Probabilistic prediction of extreme temperatures using NMME

Nir Y Krakauer

Associate Professor, Department of Civil Engineering and NOAA-CREST, The City College of New York
https://nirkrakauer.net ; mail@nirkrakauer.net

Summary

The global frequency of new records for hottest month recorded in a particular location has increased in recent years due to anthropogenic global warming, compared to the expected decrease in the frequency of new records under a stationary climate. I compare 3 probabilistic forecasts of whether a new record will be set in a given month: (a) baseline: assumes stationarity; (b) trend: extrapolates from the observed warming in past years; (c) NMME: adjusts the temperature based on the predictions of seasonal forecast models. Normalized mean forecast log likelihood of the outcome is used as the skill metric. Looking at 2012-2016 for all land and for North America, the trend forecast clearly outperforms the baseline. The NMME forecast outperforms the trend-based forecast, particularly using predictions issued at the beginning of the month (lag 0), with the NMME forecast maintaining a lead over the purely trend-based one up to 6 months in advance.

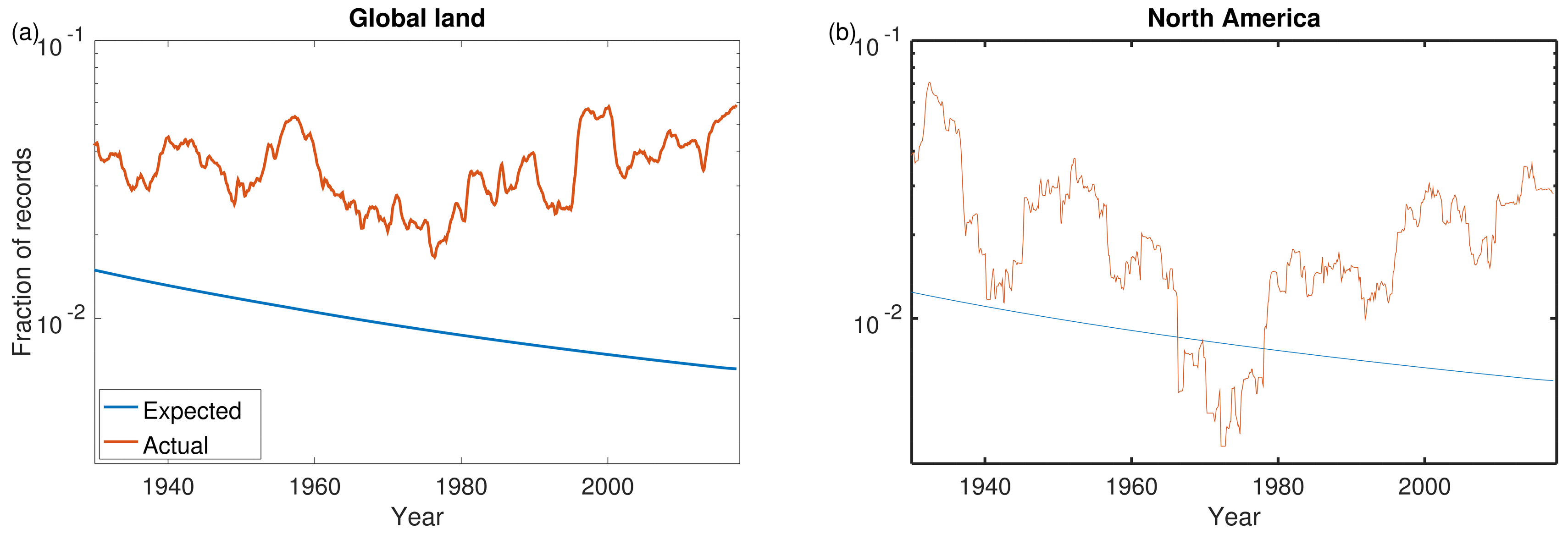

Figure 1. Actual frequency of temperature record highs versus that expected under a stationary climate, for (a) all land and (b) North America. 5-year moving average.

