

improves the predictability of short-term droughts. The PDSI incorporates temperature, precipitation, soil moisture storage, and net radiation that are relevant in the dynamics of the dry climate regime in the NE. The PDSI has been used since the 1960s to assess the severity of droughts (e.g., Palmer 1965; Alley 1984; Briffa et al. 1994; Wells et al. 2004; Dai et al. 2004; van der Schrier et al. 2011), and it was found to be an objective metric to assess the variability of short-term droughts (Lohani et al. 1998; Sims et al. 2002). In this paper, a short-term drought is defined in the intraseasonal or subseasonal temporal scale (e.g., from weeks up to three months; Lorenz et al. 2017). The PDSI metric was found to be comparable with other metrics such as the standardized precipitation index (SPI; McKee et al. 1993) and standardized precipitation and evapotranspiration index (SPEI; Vicente-Serrano et al. 2010).

This insight on the predictability of drought is not investigated based on changes in the model core dynamics or parameterizations, but by using a special statistic postprocessing treatment over all possible signals of potential predictability. Several drought studies have incorporated multimodel ensemble approaches with successful results (Yuan and Wood 2013; Infanti and Kirtman 2014; Becker et al. 2014; Bolinger et al. 2017). However, using the traditional multimodel ensemble mean does not provide additional relevant information, nor does it explore decadal signals that could improve predictability. In the early 2000s, several hypotheticals (Wilks 2006) and “real” forecast (Wilks and Hamill 2007) time series were used to test the value added of ensemble postprocessing and found positive results. However, applications using PDSI computed from a multimodel ensemble dataset trained over decadal length are not yet a common tool (Bolinger et al. 2017; Carrillo et al. 2018; Hao et al. 2018). This study uses the currently available (hindcast and forecast) North American Multi-Model Ensemble (NMME; Kirtman et al. 2014) datasets to fulfill this gap.

We then tested the ability of the PDSI–NMME to forecast short-term drought (predictability of the second order; Becker et al. 2014). We investigated whether the decadal training of the synthetic multimodel ensemble climate has some influence on enhancing this predictability. We answer whether the ensemble model output statistic (EMOS) outperforms the traditional average over a long period. If the above is true, we could propose a new design of drought forecasting that explicitly includes the multimodel ensemble signal as a framework to assess drought severity in the NE region. We hypothesize that the EMOS methods applied at the NMME multimodel ensemble will outperform the ordinary ensemble mean training of the entire time domain. We can then utilize this assessment to adjust an operational drought forecast and add value to the biased numerical prediction.

2. Datasets and methodology

a. The Northeast Regional Climate Center dataset

Temperature and precipitation datasets were obtained from the Northeast Regional Climate Center (NRCC; DeGaetano and Belcher 2007; DeGaetano and Wilks 2009) at monthly

temporal resolution from 1950 to 2016. Its original resolution is 4 km, but it was linearly regridded to 1° and 32 km. We used the NRCC dataset because of its real-time operational availability for temperature and precipitation, and the low degree of uncertainty found in the Northeast region (Bishop and Beier 2013). The NRCC temperature gridded product is based on the rapid update cycle (RUC) model output that is used to interpolate to the higher-resolution Cooperative Observer Network stations (DeGaetano and Belcher 2007). Two interpolations are performed: elevation adjustment using the RUC lapse rate, and the horizontal interpolation using a multiquadratic approach. NRCC precipitation uses a radar-based correction to adjust rain gauge data to a refined spatial resolution (DeGaetano and Wilks 2009). The method was applied to daily precipitation using an inverse-with-distance interpolation. Our analysis compares the coarse resolution to show consistency among resolution in the dataset.

b. The North American Multi-Model Experiment dataset

Temperature and precipitation forecast products were obtained from the NMME project (Kirtman et al. 2014). Five models were selected (see Table 1), and from each model 10 ensemble members were used. As some models have more than 10 ensemble members, we enforced an equal number of ensemble members in our calculation. The initial forecast on NMME starts on the first day of each month. Therefore, for a given month, we have 50 possible realizations. It was already shown that multimodel forecasting outperforms the single-model approach (Kirtman et al. 2014; Saha et al. 2014), and the details of how this superiority happens were extensively analyzed focused on error compensation and reliability of the single model (Palmer et al. 2004; Hagedorn et al. 2005; Weigel et al. 2008). The NMME focused on seasonal to interannual time scales, but we are using the three first lead time forecasting months. The original model grids are linearly interpolated to a 1° × 1° resolution. The period of the analysis is from 1982 to 2012 (31 years). Temperature and precipitation from the NMME have shown some forecast skill when analyzed with anomaly correlation (Becker et al. 2014).

c. The Palmer drought severity index

We used the PDSI to define the dryness of the NE region. PDSI has been used since the 1960s to evaluate meteorological droughts across the United States and worldwide (Palmer 1965). The observational precipitation and temperature data to compute the PDSI are from the NRCC (observed) and the NMME (forecasted) datasets. The calibration period for this PDSI is from 1950 to 1980. The PDSI is computed using the Thornthwaite (TH) and Penman–Monteith (PM) variations for the potential evapotranspiration (van der Schrier et al. 2011). To compute potential evapotranspiration needed for the water budget balance in PDSI, we use net radiation fields from the NCEP–NCAR Reanalysis-1 (Kalnay et al. 1996). However, the forecast skill analysis is only performed on TH for consistency in radiation parameters for the historical and forecast data. The available water capacity (AWC), a term used in

f. The skill score

The verification of the forecasting skill score is done for individual months of the summer season (July–September). We have evaluated 3-month forecasts with leads of 0, 1, and 2 months. The NMME initialization is from 1 July for all the cases. Three metrics are used to evaluate the forecast skill: correlation, continuous ranked probability score (CRPS; Epstein 1969), and the reduction of variance skill score (SScore; Wilks 2011). The SScore is used to quantify the improvement skill based on the climatology, and the CRPS provides a robust quantification of the probabilistic forecast when using the NGR approach. In the reduction of variance skill score (SScore),

$$\text{SScore} = \frac{\text{MSE} - \text{MSE}_{\text{clim}}}{\text{MSE}_{\text{clim}}} \times 100\%, \quad (5)$$

the metric to evaluate the forecast improvement is the mean square error [MSE; $\text{MSE} = (1/n) \sum_{k=1}^n (y_k - o_k)^2$] between observation (o_k) and forecast (y_k). The reference metric is the MSE of the climatology, $\text{MSE}_{\text{clim}} = (1/n) \sum_{k=1}^n (o_k - \bar{o})^2$, where $\bar{o}$ is the observed climatology.

