3MTO webpages)

Ranked Probability Skill Score



Heidke Skill Score



■ All Seasons ■ Cold (OND-FMA) ■ Warm (AMJ-ASO)

■ All Seasons ■ Cold (OND-FMA) ■ Warm (AMJ-ASO)

Warm season forecasts (AMJ-ASO) show more skillful performance than forecasts for cold seasons (OND-FMA).

Verification Methods for NWS Forecast Guidance

- Utilizing several verification statistics:
 - Modified Heidke Skill Score (HSS)
 - Reliability Diagram
 - Continuous Ranked Probability Skill Score (CRPSS)
- Stratification by forecast target season
- Summary for different regions and leads

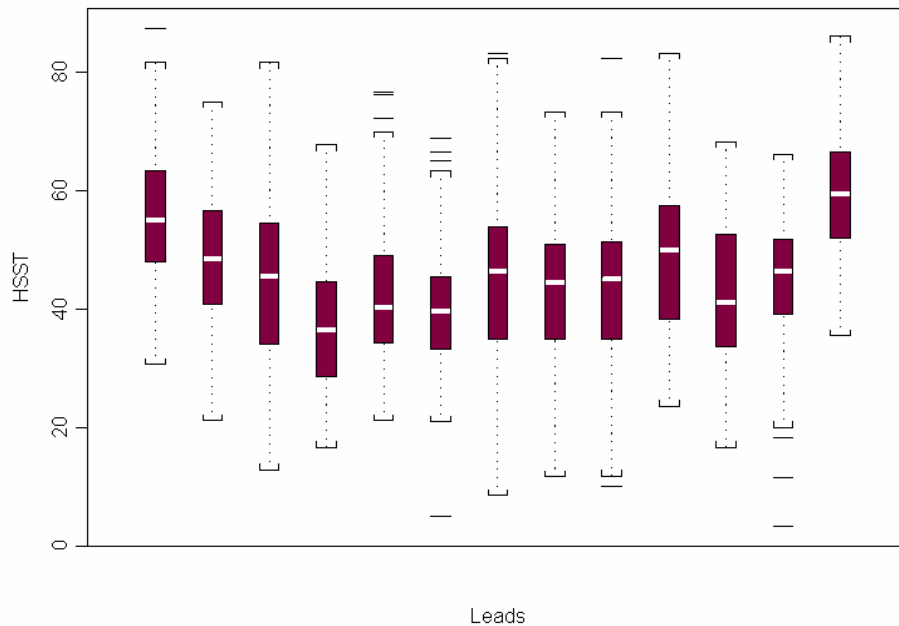
HSS definition

- **Modified Heidke Skill Score (HSS):** Measure of forecast accuracy. HSS assumes a stationary climate, therefore all observations fall into a climatological distribution, implying that 1/3 of the forecasts (EC or climatology), if randomly drawn from the same climatological distribution will be correct. EC in this case does not specifically account for the neutral (or above, or below) conditions observed. The score is computed as:

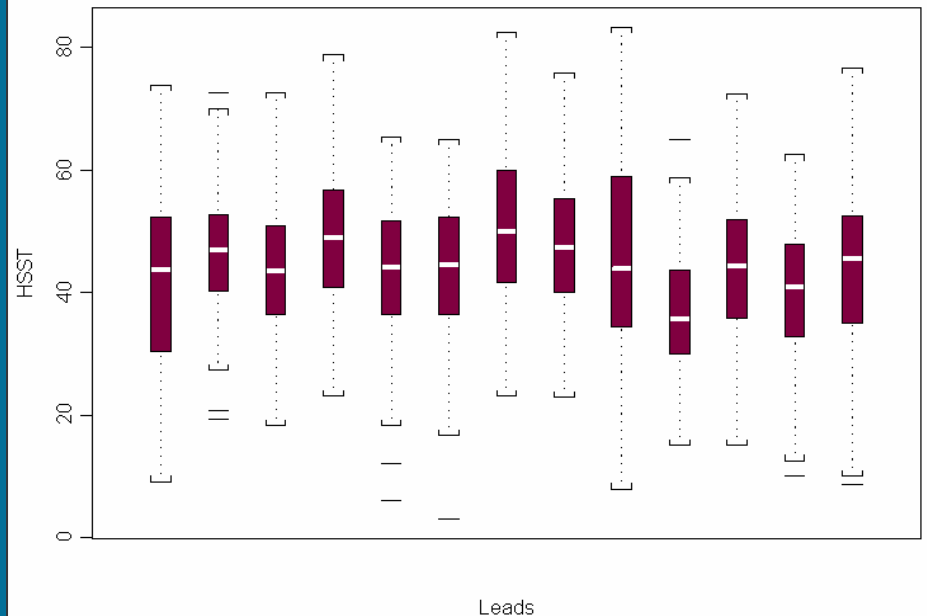
$$HSS = \frac{\text{Total Correct Forecasts} - 1/3 \text{ of Total Forecasts}}{2/3 \text{ of Total Forecasts}}$$

