Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
43rd NOAA Annual Climate Diagnostics and Prediction Workshop
Santa Barbara, CA, 23-25 October 2018

The 2018 US Heat Season and the Performance of the SEHOS Forecasting Tool at the CPC

Evan M. Oswald
Climate Prediction Center, NOAA/NWS/NCEP, and Innovim, LLC

1. Tools at the CPC

What is called the "Subseasonal Excessive Heat Outlook System", or SEHOS, was constructed in 2016 at the Climate Prediction Center (CPC) in conjunction with University of Maryland Cooperative Institute for Research in Environmental Sciences (UM CIRES). It provides a real-time probabilistic forecast of the occurrence of heat waves (or "excessive heat events") with lead times of 8 to 14 days. It is a direct and simple interpretation of any given forecast models' output in terms of the chances of a heat wave. A heat wave here consists of consecutive dates over either the 90th or 95th historical percentile values for the time of year and at this location. In addition to producing the probability of a heat wave occurring, also estimated are the heat wave start date and duration length, the forecast period's single-day and/or running average (2 or 3 days) maximum temperature and standardized anomalies, the historical 90th and 95th percentile values for the period, and the probabilities of exceedance of three absolute thresholds. All products come in both heat index and dry air temperature varieties, but they also come in a "hybrid" variety. These hybrid probabilities are the maximum probability/value between the heat index and temperature at any given location or in any of the various daily products.

The SEHOS uses forecasts of daily maximum temperature and heat index from two operational ensemble forecast systems/models: the ECMWF Ensemble Prediction System (ECENS) and the Global Ensemble Forecast System (GEFS). In addition, a third model-output is also utilized: an equal weighted blend of the two sets of SEHOS outputs (denoted "ECENS-GEFS"). The historical reforecasts of these models during the summertime from 1996-2014 are used, in conjunction with the NCEP R1 reanalysis, to calibrate the raw probabilistic forecasts. This is done using reliability mapping with the locally (both spatially and calendar-wise) sampled paired reforecasts and observations.

Table 1 The forecasts' skill historically, and during summer 2018.

WK 2[a]	*National AUC-ROC*	*National SEDI*
GEFS	0.61, 0.64	0.24, 0.30
ECENS	0.63, 0.67	0.28, 0.34
BLEND	0.63, 0.68	0.28, 0.36
WK 34[b]		
CFS	0.58	0.21
ECMWF	0.59	0.23

a For WK2, the first value corresponds to skill forecasting the 2018 heat season events, and the second value to historical period (~1999-2014). Both values correspond to the 90th percentile based events.

b For WK34, all values correspond to probabilities of 3-dates (not 6).

All three models are calibrated separately. Skill score metrics used to quantify the accuracy of the models include the area under the receiving operating curve (AUC-ROC) and the maximum symmetric external dependence index value (maximum over the range of thresholds by which one can diagnose "heat wave" or "no heat wave", but herein denoted as the "Max. SEDI"; Ferro and Stephenson, 2011). Generally speaking the models have accuracy/skill about a third of the way from "random guess" to "perfect predictor". The ECENS-

Correspondence to: Evan M. Oswald, Climate Prediction Center, NOAA/NWS/NCEP, and Innovim, LLC;
E-mail: evan.oswald@noaa.gov

GEFS model performs the best and the ECENS model the second best. Interestingly, the heat waves based on heat index-based heat waves saw more accuracy (on average, 28% vs. 34%) than the dry air heat waves did. Spatially, the models have the greatest skill in the Southern Plains.

An important tool called the "Probabilistic Extremes Tool" has been used by CPC forecasters since 2014. This tool produces daily maximum and minimum temperature forecasts during the week 2 period (days 8-14). These forecasts are formatted as the probability of exceeding a range of absolute (*e.g.* over 100F) and relative thresholds (*e.g.* below the 5th percentile, over the 90th percentile). Daily resolution probabilistic forecasts are helpful to CPC forecasters when they need support forecasting start and stop dates of heat waves, albeit potentially at the expense of skill (*i.e.* temporal averages are easier to forecast beyond week 1). This tool also provides the climatological values and percentiles associated with the various thresholds, at the forecasted time of year. With respect to air temperatures, while daily maximums and minimums are both considered, heat index is not forecasted within this tool. A limited form of this tool is also available to the public via the CPC webpage. In addition to air temperatures, this tool is also available for precipitation and 10m wind speed. The only dynamical model that currently drives the temperature forecasts from this tool is the GEFS, but forecasts from the ECENS and Canadian Ensemble Forecasts (CAN) are said to be forthcoming. Calibration of the probabilistic forecasts was accomplished via ensemble regression (Unger *et al.* 2009) and an in-house observational 2-meter above ground level (AGL) temperature dataset over the 1986-2010 period. Skill scores of the real-time forecasts were calculated using the ranked probability skill metric from 2014-2017. Skill scores vary depending on the lead time and percentile values, but generally range from about 4-16% of the way to "perfect" from "random guess".