The CRPS skill score (SS_{crps}) is defined as

$$\text{SS}_{\text{crps}} = \frac{\text{CRPS} - \text{CRPS}_{\text{ref}}}{\text{CRPS}_{\text{ref}}} \times 100\%, \quad (6)$$

which is based on the CRPS:

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dy, \quad (7)$$

where $F(y)$ is the continuous CDF of the predictand y . The term F_o is the cumulative probability step function defined by

$$F_o(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value} \end{cases}. \quad (8)$$

CRPS for a given observation o is calculated using

$$\begin{aligned} \text{CRPS}(\mu, \sigma^2, o) = \sigma \left\{ \left(\frac{o - \mu}{\sigma} \right) \left[2\Phi\left(\frac{o - \mu}{\sigma}\right) - 1 \right] \right. \\ \left. + 2\phi\left(\frac{o - \mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \right\}, \end{aligned} \quad (9)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF, respectively, of the standard Gaussian distribution.

We used the leave-one-out cross validation to fit the post-processed NMME PDSI reconstructed data before being used to compute the SScore (Wilks 2011). The test of significance used t and F distributions for the local significance, and the global significance used a nonparametric Monte Carlo distribution with 500 random permutations (Livezey and Chen 1983).

3. Results

a. Variability of drought in the U.S. Northeast

A decadal signal exists in drought variability of the United States that particularly affects the NE region. Figure 1a shows

the spatial extension of the 1962–66 drought that affected the NE (Barlow et al. 2001). A clear out-of-phase spatial pattern is noted between the NE (negative PDSI) and southeastern states (positive PDSI). The PDSI temporal variability over the NE (Fig. 1b) seems to have an interannual (2 year) and decadal signature, as highlighted by its spectrum plot (Fig. S2a in the supplemental material). A positive trend is observed (0.18 PDSI units/10 years), which can be inferred from the bump in the spectrum at 25–35 years. The 10-yr running mean average (Fig. S2b) clearly show that the 1962–66 drought is part of the low-frequency variability. One characteristic of this analysis is that both the decadal and the 2-yr signal are statistically significant. Next, we will use this information to improve the predictability of short-term drought in the NE.

Previous studies showed that the bias correction approach on PDSI can provide improved results (Carrillo et al. 2018). As we assess the predictability of an index for droughts that is constructed with precipitation and temperature, we show first the forecasting skill of temperature and precipitation for a 3-month lead time with the same initial month (e.g., 1 Jul). Using the correlation between the forecasted case and observation, temperature performed as expected for the three months in the NE region (Koster et al. 2011), with a typical reduced performance as the forecasted lead months progress (Fig. 2, left panel). The spatial pattern is consistent with the result presented in Becker et al. (2014), but the BC seems to add value in the eastern part of the domain with an emphasis in the second and the third month. Using the SScore, the spatial pattern of the skill performance for temperature is maintained but numbers are lower due to the challenge of outperforming climatology skill (Fig. S3). However, spatial patterns between the two metrics are consistent. On the other hand, forecasting skill for precipitation was a greater challenge (Fig. 3, left panel). Although correlation showed a similar pattern, there is a limited skill in the NE after the second month. Still, the relatively high score in the third month is in the NE. The low skill in precipitation is well known (Saha et al. 2014), but as PDSI is built with both variables (precipitation and temperature), it is necessary to quantify how PDSI helps to address drought predictability. Previous studies have shown that temperature and precipitation of the NMME models have a statistically significant forecast skill in North America (Becker et al. 2014; Infanti and Kirtman 2014). Using anomaly correlation to assess potential predictability, temperature at 2 m shows an average skill of 0.26 and precipitation a value of 0.16 (Becker et al. 2014). However, we have pointed out that using NMME without any postprocessing is of limited use (Carrillo et al. 2018).

b. The bias correction–spatial disaggregation approach

The BC-SD approach corrects mean and variance, and it provides higher spatial resolution while keeping the pattern of improvement unchanged (Figs. 2 and 3, right panel). Therefore, BC-SD does not increase the forecast skill along with the domain, as results are consistent with the analysis using BC (Figs. 2 and 3, left panel). The BC was done with

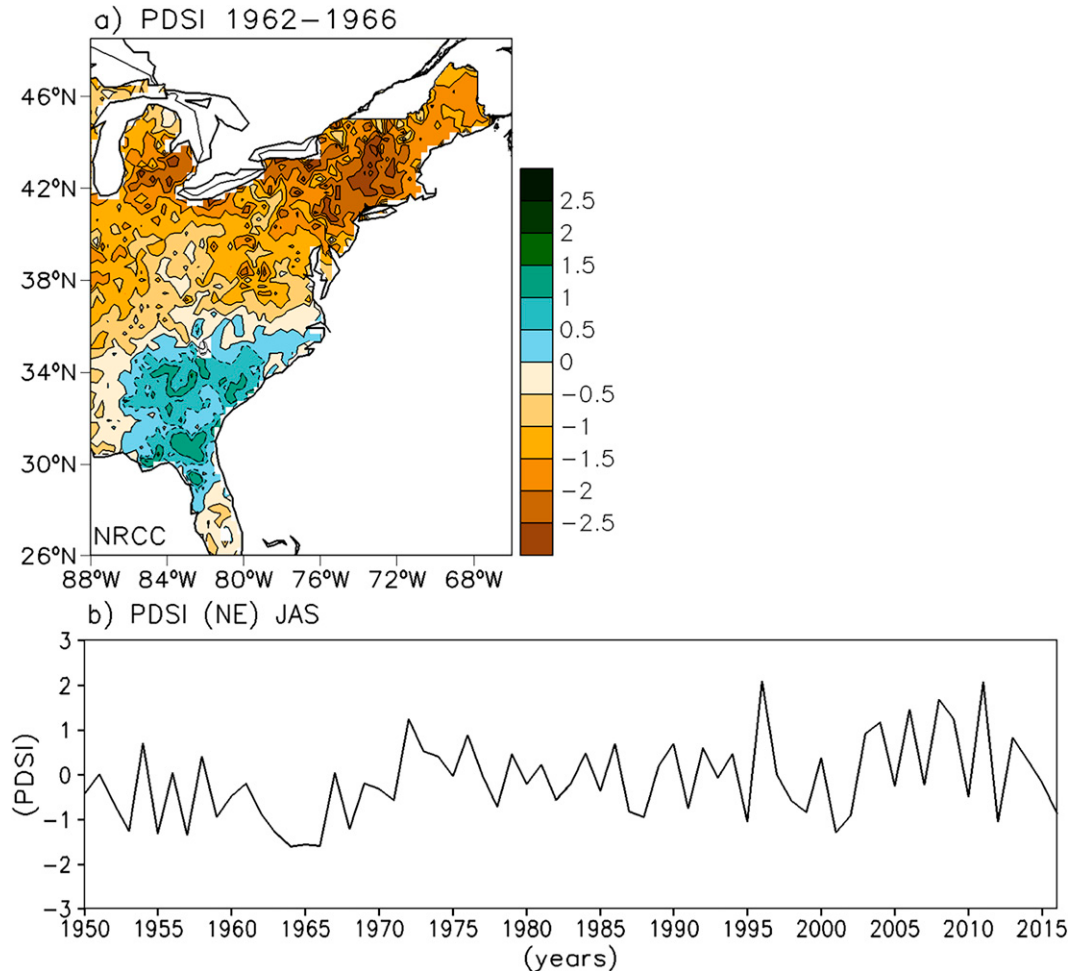


FIG. 1. (a) The Palmer drought severity index (PDSI) map averaged for the period January 1962–December 1966 that highlights the drought in the Northeast (NE), and (b) the PDSI time series for the July–September (JAS) season. The dataset to compute the PDSI is from the Northeast Regional Climate Center (NRCC), and the interannual variation of the PDSI is for the NE region (37° – 48° N, 83° – 67° W). The calibration period for the PDSI is 1950–80.