HSS (all stations) - Lead

DJF



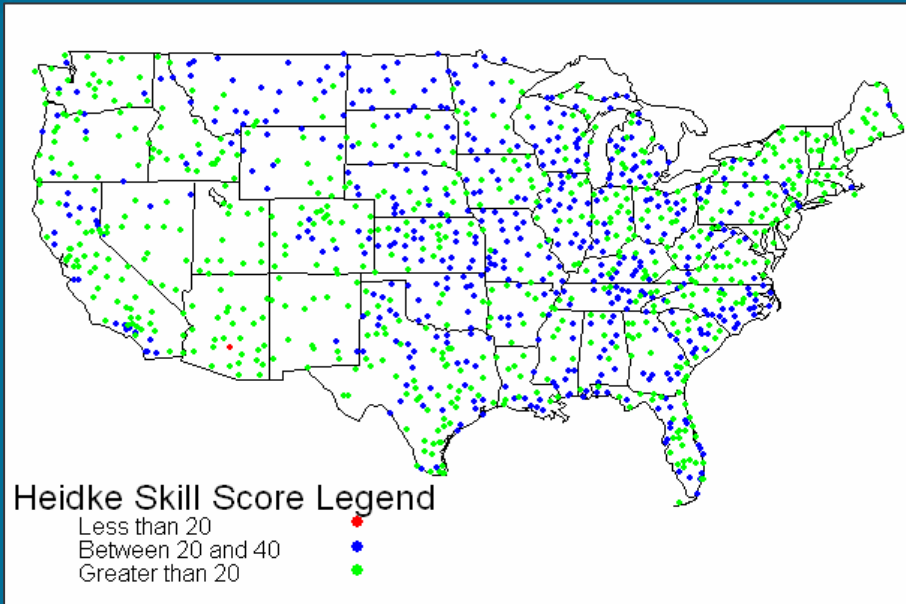
JJA



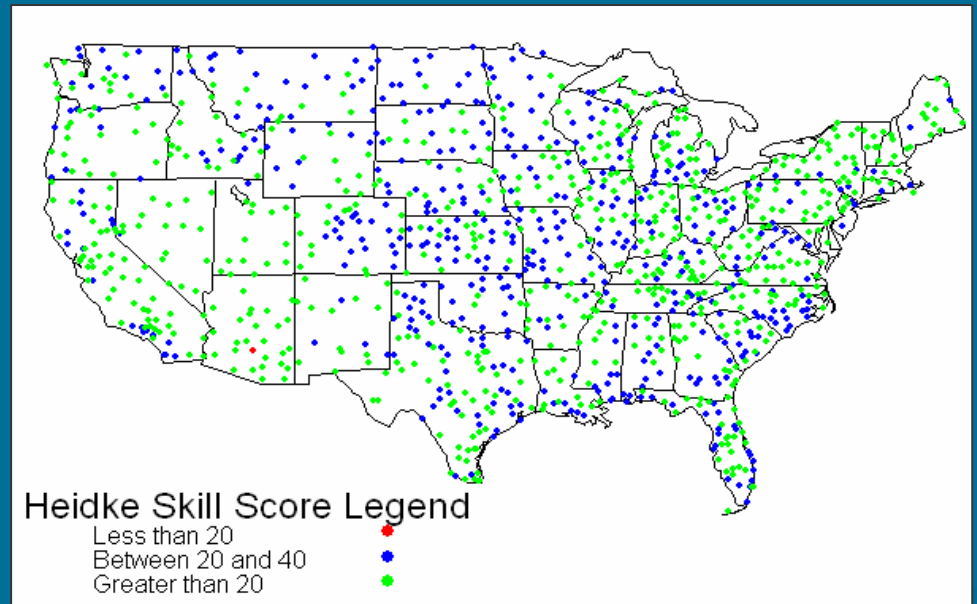
- Most stations have correct forecasts 30% (or greater) of the time, and have little variability between the 13 leads
- For all years analyzed together, shorter leads in the winter seasons have higher scores than the same in the warm seasons
- Larger spread between stations' performance in the warm seasons than cold