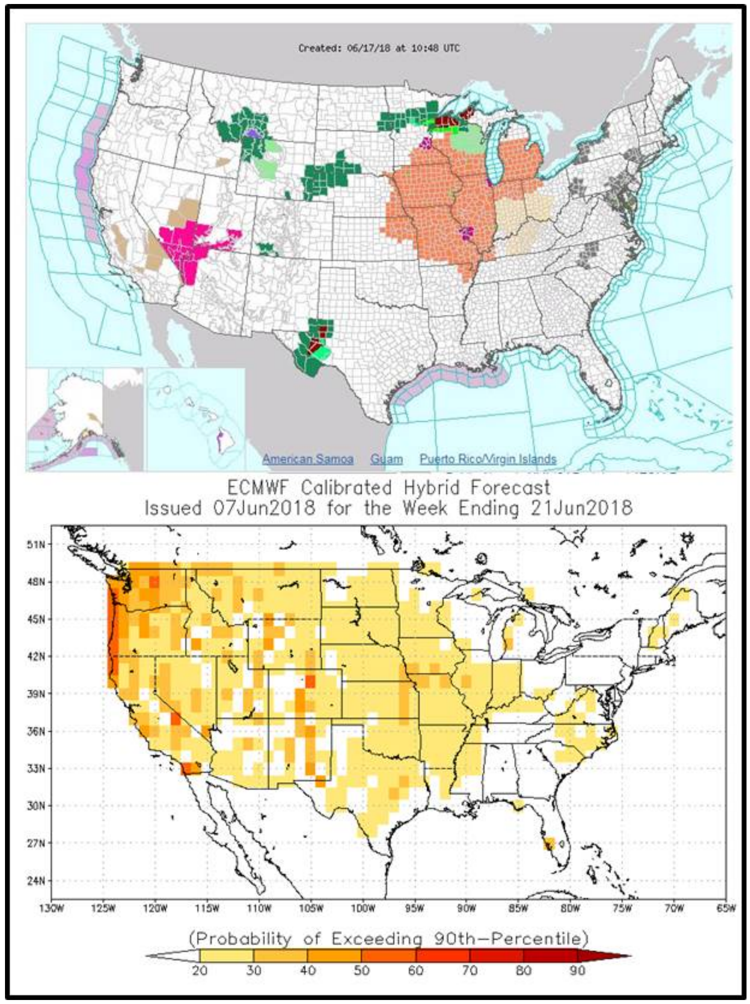

Fig. 1 (Top) National Weather Service Watches, Warnings and Advisories map the morning of June 17, 2018. Orange represents heat advisories for the afternoon, and the embedded magenta colors are excessive heat warnings. (Bottom) Probabilistic forecasts of a heat wave during the week ending on June 21st, as issued on June 7th by the ECENS powered version of the week 2 tool.

Forecasting at for the week 3-4 timescale, while still currently experimental in nature at the CPC, is currently supported by a different suite of dynamical based tools. It doesn't have a snazzy name – perhaps indicative that it is still very new (developed in 2018). This tool leverages the ECENS (Mondays and Thursdays) and the CFS model (daily) model forecast outputs in order to forecast the presence of multiple extremely high daily mean temperatures in week 3-4. Forecast guidance for this tool is less weather-based than the two previously mentioned tools: focusing on daily mean temperatures rather than daily maximum, and only requiring a number of (3 or 6) hot (over the 92.5th percentile) dates and thus dropping the *consecutiveness* aspect of the requirement. There is also provided (uncalibrated) guidance for the timing within the period for the dominant signal (*i.e.* spatial location). The CFS model is taken as a 3-day super-ensemble consisting of a meager

12 ensemble members, while the ECENS is not a super-ensemble but has over 4x (51) the number of ensemble members. Probability is taken as the percent of the ensemble members that have or exceed the number of required hot dates during the 2-week forecast period. The raw probability is then calibrated using the historical relationship (*i.e.* reliability mapping) between forecast probabilities and observed frequency over all locations and dates (*i.e.* its' not dynamic in space and calendar date).