the coarse $1^{\circ} \times 1^{\circ}$ datasets, and this BC-SD enabled us to handle a higher-resolution dataset (32 km). Nevertheless, BC-SD revealed better details in the NE region with some predictability enhanced during all the three lead months for temperature and precipitation. It can be noted that for precipitation, the correlation values increased for the lead 1 in the Massachusetts region. Although the values are local significant ($p < 0.05$), as indicated with the oblique lines in the maps, the map is not global significant ($f > 85\%$; Livezey and Chen 1983). Therefore, we could conclude that BC-SD inherits the predictability range of the BC approach. This value added is potentially due to the inclusion of higher spatial resolution in the observed input data at this 32-km resolution. This information supports that using a dynamically downscaled approach [e.g., with the Weather Research and Forecasting (WRF) Model] on the coarse-resolution NMME dataset can have an improved forecast outcome (Castro et al. 2012). The

region with negative correlation is a potential target region where the postprocessing EMOS could have a positive impact.

A similar analysis with correlation and skill score was done for PDSI (Fig. 4), which showed the spatial pattern of PDSI predictability measured with correlation (left panel) and skill score (right panel). In the NE region, a correlation of 0.4 or higher was observed, but the skill score dropped after the second month in some regions. Correlation patterns showed values on the order of (0.2, 0.3) for the majority of the domain, with the highest values in the lead-0 map. The magnitude of these numbers confirms that the BC-SD of PDSI is a useful product for a 1-month forecast window and maybe two months. However, as correlation showed only the skill of the transition and not the amplitude, a better metric that evaluates amplitude (SScore) is also shown (Fig. 4, right). Values of SScore are relatively high for the first month, and for the other lead months only in the southern states (Florida and Georgia).

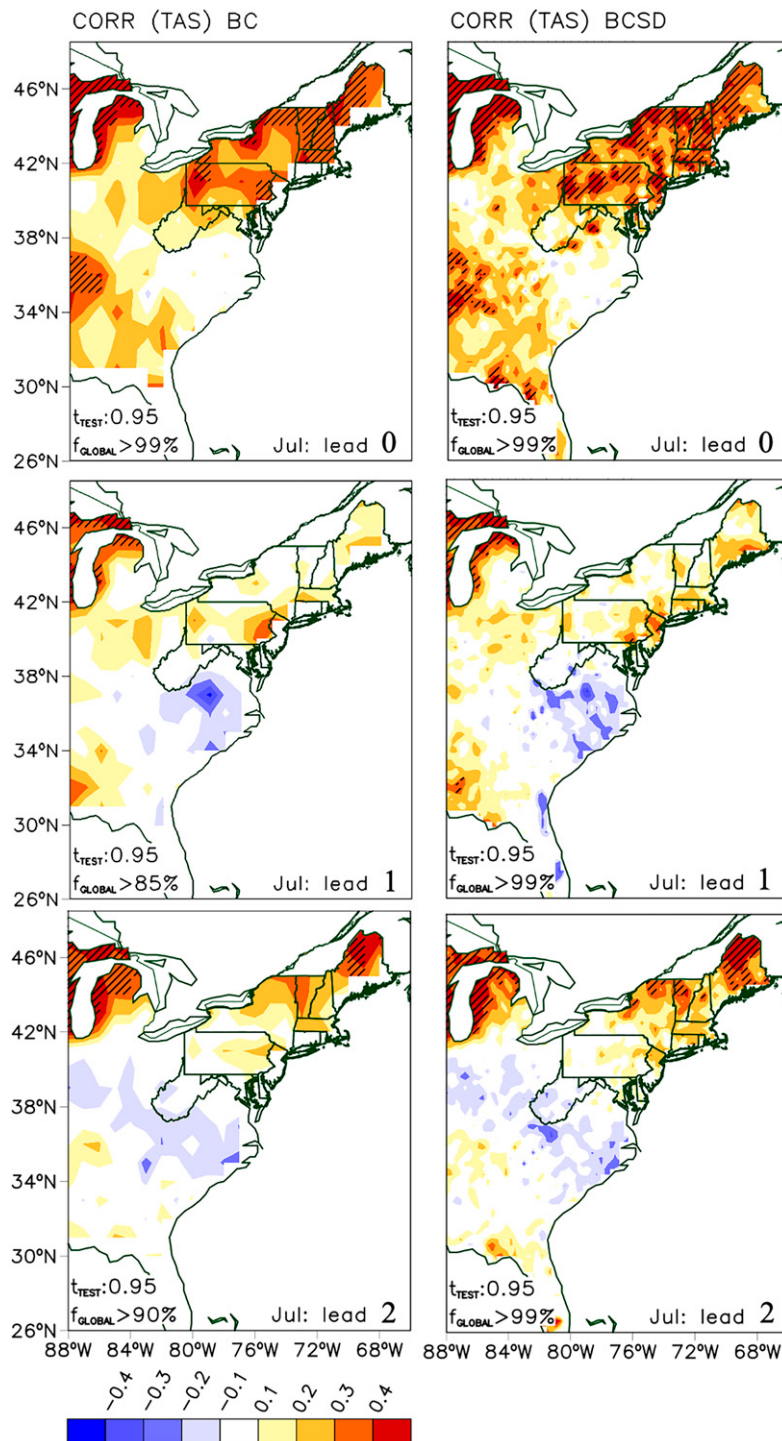


FIG. 2. (left) Spatial patterns of correlation map (CORR) for surface temperature (TAS) using a bias correction (BC) approach to correct temperature bias. Each panel shows different lead times (0, 1, and 2 months) for the same initialization on 1 Jul. The resolution of the data is $1^\circ \times 1^\circ$. (right) As in the left panels, but this case used a bias correction-spatial disaggregation (BC-SD) approach for a target resolution of 32 km. Local significance uses t distribution and is shown with oblique lines; global significance uses a nonparametric Monte Carlo distribution with 500 random permutations and is shown in percentage (Livezey and Chen 1983).