HSS - map

DJF



JJA



- For each target season, the maps show the HSSs for all leads and years combined
- Most stations in the Great Basin and Southwest have high (>40%) HSS
- Most variations present in the station performance in the coastal areas and mid-west

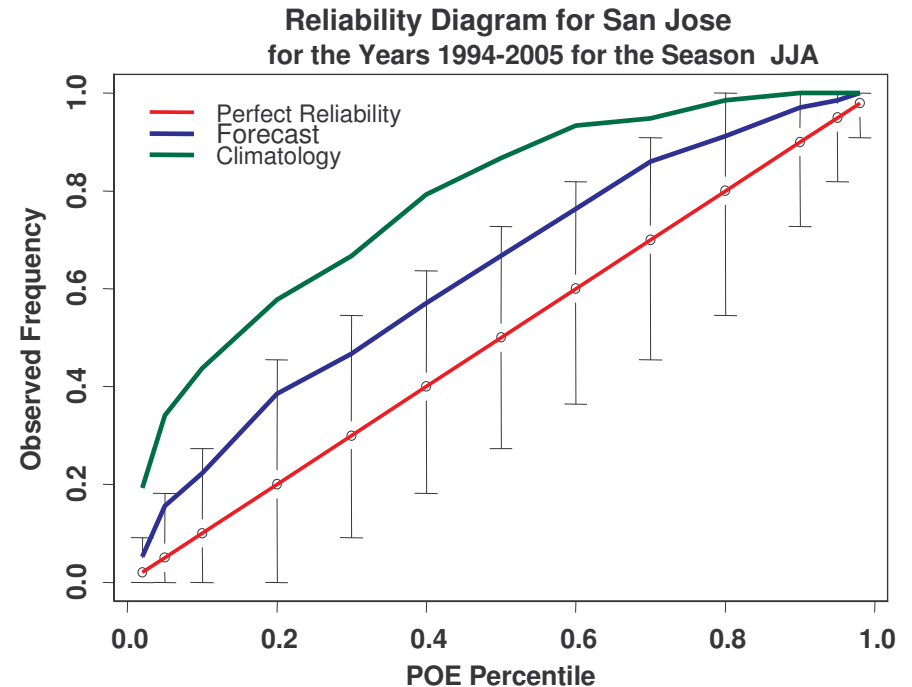
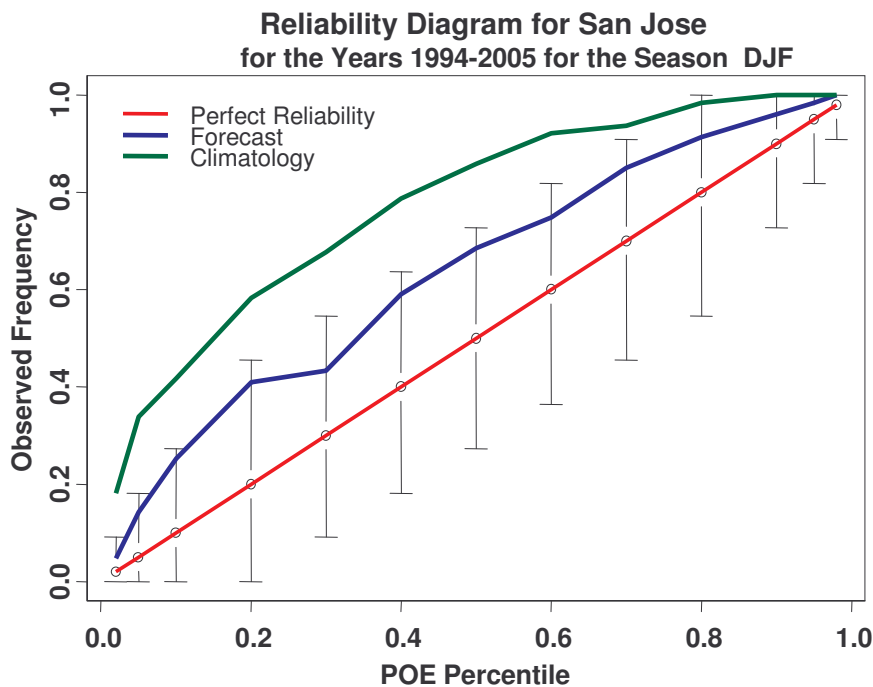
Reliability Diagram