The week 3-4 tool suite skill was measured using the AUC-ROC and SEDI metrics. Nationally, the ECENS had a 0.59 AUC-ROC and the CFS had a 0.58 when forecasting 3-or-more hot dates during the 2-week period; so about 16-17% of the way to "perfect" from "random guess". The same metric, but for 6-or-more days, was either the same or a very similar value. In regards to the nation-wide SEDI metric the ECMWF had a 0.23 and the CFS had a 0.21 when forecasting 3 or more hot dates. Therefore, the skill of this tool is roughly 22% of the way to "perfect" from "random guess".

2. Overview of the 2018 season

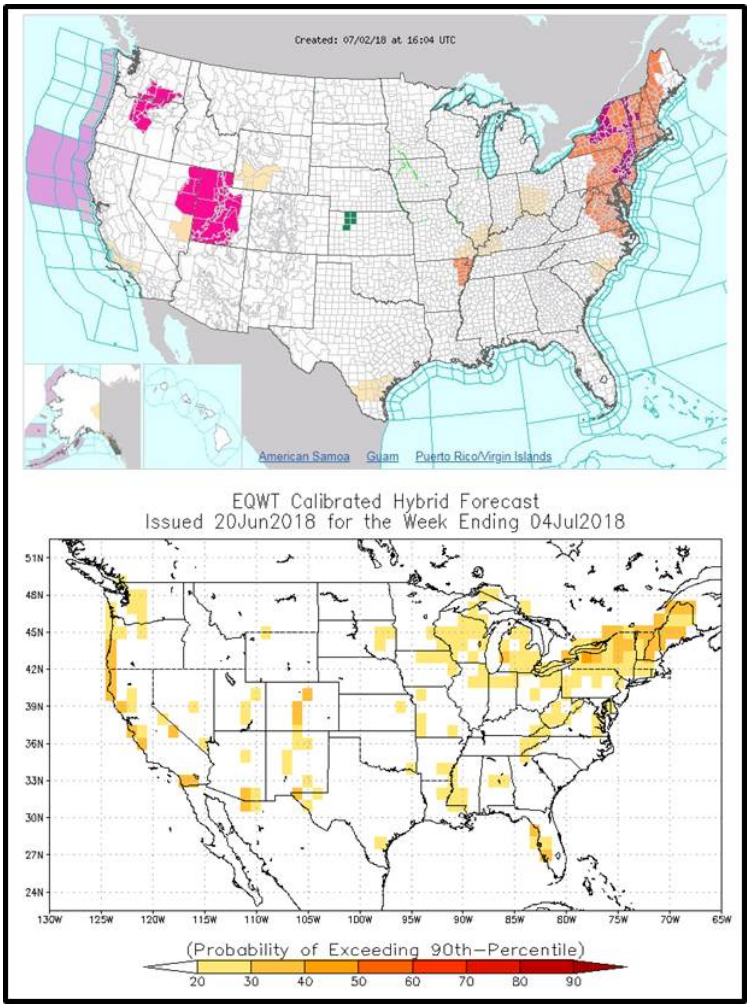

Fig. 2 (Top) Same as Figure 1 but for the morning of July 2nd, 2018. (Bottom) Probabilistic forecasts of a heat wave during the week ending on July 4th, as issued on June 20th by the GEFS-ECENS equal-weighted blend forecast powered version of the week 2 tool.

The summer of 2018 was very hot compared to historical norms, especially in the Southwestern, Southern Plains and Northeastern regions. Nationally, it compared well with other recent warm summers such as 2011, 2012, and 2016. Examination of nation-wide "google search frequency" for terms such as "heat wave" and "hot weather" illustrated 5-6 prominent spikes throughout the summer with one such extremely large episode in early July. While it is subjective, our analysis delineated about 15 warm-season heat waves in the US in 2018. At the end of this article we will discuss in detail a handful of the more prominent episodes.

The skill of the forecasts in the 2018 heat season was comparable to forecasts from the model over the recent historical period. Skill metrics (Table 1) were calculated for all the models against historical observations (NCEP/NCAR R1 dataset) over the 1999-2014 period, at a minimum, and with sample sizes ranging from 400-4,400. The week 2 models (GEFS, ECENS, GEFS-ECENS) demonstrate more skill than the Week 34 models (CFS, ECENS). The former are just under a third of the way towards "perfect predictor" from "random guess", while the latter are just under a fifth of the way towards "perfect predictor". However, for the 2018 period, we saw a modest decrease in that skill (in regards to week 2 forecasting) to just over a quarter of the way towards perfect prediction.