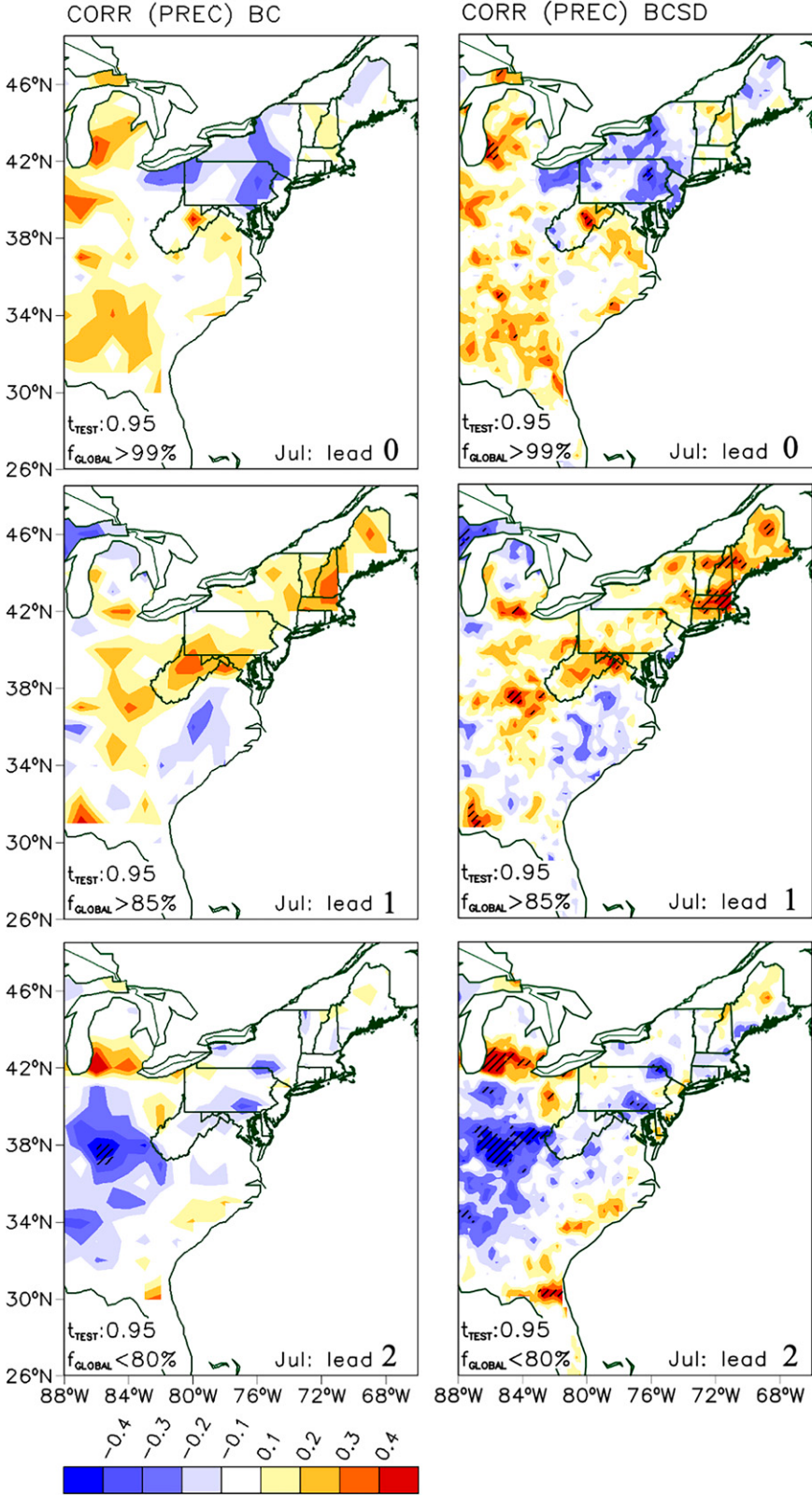


FIG. 3. As in Fig. 2, but for precipitation (PREC).

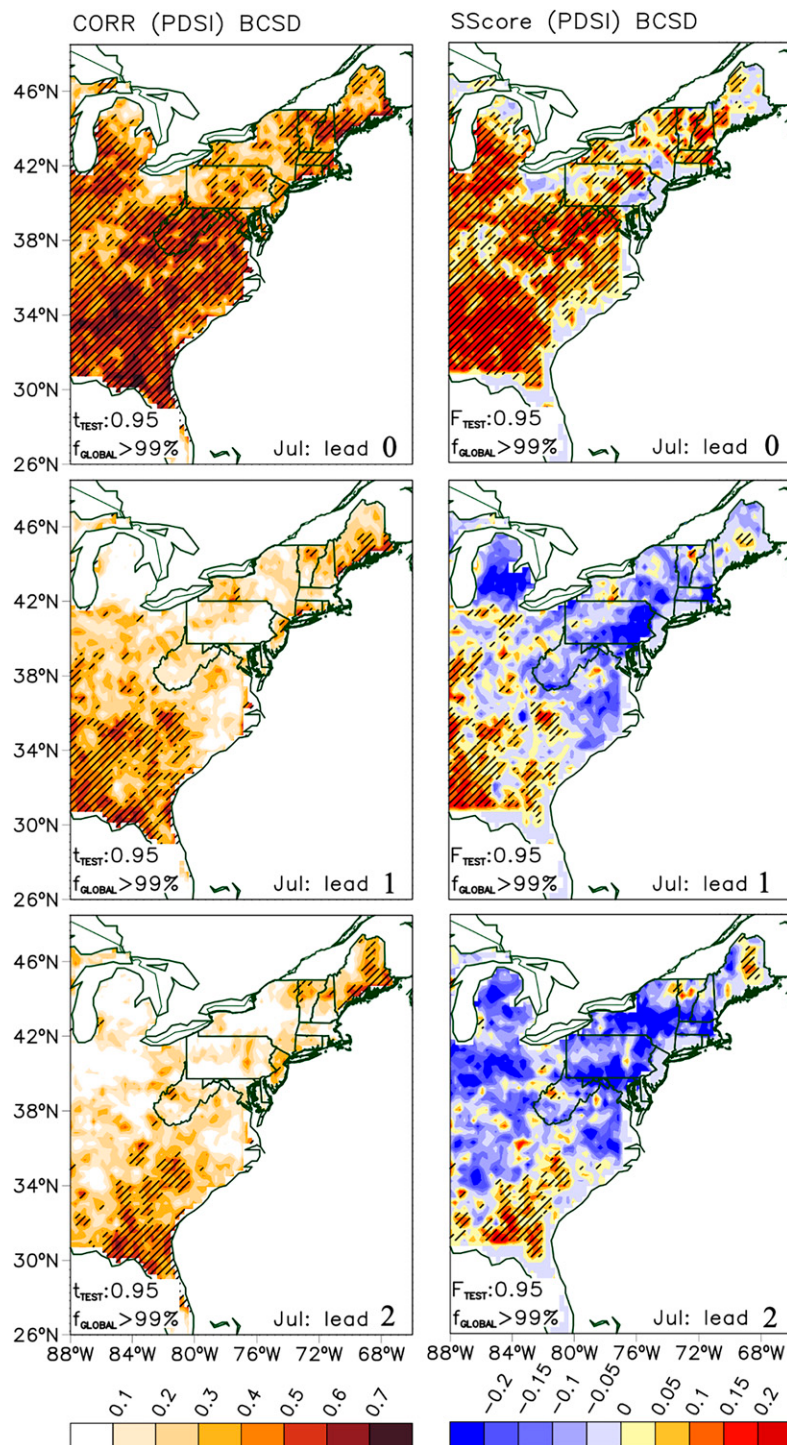


FIG. 4. (left) Spatial patterns of correlation map (CORR) for PDSI using a BC-SD approach to correct PDSI bias. Each subplot in this panel shows different lead times (0, 1, and 2 months) for the same initialization on 1 Jul. (right) As in the left panels, but for the reduction of variance skill score (SScore) instead of correlation. The resolution for both panels is $32 \text{ km} \times 32 \text{ km}$. Local significance—using t distribution (left) and F distribution (right)—is shown with oblique lines; global significance uses a nonparametric Monte Carlo distribution with 500 random permutations and is shown in percentage (Livezey and Chen 1983).