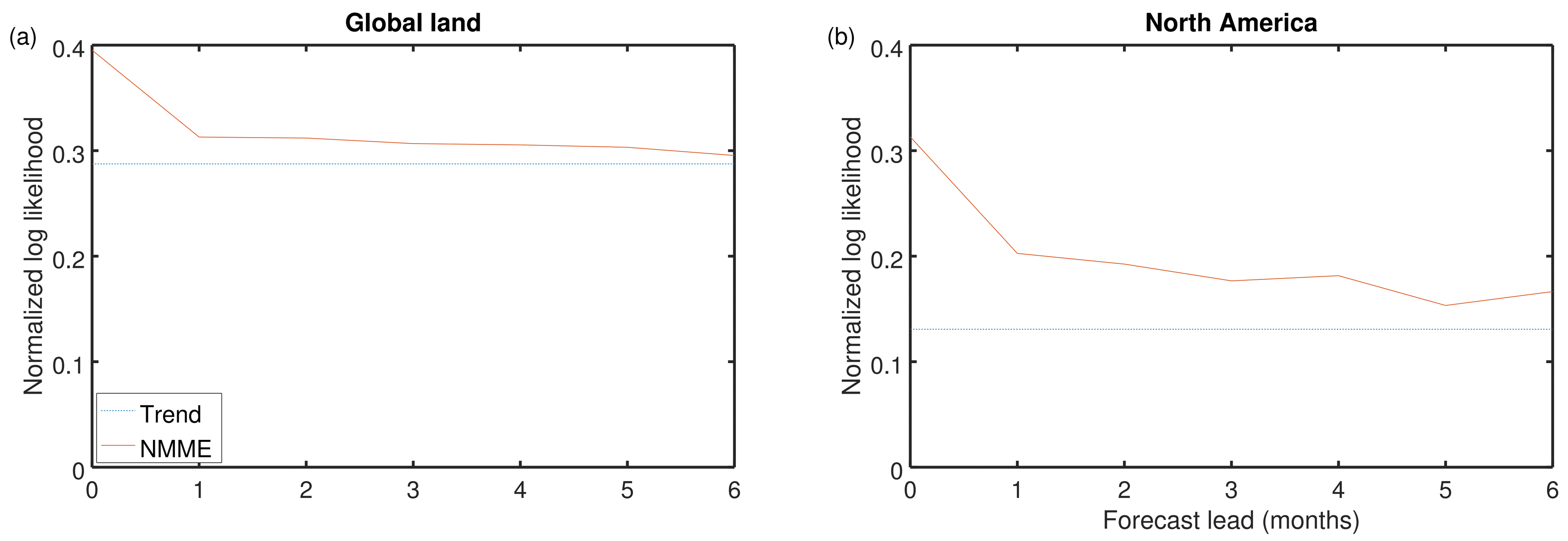

Figure 2. Normalized mean skill of probabilistic prediction of temperature record highs, for (a) all land and (b) North America over 2012-2016. Trend and NMME -based forecasts are compared. A baseline forecast based on assuming a stationary climate is assigned 0 skill, while a perfect forecast would have a skill of 1.

Further reading

NY Krakauer (2017), Temperature trends and prediction skill in NMME seasonal forecasts, *Climate Dynamics*.

NY Krakauer (2016), SeFo: A package for generating probabilistic forecasts from NMME predictive ensembles, *Journal of Open Research Software*, 4(1): e19.

H Aizenman, MD Grossberg, NY Krakauer, I Gladkova (2016) Ensemble forecasts: probabilistic seasonal forecasts based on a model ensemble, *Climate*, 4(2): 19.

NY Krakauer, N Devineni (2015), Up-to-date probabilistic temperature climatologies, *Environmental Research Letters*, 10(2): 024014.

Acknowledgments

This work has been supported through NOAA grants NA11SEC4810004 and NA15OAR4310080.