3. **Case studies of 2018**

An extremely hot period in the central of the US on Father's day marked the first extensive event of the heat season, which (depending on location) spanned June 14/16-18/20. It covered several states (*e.g.* from the

Central Plains through to the Great Lakes; Fig. 1, top) and several population centers, lasted 4-5 dates, and bore extremely hot temperatures in both daytime and nighttime temperatures. For instance, in Lincoln, Nebraska, there were 5 consecutive dates breaking the daily record. Additionally, high dewpoint temperatures drove the heat index values into the low 100's. For this event, drought was not present in the region. Air quality ranged from good to unsafe-for-sensitive-groups, with worse air quality in the eastern part of the region.

Analysis of the news headlines throughout the summer suggested the social disruption of this event was moderate compared to the other events in 2018. Since many schools were just wrapping up the school year, school closings comprised the bulk of the headlines. This region has a large number of residents and therefore any heat wave can make headlines, somewhat heedlessly. However relatively speaking, this region in mid-to-late June is not especially vulnerable to episodes of mid-90s temperature and high (upper 60s) humidity (modest hospital admission increases, no infrastructure damage, etc.).

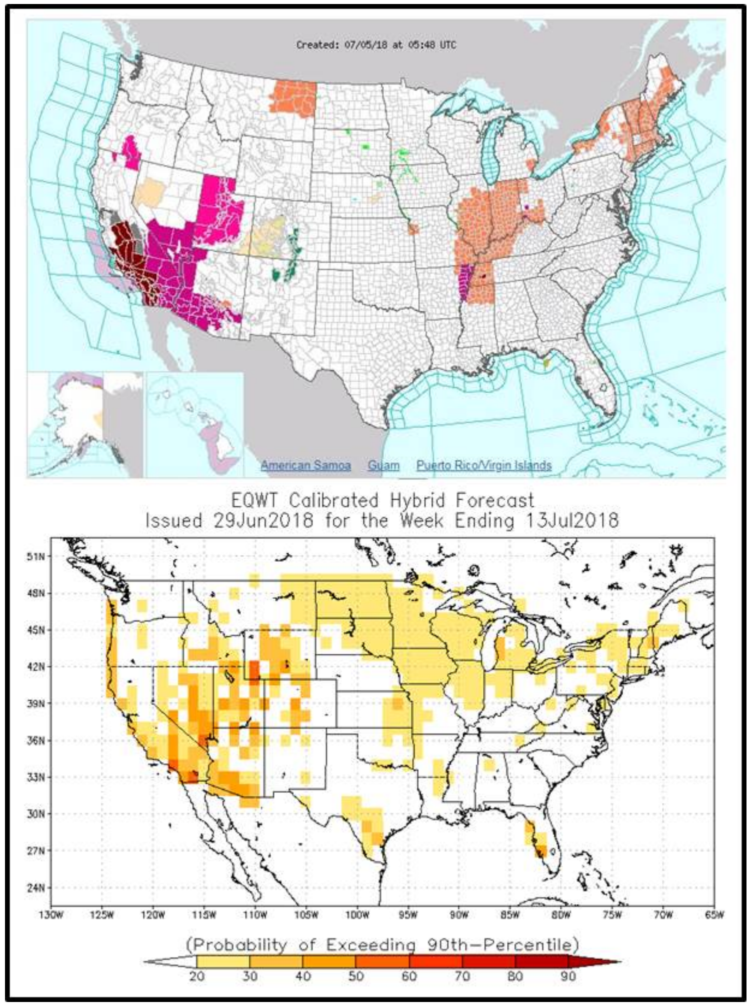

Fig. 3 (Top) Same as Fig. 1 but for the morning of July 5th, 2018. (Bottom) Probabilistic forecasts of a heat wave during the week ending on July 13th, as issued on June 29th by the GEFS-ECENS equal-weighted blend forecast powered version of the week 2 tool.

For consistency from event-to-event and across models, all forecasts are evaluated from the perspective of a 9-day lead to the beginning of the event. Additionally, we will focus on the Week 2 tools (SEHOS). Overall the forecast accuracy was mediocre when the 16th was 9 days the lead of the forecasts, and the skill varied with model. On one hand the GEFS-based forecasts completely failed to indicate a threat in the region with probabilities almost all sub-20%. The spatial pattern suggested an event in the Southern Plains. Conversely the ECMWF had probabilities of mostly 20-30% (30-40% maximum) and a spatial pattern with some resemblance of the observed event (Fig. 1, bottom). The resulting blended forecast had an even-handed mix of both sub-20% and 20-30% probabilities in the impacted region, but the spatial pattern still suggested a greater chance of a heat wave in the Southern Plains rather than the impacted region.