For each station and each target season:

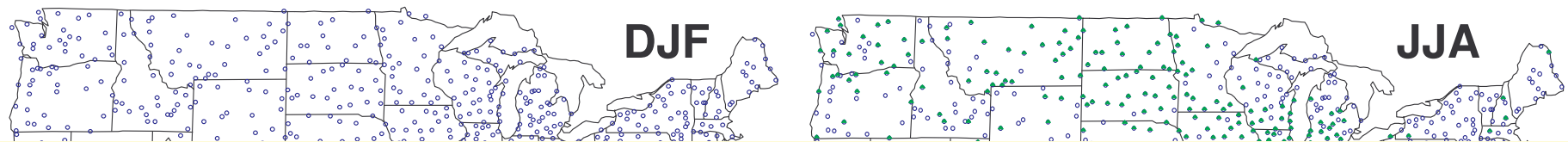
- **The forecast RD was compared with the corresponding climatology's RD, using 13 %tiles (98%, 95%, 90%, 80%,...,20%, 10%, 5%, 2%) and perfect reliability confidence intervals at 95%**
- **Summary maps show:**
 - **Forecast are fit to the confidence limits of perfect reliability**
 - **Distance measure of the forecast and climatology from the perfect forecast line**

Reliability Diagram

- For each station and forecast target season a reliability diagram was created to analyze each percentile's performance.
- Confidence limits around the perfect forecast line, were drawn from a binomial distribution at 95% confidence.
- The forecast was compared with 1971-2000 climatology performance



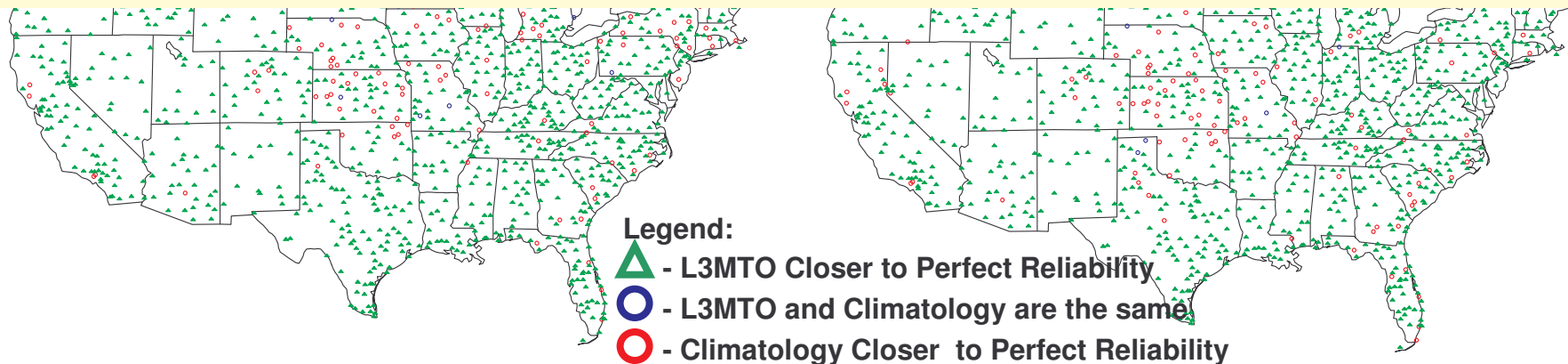
Reliability Diagram



The maps below show a comparison of the forecast performance to the 1971-2000 climatology, using a distance measure from the perfect forecast line.

