Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
42nd NOAA Annual Climate Diagnostics and Prediction Workshop
Norman, OK, 23-26 October 2017

Short-Term Climate Extremes: Probabilistic Forecasts from a Multi-Model Ensemble

Emily J. Becker and Huug van den Dool
Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD
and
Innovim, LLC, Greenbelt, MD

1. Introduction

This study explores the potential for probabilistic forecasts of extremes in monthly mean 2 m temperature and precipitation rate using the North American Multi-Model Ensemble (NMME; Kirtman *et al.* 2014), an ensemble of state-of-the-art coupled global climate models. Extremes are where the real impact of weather and climate are felt, yet there are currently very few forecasts for short-term climate extremes (STCE). Aggregate skill of deterministic forecasts of STCE (as assessed using the anomaly correlation) has previously been found to be higher than the aggregate skill of all forecasts (Becker *et al.* 2013; Becker 2017). This study examines the skill of a proposed forecast system providing probabilities of occurrence of the 15th and 85th percentile events, based on climatology.

2. Methodology

The NMME currently provides real-time guidance for NOAA's operational short-term climate forecasts, and includes a database of retrospective forecasts (1982-2010), used for bias correction, calibration, and skill studies. Seven models from the NMME contribute to this study: NCEP-CFSv2, Environment Canada's CanCM3 and CanCM4, GFDL's CM2.1 and FLOR, NASA-GEOS5, and NCAR-CCSM4. This study spans the hindcast and forecast period from 1982-2016. Verification data is obtained from the GHCN+CAMS station-to-grid 2 m surface temperature dataset (Fan and van den Dool 2008) and the CMAP precipitation rate dataset (Xie and Arkin 1997).

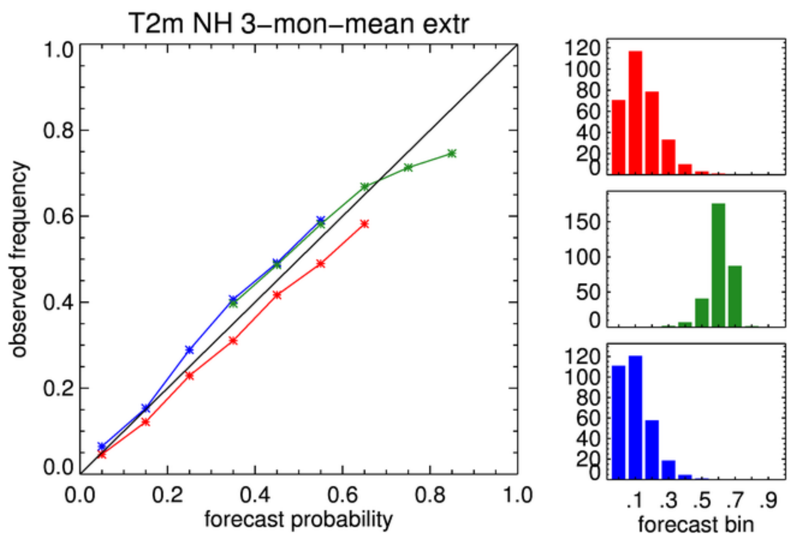

Fig. 1 2 m temperature lead-1 3-month-mean reliability diagram showing forecast reliability aggregated over the northern hemisphere and all 12 initial conditions. Forecasts are binned by 10% increments. Histograms show sharpness diagrams, indicating how often each forecast bin was used.

