BAMS Article

A Decade of the North American Multimodel Ensemble (NMME)

Research, Application, and Future Directions

Emily J. Becker, Ben P. Kirtman, Michelle L'Heureux, Ángel G. Muñoz, and Kathy Pegion

ABSTRACT: Ten years, 16 fully coupled global models, and hundreds of research papers later, the North American Multimodel Ensemble (NMME) monthly-to-seasonal prediction system is looking ahead to its second decade. The NMME comprises both real-time, initialized predictions and a substantial research database; both retrospective and real-time forecasts are archived and freely available for research and development. Many U.S.-based and international entities, both private and public, use NMME data for regional or otherwise tailored forecasts. The system's built-in evolution, with new models gradually replacing older ones, has been demonstrated to gradually improve the skill of 2-m temperature and sea surface temperature, although precipitation prediction remains a difficult problem. This paper reviews some of the NMME-based contributions to seasonal climate prediction research and applications, progress on scientific understanding of seasonal prediction and multimodel ensembles, and new techniques. Several prediction-oriented aspects are explored, including model representation of observed trends and the underprediction of below-average temperature. We discuss potential new directions, such as higher-resolution models, hybrid statistical–dynamical techniques, or prediction of environmental hazards such as coastal flooding and the risk of mosquito-borne diseases.

KEYWORDS: Climate prediction; Ensembles; Forecasting techniques; Seasonal forecasting; Climate models

https://doi.org/10.1175/BAMS-D-20-0327.1

Corresponding author: Emily J. Becker, emily.becker@rsmas.miami.edu

In final form 14 December 2021

AFFILIATIONS: Becker and Kirtman—Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida; **L'Heureux**—NOAA/NWS/NCEP/Climate Prediction Center, College Park, Maryland; **Muñoz**—International Research Institute for Climate and Society, The Earth Institute, Columbia University, Palisades, New York; **Pegion**—Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, Virginia

In September of 2010, the National Research Council recommended that "multi-model ensemble (MME) forecast strategies should be pursued" to improve the building blocks of intraseasonal-to-interannual prediction (National Research Council 2010). On 8 August 2011, the first monthly predictions from the National Multimodel Ensemble (NMME), as it was called at the time, were provided to NOAA's operational climate prediction forecasters and to the public. The NMME of 2011 captured all available U.S. coupled global climate models—seven models from five operational and research centers. Implementation requirements were designed to optimize the information provided to forecasters at NOAA's Climate Prediction Center (CPC), while all forecast data would be published in real time for research, development, and prediction. The NMME (now the North American Multimodel Ensemble; Kirtman et al. 2014) turned 10 years old in 2021, and we mark this milestone by looking back at some of the scientific contributions of the project and forward to future directions.

The NMME protocol designed in 2011 established requirements for horizontal resolution (1° longitude × 1° latitude), number of lead months (at least 9), retrospective forecasts (1982–2010), and delivery date (eighth of each month). The retrospective forecasts, also called hindcasts, were required to use the same version and initialization as the real-time forecast model version, allowing for analysis of model forecast characteristics and correction of systematic bias. Also specified was that the monthly-mean forecast variables—2-m temperature, precipitation rate, and sea surface temperature—were provided as total fields, i.e., not anomalies, so that the forecast users could apply bias correction as they found appropriate. Beyond this, all details of the models were left up to the modeling centers, including model physics, initialization and ensembling strategies, and native resolution.

This relatively spare protocol encouraged as much model diversity as possible, while minimizing the technical burden on CPC and other forecast data users. The modeling centers were able to run retrospective forecasts during the spring and early summer of 2011, and the first real-time forecasts were provided in early August. That first ensemble comprised seven models from five centers: two each from NOAA's National Centers for Environmental Prediction (NCEP) and the International Research Institute for Climate and Society (IRI) at Columbia University, and one each from NASA's Global Modeling and Assimilation Office (GMAO), the University of Miami, and NOAA's Geophysical Fluid Dynamics Laboratory (GFDL). See Table 1 for model details and active years. CPC's NMME website published images for each model's anomaly forecasts and the multimodel ensemble-mean anomaly, while the IRI Data Library made all forecast data available to researchers and the public.

In August of 2012, the IRI models were retired, and Environment and Climate Change Canada (ECCC) joined the NMME, which was renamed the North American Multimodel Ensemble. To expedite the transmission of some retrospective forecast data, ECCC copied the data for the two models that form their Canadian Seasonal-to-Interannual Prediction System (CanSIPS; Merryfield et al. 2013) onto surplused hard drives and shipped them to NCAR.