Some heat waves have a migratory nature to them, traveling slowly from one part of the country to another. The subject of the second case study was one of these, and it was actually broken into 4 "chapters" in our internal analysis: June 22-25 in the Texas-Southern Plains, June 24/26-June 27/July 2 in the Deep South, June 27-June 30/July 1 in the Middle Mississippi-Great Lakes, and finally June 29-July 5 in the Northeast-Mid-Atlantic. The beginning of the event brought 100°F degree weather to parts of Texas for the first time in the summer. Air quality was not an issue, and the duration was not overly long. However, the temperatures were very high and the region was predominantly in moderate drought. After Texas, the event baked the Lower Mississippi Valley in the Deep South for 3.5-7.5 dates – depending on location. High dewpoints contributed to heat index values over 105°F. Air quality was not an issue during this part of the event. Then this large, migrating heat wave moved into the center of the US for a few days. Dewpoints were high (low 70s) and pushed heat index values over 105°F in many locations. Overnight lows were high-to-extreme, depending on location, including into the low 80’s on some evenings in some locations (*e.g.* Kansas City). Strong small-scale variability in the strength of the anomalies seemed to reflect the soil moisture patterns; this region ranged from extreme drought to normal soil moisture. After baking the central US, the heat wave settled into the Great Lakes region and then the densely-populated Northeastern US (Fig. 2, top). The degree of temperature extremes (e.g. anomalies) in the Northeastern US were the greatest this migrating heat wave recorded. The Great Lakes region had typical soil moisture content, but much of the Northeast was dry or in (low-level) drought. Mild air quality issues accompanied the event, except for dramatic air pollution in the NYC-Philly-NJ area for a single afternoon.

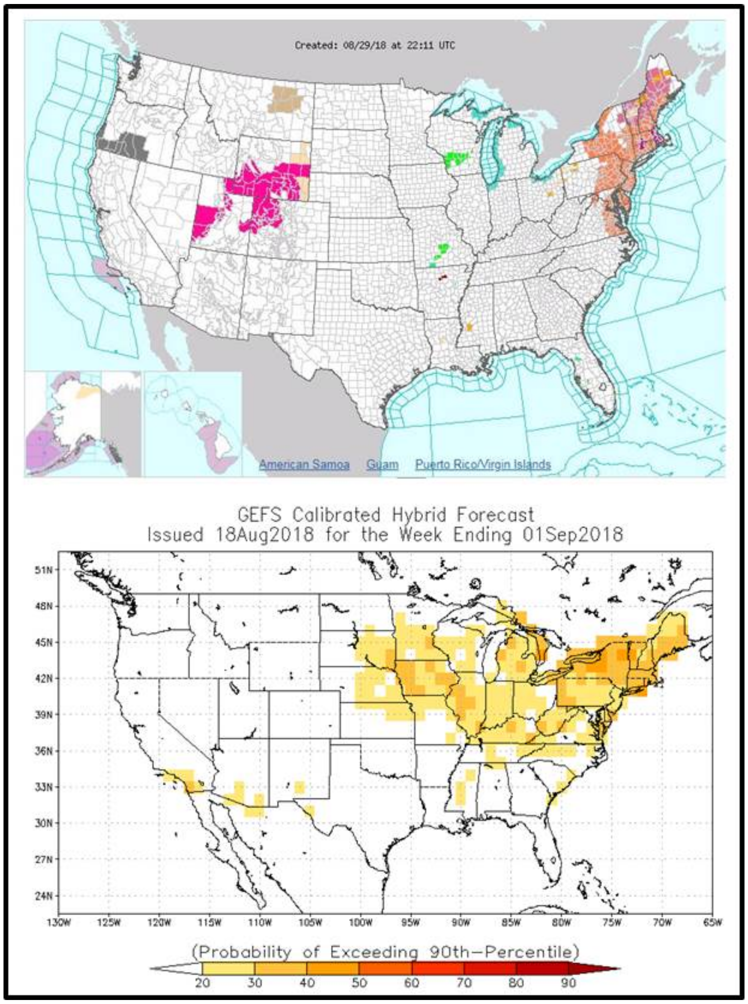

Fig. 4 (Top) Same as Figure 1 but for the morning of August 29th, 2018. (Bottom) Probabilistic forecasts of a heat wave during the week ending on September 1st, as issued on August 18th by the GEFS forecast powered version of the week 2 tool.