Temperature extremes are herein defined as greater than 1 standard deviation or less than -1 standard deviation away from the mean of the 1982-2010 historical record at each gridpoint. This is very close to the 85th and 15th percentile. A Gaussian distribution is assumed, but may not be the most accurate fit; this is a point that requires further examination. This study assesses forecast verification, that is, the question of "did the forecast come true?" at a one-month lead for the 3-month mean, over all initial conditions. Forecast probabilities are provided in a 3-category system, where a probability of occurrence is provided for each of the three unequal categories: < -1 standard deviation (SD); > -1 SD and < 1 SD; > 1 SD, very close to 15%/70%/15% climatological likelihood. Observed probabilities are either 1 or 0. Results are compared to

Correspondence to: Emily J. Becker, Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD and Innovim, LLC, Greenbelt, MD; E-mail: Emily.Becker@noaa.gov

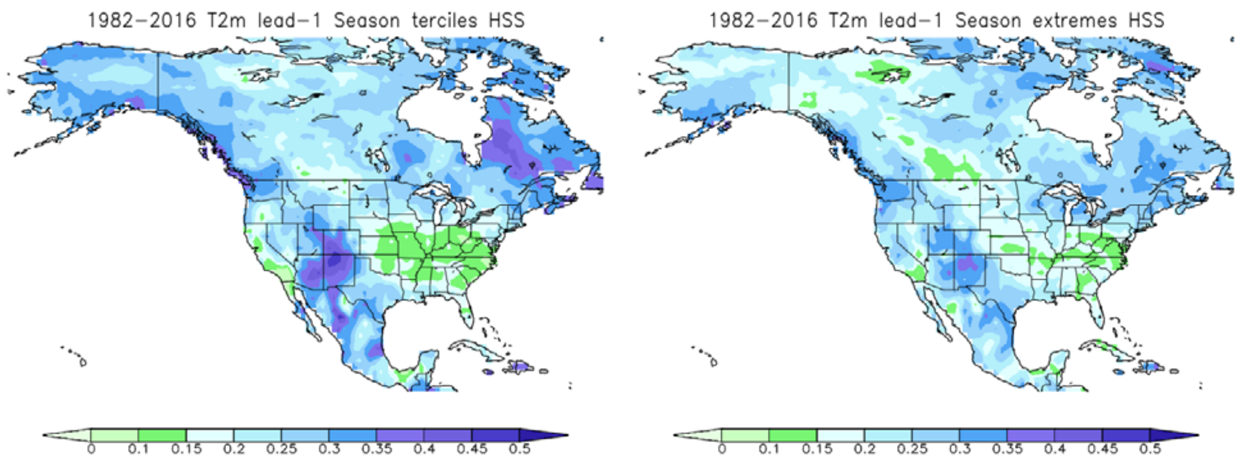

Fig. 2 Heidke skill score of lead-1 3-month-mean 2 m temperature forecasts, aggregated over all 12 initial conditions. Left panel shows results for the tercile system, right panel for the extremes.