In most cases, even the marginal forecasts are outperforming climatology.

Majority of cool season forecasts show more stations with marginal forecast results, while warm seasons show more stations with skillful performance.



Conclusions

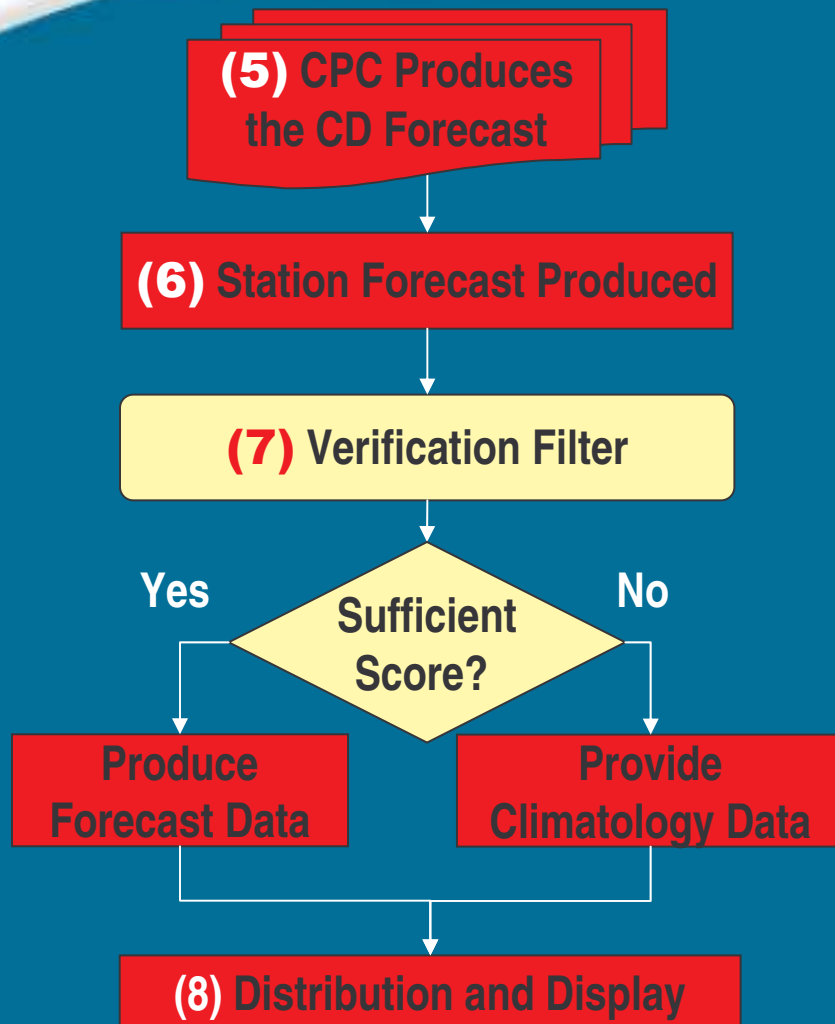
- **Using multiple verification scores in climate forecast evaluation provides the most comprehensive assessment of long term forecast performance**
- **Verification analysis identified locations and 3-month periods with satisfactory forecast performance, and their spatial and temporal variability**
- **Overall, for the majority of stations and seasons, the L3MTO indicates higher skill than the 1971-2000 climatology**

Lessons Learned

- Although L3MTO in general proves to be more skillful than 1971-2000 climatology, it is still important to identify station/season combinations where L3MTO does not have long term skill. At such stations/seasons a Verification Filter will be applied.
- The verification filter is the test that compares the correlation significance test probability (p-value) for the past forecasts and observations for each lead and season, with the passing p-value: 95% confidence. If the p-value is greater than passing, then the forecast is reported, if less, then 1971-2000 climatology is reported.

Verification Filter

Monthly Routine:



5. CPC produces CD forecasts monthly
6. CD outlook together with bi-variate statistics reported from the regression analysis, station forecast is produced
7. Verification filter is the test that compares p.value for correlation skill score for each lead and season with the passing p.value at 95% confidence. If p.value > passing, forecast is reported, if less climatology is reported
8. Data distribution and display are provided at the L3MTO web site



Thank You!!!