Albeit with its fair share of stories regarding children being trapped in hot cars, heat exhaustion of athletes, stressed agriculture, opening of cooling centers; the level of societal impacts during the first half of the episode was fairly typical for mid-summer heat waves. However the latter half had notably large societal impacts, the most conspicuous being 70-plus deaths in Quebec as well as in the Great Lakes an alarming number (over 12) of road (highway) closures due to cement buckling. Mild air quality issues accompanied the event, except for dramatic air pollution in the NYC-Philly-NJ area on one afternoon. A rash of daily historical records were also broken during the last leg of the heat wave from Vermont to New York City.

The beginning of the episode, in Texas-Southern Plains, was not well forecasted by either dynamical model. Albeit the quality of the forecasts improved as the episode progressed and therefore the forecasts were generally OK for the Middle Mississippi-Great Lakes of the US. For example, the blended forecast has 20-30% coverage in most of the impact region, with some 30-40% and some sub 20%. By the end of the event, when the heat was in the Northeast, forecasts from the ECMWF for most of the impacted region showed 20-30% probabilities and the spatial pattern resembles the observed region impacted (Fig. 2, bottom).

A four-day episode of record breaking heat hit Southern California and parts of the Desert Southwest the weekend following the 4th of July (July 5-8; Fig. 3, top). For example, San Diego and Las Angeles both saw multiple/consecutive single-day records fall. Several single-day records fell in the desert, too. At the desert southwest locations absolute thresholds of 110-115 were met, in lieu of relative thresholds. Air quality was an issue, usually only moderately but very elevated risk existed in the Los Angeles area. Drought was an issue, ranging from moderate (Southern California (SoCal)) to extreme (Desert Southwest). Societal impacts were modestly larger than usual and included a postal worker dying in their mail-truck, and over 20,000 customers going without power.

The forecasts were relatively good for this event. The GEFS- and ECMWF-based forecasts predominantly had probabilities of 20-30% of a heat event in SoCal, and chances of 110°F exceedance over 50% in the desert southwest locations. The blended forecast (Fig. 3, bottom) was slightly superior to either individual model-based forecast, and for both SoCal locations (relative thresholds) and the Desert Southwest locations (absolute thresholds).

The weekend before Labor Day displayed a substantial 3.5-day heat wave that impacted an estimated 68 million Americans from Chicago to New York City. The western portion of the episode hit the Midwest and Great Lakes for 4 dates (August 25-28). It was accompanied by very high dewpoints (>=74°F average in Indianapolis) that drove extremely high nighttime temperatures and heat index values. Daily high heat index values ran in the upper 90’s to mid-100’s, which is very hot for this time of year and part of the country. It had moderate air quality levels, any issues due exclusively to PM2.5. The only drought in the region impacted a substantial portion of Michigan, with moderate-to-severe drought. The eastern portion of the heat wave hit New England and the Mid-Atlantic states for a couple dates (August 27/28-29/30; Fig. 4, top). Again very high dewpoints across the region drove heat index values over 100-105°F and several single-day-record nightly low temperatures to be broken. The eastern chapter of this event also saw locations (VT, PA, RI, NY) with extreme daily high air temperatures. Drought did impact some of the northern areas (upstate NY, VT, NH, ME).

While certainly some daily records that fell, the numbers of societal impacts were minimal. Perhaps many residents still had the knowledge and resources to beat the heat. Some headlines included school closings and the US Open needing to be shut down for a day.

The western portion of the episode was not as well forecasted as the eastern portion, as this was primarily due to the ECENS’s ability to forecast the episode. The GEFS had good skill with the western portion of the impacted region showing half 20-30% and half 30-40%, and in the eastern portion they were predominantly 30-40% (Fig. 4, bottom). The ECMWF did much better for the eastern portion of the event.

References

Ferro, C. A. T., and D. B. Stephenson, 2011: External dependence indices: improved verification measures for deterministic forecasts of rare binary events. *Wea. Forecasting*, **26,** 699-713. doi: 10.1175/WAF-D-10-05030.1.

Unger, D. A., H. van den Dool, E. O’Lenic, and D. Collins, 2009: Ensemble regression. *Mon. Wea. Rev.*. **137**, 2365-2379. doi: 10.1175/2008MWR2605.1.