Here, positive values define a percentage of improvement, so any positive value indicates a positive performance. The forecast failed to get the signal of drought according to the SScore with an emphasis in the northern region of the domain for lead 1 and 2. Therefore, PDSI forecast postprocessed with BC-SD can be of valuable use for the first month. The skill is better observed in the southern states than in the northern states. This might be different for other places such as the U.S. Southwest due to the sensitivity difference in the approach used to compute the potential evapotranspiration (van der Schrier et al. 2011). Also, the signal decayed in time, and a portion of the skill could be due to the soil storage residual memory in the PSDI (Palmer 1965). We argue next that the role of the ensemble in these results is significant, but it is revealed only with the use of EMOS. In other words, multimodel ensemble could be used to add value to the forecasts.

c. The forecast skill due to the EMOS postprocessing

The variability in the NE and the limited results shown with the correlation and SScore metrics directed our attention to evaluate whether an EMOS postprocessing approach can help to incorporate new aspects into the drought predictability. If the decadal effect on predictability is estimated using EMOS, what we are developing here is the engineering to bridge the gap between multiple scales in the climate system (intraseasonal to decadal variability) and implementing it in an operational framework.

First, we performed the EMOS to the computed PDSI with temperature and precipitation without bias correction, only including the SD to enhance the resolution. Using all model ensembles (10 ensemble members for each model from a total of five models makes a total of 50 cases per initialized reforecast), the performance of EMOS on PDSI using correlation is shown in Fig. 5 (left). It indicates a very significant value added of the EMOS approach when compared against BC-SD, which is noted for all the three forecast target months. For the three lead times, the correlation value overall is between 0.4 and 0.5 in the majority of the domain, which is significantly higher than the BC-SD approach (0.2–0.3). For the first lead month, correlation values higher than 0.7 are observed in large patches. Also, an improvement is observed in the skill score of PDSI using SScore (Fig. 5, right). However, this improvement happens mostly in the NE region. The results presented here support the conclusion that there is an improvement in the low forecast skill from previous approaches (i.e., BC and BC-SD) up to the second month (lead 1) in the NE, which is also the case for analysis of the entire United States (Fig. S4). According to EMOS, the western U.S. region has better forecast results, but that is not the case for the Great Plains, where the latent and sensible heat of the diurnal cycle and synoptic scale have a higher interaction with the climate system. The EMOS approach is a powerful tool, as it assessed, objectively, which model better contributes to the predictability signal. Then, extrapolating this result with the EMOS metric, a similar approach can be used to assess

uncertainty of drought in climate projection datasets (e.g., CMIP3, CMIP5, or CMIP6).

Second, we modified the training period (15, 20, 25, and 30 years) of the EMOS-NGR function with the idea that decadal training periods in the order of the decadal should outperform other ranges—the previous analysis was done with a training period of 30 years. Figure 6 shows the analysis for PDSI for a region of $5^\circ \times 5^\circ$ center at the point 42°N , 77°W using two metrics: correlation (top) and CRPS (bottom). Using correlation, for the first lead month, the 25- and 30-yr training outperformed the other 15- and 20-yr case, which is also valid for the second and third lead month. The signal of outperforming with a long-term training is confirmed with the continuous ranked probability score (CRPS) metric, where the training period close to 25 years showed the best results for the three lead months (lower values of CRPS represent relative better skill). For the 0 and 1 lead months, using the same CRPS metric, the signal of improvement is much clearer with both training periods: 25 and 30 years. Is this decadal signal responsible for the improvement? If yes, the calculation of the SS for the entire domain (Fig. 5, right panel) should suggest an improvement that matches the spatial pattern of the decadal PDSI (Fig. 1a), and this matching pattern is indeed noted in Fig. 5, at least for the first and second lead time. This outcome can be seen as an indirect validation that the decadal variability is being assimilated in the training period. However, the training record is short (30 years) to fully confirm this observation.

How is EMOS able to provide this improvement, and how can we use it in a drought monitoring tool? The EMOS-NGR adjusts the probability forecast of PDSI by including the variability of the multiple ranges in the NE. The results presented here (Figs. 4–6) show the EMOS effect on the decadal range that improves the forecasts. Our analysis strongly suggests that this is the effect of the decadal variability. It was shown by Woollings et al. (2015) that temperature and precipitation in the NE are affected by a decadal variation of the climate. Our results show that the EMOS technique is able to incorporate the decadal signal into the forecast postprocessing. These results provide good insight that the signal of predictability exists up to a 2-month lead time. In addition, using the findings of previous works that have shown the value added of using a dynamically downscaled approach, we could add value to the forecast with this approach, but objective proof of this can be shown at the cost of a higher computational effort.

4. Conclusions

This study evaluated the predictability (forecast skill) of drought in the Northeast United States. The central elements are the merged information among PDSI, NMME, and EMOS. The PDSI was used as a metric to define drought variability at monthly sampling. We used the reforecast data from the NMME, which is the most comprehensive set of multimodel ensemble simulations currently available, to compute PDSI using the Thornthwaite relationship for the estimation of evapotranspiration. We hypothesize that by adding