Results

Whereas under a stationary climate the frequency of new records would be expected to gradually diminish with time, the frequency of new record high monthly temperature has not diminished and averages several times larger than expected (Figure 1), e.g. 5.0% for all land and 3.4% for North America over 2012-2016, compared to 0.65% under a stationary climate.

Forecast skill, as measured by normalized log likelihood, was some 29% for the trend forecast globally (13% for North America). NMME at lag 0 had substantially greater skill (40% globally, 31% for North America). At lags of 1 month and more, NMME skill dropped off, but remained slightly better than the trend forecast even at lag 6 (Figure 2).

The results are encouraging that even the likelihood of new temperature records (i.e. the right tail of the probability distribution) can be adequately approximated by assuming that anomalies follow a normal distribution. Probabilistic forecasts derived from the NMME means could supplement the existing operational ones, which focus on per-tercile probabilities and generally do not provide direct information on the likelihood of extreme values.

Methods

Data

Temperature records (monthly since 1850) for training and evaluation were taken from the gridded Berkeley Earth Land + Ocean dataset. To simplify averaging, I used the equal-area grid, which has a nominal resolution of 1°at the Equator. Forecast skill was averaged either over all land (grid cells at least 50% land) or over North America (20-50 °N, 60-130 °W).

Forecasts

Each forecast is in the form of a probability p for each month and grid cell exceeding the highest temperature previously recorded there.

The baseline forecast is the same globally, and consists of setting $p = 1/(N+1)$, where N is the mean number of years of available temperature data. It is expected to be close to optimal if there is no trend or predictability to the occurrence of temperature records.

The trend and NMME forecasts both assume that temperature anomalies are normally distributed, generating a t distribution for temperature whose parameters are estimated by linear regression. The exceedance probability p is derived from this as 1 minus the quantile of the t distribution corresponding to the previous record value.

The mean and standard deviation for trend forecasts are based on an exponentially weighted moving average (with 15-year timescale, using data since 1957). This includes a globally constant additive adjustment to minimize bias of the moving average due to recent warming.

North American Multi-Model Ensemble (NMME) predictions were available since 1982 (including hindcasts) and were based on the means of ensemble values for 8 models: CMC1-CanCM3, CMC2-CanCM4, COLA-RSMAS-CCSM4, GFDL-CM2p1-aer04, GFDL-CM2p5-FLOR-A06, GFDL-CM2p5-FLOR-B01, NASA-GMAO-062012, NCEP-CFSv2. The forecast mean was a linear combination of these values, with the (spatially constant) weights determined by linear regression over previous prediction-verification pairs. The regression model also included a linear time trend to correct for any misspecification of warming in the NMME models. The forecast standard deviation was determined by linear regression over past values for each grid point separately. The NMME predictions used were either for the current month (lag 0) or up to 6 months in advance (lags 1-6).

Skill metric

Probabilistic forecasts perform better the higher their mean log likelihood, which for a binary outcome has the form

$$\mathrm{LL} = \langle I_i \log p_i + \neg I_i \log(1 - p_i) \rangle, \tag{1}$$

where I is an indicator function equal to 1 when a record does occur and 0 when it does not. The average was taken over land (or North America) BEST grid points for forecasts of each month in 2012-2016 (60 months).

A perfect forecast would have LL of 0, since the forecast probability of the actual outcome would always be 1. If we also calculate the LL for the baseline forecast, we can normalize the LL for other forecasts so that 1 is perfect and 0 is equal to baseline:

$$\mathrm{LL}_{\mathrm{norm}} = \frac{\mathrm{LL} - \mathrm{LL}_{\mathrm{baseline}}}{-\mathrm{LL}_{\mathrm{baseline}}}. \tag{2}$$

Here, $\mathrm{LL}_{\mathrm{norm}}$ is found for the trend forecast and for the NMME forecast (at lags 0-6), as averaged over either all land or North America.