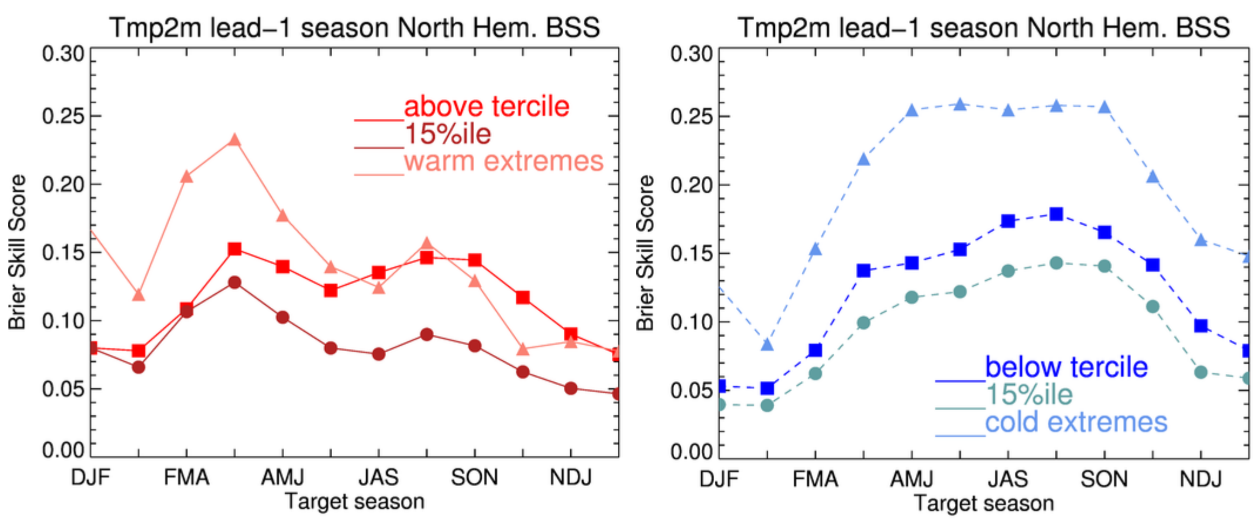

Fig. 3 Brier skill score of lead-1 3-month-mean 2 m temperature forecast aggregated over the northern hemisphere land. Left panel shows results for the above average tercile, all forecasts in the 15th percentile, and "warm extremes," where forecasts indicated > 25% probability of an extreme. Right panel shows the same for the below average/cold extremes.

the 3-category, equally likely tercile forecast system that is in use by the Climate Prediction Center: each category has a climatological likelihood of 33.33%. Metrics used in this study include the Brier skill score (BSS), Heidke skill score (HSS), and reliability diagram.

3. Results

Cold temperature extremes in the Northern Hemisphere are found to be slightly underforecast in the 1982-2016 data set, while warm extremes are slightly overforecast (Fig. 1); overall, however, the reliability is very good, indicating that forecast probabilities match up with observed frequencies. These are similar patterns to those found when the reliability of the tercile system is assessed. The Heidke skill score for extremes, examined over North America, shows similar annual-aggregate patterns to the terciles, with an area of low skill over the southeastern United States (Fig. 2). In general, HSS of extremes is slightly lower than for terciles; this is not unexpected, given the lower base rate of the extremes (15% climatological threshold). However, in some areas, such as the lower-skill southeastern US, the HSS is slightly higher for the extreme system. When comparing the BSS for the 15/70/15 category system to the BSS of the tercile system, the extremes system results in somewhat lower BSS (Fig. 3). However, when forecasts for extremes are isolated, in this case by examining only the subset of forecasts where the probability in the extreme category was > 25%, the BSS is found to be generally higher than for forecasts in the tercile system (Fig. 3).

Results for land-only precipitation rate over the northern hemisphere show very low skill for both the extreme and tercile system. Some regions and times show statistically significant positive skill, however. As a forecast for extremes could be issued only when there is reason for confidence, it's possible that a forecast system could be developed for some regions or seasons.

4. Concluding remarks

This is a preliminary study that demonstrates that there is some potential for skillful probabilistic forecasting of extremes. Further experimentation will examine the definition of "extreme", including possible use of absolute temperature thresholds and the relationship between extreme temperature and precipitation. A large ensemble such as the NMME is valuable in constructing probabilistic forecasts, and further analysis will be necessary to discover valid thresholds for triggering an extreme forecast.

References

Becker, E. J., H. van den Dool, and M. Pena, 2013: Short-term climate extremes: Prediction skill and predictability. *J. Climate*, **26**, 512-531.

Becker, E. J., 2017: Prediction of short-term climate extremes with a multimodel ensemble. *Climate Extremes: Patterns and Mechanisms*, S.-Y. Wang *et al.* Eds., John Wiley & Sons, Inc. 347-359, doi: 10.1002/9781119068020.ch21

Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948-present, *J. Geophys. Res.*, **113**, D01103, doi:10.1029/2007JD008470.

Kirtman, B. P., D. Min, J. M. Infanti, J. L. Kinter, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, Y.-K. Lim, Z. E. Li, J. Tribbia, K. Pegion, W. Merryfield, B. Denis and E. Wood, 2014: Phase-1 seasonal to interannual prediction, phase-2 toward developing intra-seasonal prediction. *Bull. Amer.Meteor. Soc.*, **95**, 585–601, doi:10.1175/BAMS-D-12-00050.1.

Xie P., and P. A. Arkin, 1997: Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539-2558.