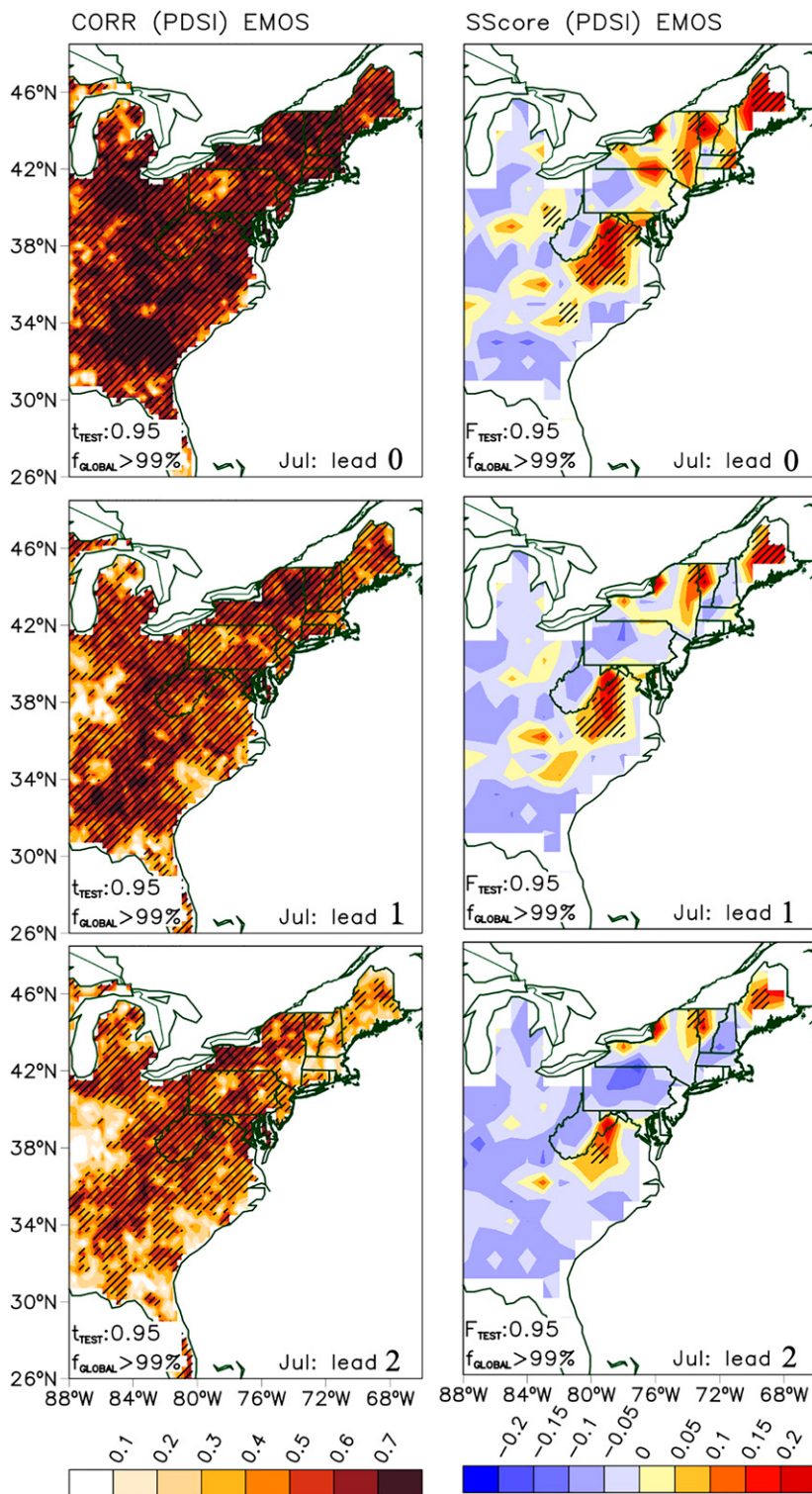


FIG. 5. (left) Spatial pattern of correlation (CORR) and (right) the reduction of variance skill score (SScore) of the PDSI. Each panel shows different lead times (0, 1, and 2 months) for the same initialization on 1 Jul. In both cases the EMOS was applied to the following NMME models: CanCM3, CanCM4, CESM1, FLOR01, and GEOS-5. The number of ensemble members per model is 10. Local and global significance were calculated as in Fig. 4.

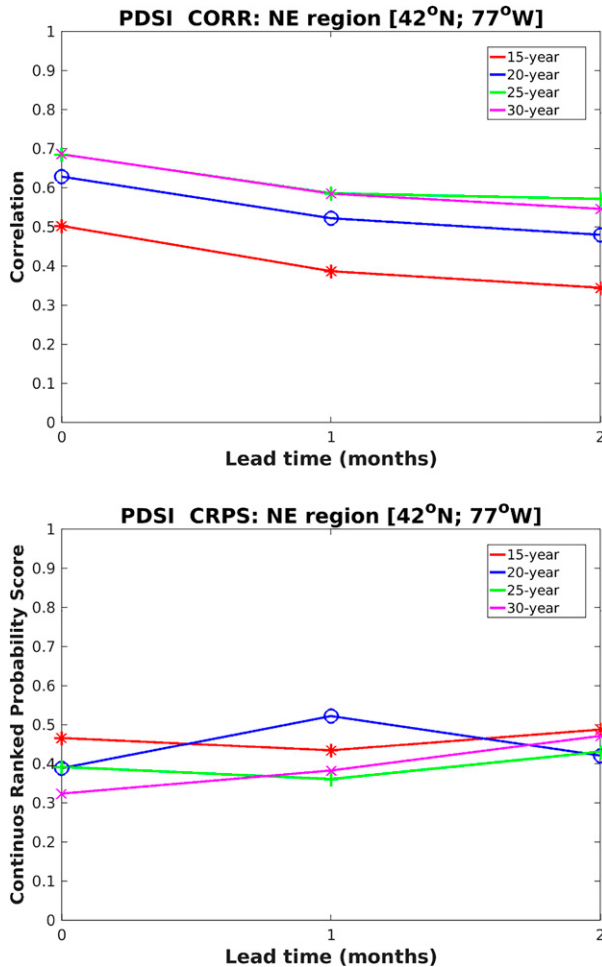


FIG. 6. (top) Correlation (CORR) and (bottom) the continuous ranked probability score (CRPS) metrics for the PDSI forecast. The x axis is the lead time in months from 1 to 3. Each line shows different training periods (15, 20, 25, and 30 years) for the same initialization (1 Jul). In both cases the EMOS was applied to the following NMME models: CanCM3, CanCM4, CESM1, FLOR01, and GEOS-5 for an area of $5^\circ \times 5^\circ$ center at the grid point in the Northeast region (42°N , 77°W).

long-term training of the variability of drought in the NE, it can have a positive influence on the predictability of drought in the region. Two postprocessing techniques were used. 1) The BC-SD method improves spatial resolution that allows using refined soil information introduced in the available water capacity (AWC) of the PDSI calculation to better assess water deficit that better estimates drought variability. 2) The EMOS approach, known as nonhomogeneous Gaussian regression (Gneiting et al. 2005), systematically includes the decadal information from the multimodel ensemble simulations.

By using the BC-SD approach, we created a baseline of high-resolution (32 km) PDSI for the two databases: observed (NRCC) and reforecasted (NMME). A comparison of the forecast skill in two resolutions using BC ($1^\circ \times 1^\circ$)

and BC-SD ($32 \text{ km} \times 32 \text{ km}$) shows that the statistically downscaled approach is consistent and replicable. The lead time of PDSI predictability with BC-SD is on the order of one month, which is consistent when compared with other studies using precipitation and temperature (Koster et al. 2011; Becker et al. 2014; Infanti and Kirtman 2014). However, the values shown for PDSI are provided for the first time in this study. The BC-SD does not include information on the multimodel ensemble simulations, as it only uses the mean average contained in the ensemble realizations.

The most relevant outcome of this study is the improved forecasting skill of PDSI when using EMOS. Following previous work (e.g., Carrillo et al. 2018), this study shows that the postprocessing with EMOS using the sophisticated training at the proper length of time provides a better estimation of mean and dispersion errors (Wilks 2018) that could be removed in the operational forecast.

As hypothesized, the cases with a longer training period (e.g., decadal) show a significant improvement in comparison with a shorter period. Previous studies using EMOS showed that longer training periods and longer ensemble datasets could produce high skill scores, and this seems to be a logical idea. In a related study, we showed this by evaluating the skill forecast of an index for the spring onset in North America (Carrillo et al. 2018). However, this study suggests that there is an exception when a signal of quasiperiodic variation (e.g., decadal) exists, and it is important to explain the variability of the climate regime in the region. This could potentially happen if two aspects of the climate system are at play. First, the region selected should have a clear quasiperiodic signal (e.g., decadal) for the results presented). Second, employing a multimodel ensemble that allows us to perform an EMOS that helps to use this low-frequency signal to train postprocessing on a probabilistic forecast. In this study, the EMOS approach performs a kind of weighting function of the multiple models or a rank selection of the model.

The implication of this study is that the predictability forecast of drought (PDSI) can be extended without any change in the core dynamics of the model but instead by using a sophisticated postprocessing technique. However, a few caveats in this study are disclosed. 1) The dataset, although unique, is accessed at a monthly sample. Then, the intraseasonal signal (30–60 day) where most of the predictability lies might be highly reduced, and therefore, an easy improvement is to do the analysis on a daily sample. This temporally increased resolution could be clearly noted in the location of the low skill score, which follows the synoptic–intraseasonal scale interaction. 2) The repetition of models (two from a total of five) and a limited ensemble member population (10 ensemble members per model). Also, this study clarifies that the two models are improved versions of the same model at different generations. Could this model repetition create a bias in the final result? Yes, it could, if the data product is treated as a simple multimodel ensemble average, but not here because the EMOS–NGR is used to compute the distribution of both the bias and dispersion error. In the EMOS–NGR approach, adding a new model generation helps to detect persistent errors that were carried from an earlier version of the same

model (e.g., CanCM3 and CanCM4). How predictable is short-term drought in the northeastern United States? From previous analysis on temperature and precipitation, 1 month is the most extended range we can expect (Kirtman et al. 2014; Saha et al. 2014), which is below the range of the seasonal scale presented. Here, with EMOS, this range is 2 months. These results can guide us to modify and tune other synoptic–intraseasonal modeling tools.

Acknowledgments. The authors thank the North American Multi-Model Ensemble (NMME) project for providing the dataset. NMME project is supported by NOAA, NFS, NASA, and DOE. This work was partially supported by USDA Grant 1010630 (Project NYC-124439).

Data availability statement. All NMME data used during this study are openly available from NOAA Climate Prediction Center (CPC) and distributed by the NMME project at <https://www.cpc.ncep.noaa.gov/products/NMME/data.html> as cited by Kirtman et al. (2014). Dissemination of the data archive is supported by NOAA, NSF, and DOE. Continuous updating and maintaining is provided by BCEP, IRI, and NCAR personnel.

REFERENCES

- Alley, W. M., 1984: The Palmer drought severity index: Limitations and assumptions. *J. Climate Appl. Meteor.*, **23**, 1100–1109, [https://doi.org/10.1175/1520-0450\(1984\)023<1100:TPDSIL>2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023<1100:TPDSIL>2.0.CO;2).
- Barlow, M., S. Nigam, and E. H. Berbery, 2001: ENSO, Pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow. *J. Climate*, **14**, 2105–2128, [https://doi.org/10.1175/1520-0442\(2001\)014<2105:EPDVAU>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<2105:EPDVAU>2.0.CO;2).
- Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, <https://doi.org/10.1175/JCLI-D-13-00597.1>.
- Bishop, D. A., and C. M. Beier, 2013: Assessing uncertainty in high-resolution spatial climate data across the US Northeast. *PLOS ONE*, **8**, e70260, <https://doi.org/10.1371/journal.pone.0070260>.
- Bolinger, R. A., A. D. Gronewold, K. Kompoltowicz, and L. M. Fry, 2017: Application of the NMME in the development of a new regional seasonal climate forecast tool. *Bull. Amer. Meteor. Soc.*, **98**, 555–564, <https://doi.org/10.1175/BAMS-D-15-00107.1>.
- Briffa, K. R., P. D. Jones, and M. Hulme, 1994: Summer moisture variability across Europe, 1892–1991: An analysis based on the palmer drought severity index. *Int. J. Climatol.*, **14**, 475–506, <https://doi.org/10.1002/joc.3370140502>.
- Carrillo, C. M., T. R. Ault, and D. S. Wilks, 2018: Spring onset predictability in the North American multimodel ensemble. *J. Geophys. Res. Atmos.*, **123**, 5913–5926, <https://doi.org/10.1029/2018JD028597>.
- Castro, C. L., H. Chang, F. Dominguez, C. Carrillo, J.-K. Schemm, and H.-M. H. Juang, 2012: Can a regional climate model improve warm season forecasts in North America? *J. Climate*, **25**, 8212–8237, <https://doi.org/10.1175/JCLI-D-11-00441.1>.
- Dai, A., K. E. Trenberth, and T. Qian, 2004: A global data set of Palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming. *J. Hydrometeor.*, **5**, 1117–1130, <https://doi.org/10.1175/JHM-386.1>.
- DeGaetano, A. T., and B. N. Belcher, 2007: Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *J. Appl. Meteor. Climatol.*, **46**, 1981–1992, <https://doi.org/10.1175/2007JAMC1536.1>.
- , and D. S. Wilks, 2009: Radar-guided interpolation of climatological precipitation data. *Int. J. Climatol.*, **29**, 185–196, <https://doi.org/10.1002/joc.1714>.
- Deser, C., and M. L. Blackmon, 1995: On the relationship between tropical and North Pacific sea surface temperature variations. *J. Climate*, **8**, 1677–1680, [https://doi.org/10.1175/1520-0442\(1995\)008<1677:OTRBTA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1677:OTRBTA>2.0.CO;2).
- Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987, [https://doi.org/10.1175/1520-0450\(1969\)008<0985:ASSFPF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2).
- Fan, F., R. S. Bradley, and M. A. Rawlins, 2015: Climate change in the Northeast United States: An analysis of the NARCCAP multimodel simulations. *J. Geophys. Res. Atmos.*, **120**, 10 569–10 592, <https://doi.org/10.1002/2015JD023073>.
- Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, <https://doi.org/10.1175/MWR2904.1>.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus*, **57A**, 219–233, <https://doi.org/10.1111/j.1600-0870.2005.00103.x>.
- Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, [https://doi.org/10.1175/1520-0493\(2004\)132<1434:ERIMFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2).
- Hao, Z., V. P. Singh, and Y. Xia, 2018: Seasonal drought prediction: Advances, challenges, and future prospects. *Rev. Geophys.*, **56**, 108–141, <https://doi.org/10.1002/2016RG000549>.
- Hayhoe, K., and Coauthors, 2007: Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dyn.*, **28**, 381–407, <https://doi.org/10.1007/s00382-006-0187-8>.
- Hurrell, J. W., Y. Kushnir, and M. Visbeck, 2001: The North Atlantic oscillation. *Science*, **291**, 603–605, <https://doi.org/10.1126/science.1058761>.
- Infanti, J. M., and B. P. Kirtman, 2014: Southeastern U.S. rainfall prediction in the North American Multi-Model Ensemble. *J. Hydrometeor.*, **15**, 529–550, <https://doi.org/10.1175/JHM-D-13-072.1>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYP>2.0.CO;2).
- Kern, J. D., Y. Su, and J. Hill, 2020: A retrospective study of the 2012–2016 California drought and its impacts on the power sector. *Environ. Res.*, **15**, 094008, <https://doi.org/10.1088/1748-9326/ab9db1>.
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Knighton, J., G. Pleiss, E. Carter, S. Lyon, M. T. Walter, and S. Steinschneider, 2019: Potential predictability of regional precipitation and discharge extremes using synoptic-scale climate

- information via machine learning: An evaluation for the eastern continental United States. *J. Hydrometeorol.*, **20**, 883–900, <https://doi.org/10.1175/JHM-D-18-0196.1>.
- Koster, R. D., and Coauthors, 2011: The second phase of the Global Land-Atmosphere Coupling Experiment: Soil moisture contributions to subseasonal forecast skill. *J. Hydrometeorol.*, **12**, 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright, 1998: Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.*, **9**, 112–147, <https://doi.org/10.1137/S1052623496303470>.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59, [https://doi.org/10.1175/1520-0493\(1983\)111<0046:SFAID>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<0046:SFAID>2.0.CO;2).
- Lohani, V. K., G. V. Loganathan, and S. Mostaghimi, 1998: Long-term analysis and short-term forecasting of dry spells by Palmer Drought Severity Index. *Hydrol. Res.*, **29**, 21–40, <https://doi.org/10.2166/nh.1998.0002>.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. ECMWF Seminar on Predictability*, Vol. I, Reading, United Kingdom, ECMWF, 1–18.
- Lorenz, D. J., J. A. Otkin, M. Svoboda, C. R. Hain, M. C. Anderson, and Y. Zhong, 2017: Predicting the U.S. Drought Monitor using precipitation, soil moisture, and evapotranspiration anomalies. Part II: Intraseasonal drought intensification forecasts. *J. Hydrometeorol.*, **18**, 1963–1982, <https://doi.org/10.1175/JHM-D-16-0067.1>.
- Maurer, E. P., and H. G. Hidalgo, 2008: Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods. *Hydrol. Earth Syst. Sci.*, **12**, 551–563, <https://doi.org/10.5194/hess-12-551-2008>.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. *Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–184.
- Notaro, M., W.-C. Wang, and W. Gong, 2006: Model and observational analysis of the Northeast U.S. regional climate and its relationship to the PNA and NAO patterns during early winter. *Mon. Wea. Rev.*, **134**, 3479–3505, <https://doi.org/10.1175/MWR3234.1>.
- ORNL DAAC, 2017: Spatial Data Access Tool (SDAT). ORNL DAAC, accessed 21 April 2017, <https://doi.org/10.3334/ORNLDAAAC/1388>.
- Palmer, W. C., 1965: Meteorological drought. U.S. Weather Bureau Research Paper 45, 58 pp., https://www.droughtmanagement.info/literature/USWB_Meteorological_Drought_1965.pdf.
- Palmer, T. N., and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872, <https://doi.org/10.1175/BAMS-85-6-853>.
- Palmer, J. G., and Coauthors, 2015: Drought variability in the eastern Australia and New Zealand summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific Oscillation. *Environ. Res. Lett.*, **10**, 124002, <https://doi.org/10.1088/1748-9326/10/12/124002>.
- Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.
- Power, S., F. Tseitkin, S. Torok, and B. Lavery, 1998: Australian temperature, Australian rainfall and the Southern Oscillation, 1910–1992: coherent variability and recent changes. *Aust. Meteor. Mag.*, **47**, 85–101.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Sims, A. P., D. S. Niyogi, and S. Raman, 2002: Adopting drought indices for estimating soil moisture: A North Carolina case study. *Geophys. Res. Lett.*, **29**, 1183, <https://doi.org/10.1029/2001GL013343>.
- Sweet, S. K., D. W. Wolfe, A. DeGaetano, and R. Benner, 2017: Anatomy of the 2016 drought in the northeastern United States: Implications for agriculture and water resources in humid climates. *Agric. For. Meteorol.*, **247**, 571–581, <https://doi.org/10.1016/j.agrformet.2017.08.024>.
- Tortajada, C., M. J. Kastner, J. Buurman, and A. K. Biswas, 2017: The California drought: Coping responses and resilience building. *Environ. Sci. Policy*, **78**, 97–113, <https://doi.org/10.1016/j.envsci.2017.09.012>.
- van der Schrier, G., P. D. Jones, and K. R. Briffa, 2011: The sensitivity of the PDSI to the Thornthwaite and Penman-Monteith parameterizations for potential evapotranspiration. *J. Geophys. Res.*, **116**, D03106, <https://doi.org/10.1029/2010JD015001>.
- Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Climate*, **23**, 1696–1718, <https://doi.org/10.1175/2009JCLI2909.1>.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2008: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quart. J. Roy. Meteor. Soc.*, **134**, 241–260, <https://doi.org/10.1002/qj.210>.
- Wells, N., S. Goddard, and M. J. Hayes, 2004: A self-calibrating Palmer drought severity index. *J. Climate*, **17**, 2335–2351, [https://doi.org/10.1175/1520-0442\(2004\)017<2335:ASPDSEI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2335:ASPDSEI>2.0.CO;2).
- Wilks, D. S., 2006: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteor. Appl.*, **13**, 243–256, <https://doi.org/10.1017/S1350482706002192>.
- , 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 704 pp.
- , 2018: Univariate ensemble postprocessing. *Statistical Postprocessing of Ensemble Forecast*, 1st ed. S. Vannitsem, D. S. Wilks, and J. W. Messner, Eds., Elsevier, 49–89, <https://doi.org/10.1016/B978-0-12-812372-0.00003-0>.
- , and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390, <https://doi.org/10.1175/MWR3402.1>.
- Woollings, T., C. Franzke, D. L. R. Hodson, B. Dong, E. A. Barnes, C. C. Raible, and J. G. Pinto, 2015: Contrasting interannual and multidecadal NAO variability. *Climate Dyn.*, **45**, 539–556, <https://doi.org/10.1007/s00382-014-2237-y>.
- Yang, X., X. Xu, A. Stovall, M. Chen, and J.-E. Lee, 2021: Recovery: Fast and slow—Vegetation response during the 2012–2016 California drought. *J. Geophys. Res. Biogeosci.*, **126**, e2020JG005976, <https://doi.org/10.1029/2020JG005976>.
- Yuan, X., and E. F. Wood, 2013: Multimodel seasonal forecasting of global drought onset. *Geophys. Res. Lett.*, **40**, 4900–4905, <https://doi.org/10.1002/grl.50949>.
- Zhang, Y., J. Norris, and J. Wallace, 1998: Seasonality of large-scale atmosphere-ocean interaction over the North Pacific. *J. Climate*, **11**, 2473–2481, [https://doi.org/10.1175/1520-0442\(1998\)011<2473:SOLSAO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2473:SOLSAO>2.0.CO;2).