Table 1. All models that have participated in the North American Multimodel Ensemble (NMME) 2011–21. "Years active" indicates the years during which that model provided real-time forecast data to the NMME, while "hindcast years (No. ens. mems)" shows the total number of years of retrospective forecasts and the number of ensemble members. In the case of models with differing ensemble size between the hindcast and real time, the real-time size is denoted with "RT." The "arrangement of ensemble members" column indicates the date and time of initialization of each ensemble member.

Model	Years active	Hindcast years (No. ens. mems)	Arrangement of ensemble members	Leads (months)	Atmosphere	Ocean	Reference
IRI-ECHAM4f	2011–12	1982–2010 (12)	All first of the month 0000 UTC	0–7	T42L19	MOM3L25, 1.5° × 0.5°	DeWitt (2005)
IRI-ECHAM4a	2011–12	1982–2010 (12)	All first of the month 0000 UTC	0–7	T42L19	MOM3L25, 1.5° × 0.5°	DeWitt (2005)
NCEP-CFSv1	2011–12	1982–2010 (15)	First 0000 UTC ± 2 days, twenty-first 0000 UTC ± 2 days, eleventh 0000 UTC ± 2 days	0–8	T62L64	MOM3L40, 0.30°Eq	Saha et al. (2006)
NCAR-RSMAS-CCSM3	2011–14	1982–2010 (6)	All first of the month 0000 UTC	0–11	T85L26	POPL42, 0.3°Eq	Kirtman and Min (2009)
NASA-GMAO-GEOS5	2011–17	1982–2010 (11)	4 members/5 days; 7 last days of last month	0–9	GEOS5 AGCM, 1° × 1.25°, L72	MOM4, L40, 0.5°Eq	Borovikov et al. (2019)
GFDL-CM2.1	2011–20	1982–2010 (10)	All first of the month 0000 UTC	0–11	CM2.1, 2° × 2.5°, L24	MOM4, L50, 0.3°Eq	Delworth et al.(2006)
NCEP-CFSv2	2011–present	1982–2010 (24; 28 Nov)	4 members (0000, 0600, 1200, 1800 UTC) every fifth day	0–9	GFS, T126L64	MOM4, L40, 0.25°Eq	Saha et al. (2014)
ECCC-CanCM3 "CMC1"	2012–19	1982–2011 (10)	All first of the month 0000 UTC	0–11	CanAM3, T63L31	CanOM4, L40, 0.94°Eq	Merryfield et al. (2013)
ECCC-CanCM4 "CMC2"	2012–19	1982–2011 (10)	All first of the month 0000 UTC	0–11	CanAM4, T63L35	CanOM4, L40, 0.94°Eq	Merryfield et al. (2013)
GFDL-CM2.5 "FLOR"	2014–20	1982–2013 (24)	All first of the month 0000 UTC	0–11	CM2.5, C18L32, 50km	MOM5, L50, 0.30°Eq, 1° Polar1.5	Vecchi et al. (2014)
NCAR-SMAS-CCSM4	2014–present	1982–2013 (10)	All first of the month 0000 UTC	0–11	CAM4, 0.9° × 1.25°, L26	POPL60, 0.25°Eq	Infanti and Kirtman (2016)
NCAR-CESM	2015–(research only)	1982–2010 (10)	All first of the month 0000 UTC	0–11	0.9 × 1.25° L72	POPL60, 0.25°Eq	Small et al. (2014)
NASA-GMAO-GEOSS2S	2018–present	1982–2017 (4; 10 RT)	1 member every 5 days	0–9	GEOS5 AGCM, 0.5°, L72	MOM5, L40, 0.5°Eq	Molod et al. (2020)
ECCC-CanCM4i	2019–present	1982–2018 (10)	All first of the month 0000 UTC	0–11	CanAM4, T63L31	CanOM4, L40, 0.94°Eq	Merryfield et al. (2013)
ECCC-GEM-NEMO	2019–present	1982–2018 (10)	All first of the month 0000 UTC	0–11	GEM, 256 × 128	NEMO 1° × 1°, 1/3° Eq	Lin et al. (2020)
GFDL-SPEAR	2021–present	1991–2020 (15; 30 RT)	All first of the month 0000 UTC	0–11	AM4.0, 0.5°, 33 levels	MOM6, 1°, tropical refinement to 0.3°, 75 levels, hybrid vertical coordinate	Delworth et al. (2020)

The resulting six-model NMME configuration (NCEP retired CFSv1 in the fall of 2012) led to a substantial increase in hindcast global-averaged land and ocean surface temperature anomaly correlation over the 2011 model suite, despite having fewer ensemble members overall (Becker et al. 2020).

Over the next 8 years, most models were upgraded to newer versions by the modeling centers, and, by 2021, only NCEP's CFSv2 continued from the original suite. Individual

NMME in a prediction application: The NextGen approach

NMME seasonal prediction information is employed by the IRI and partners via the NextGen methodology, a systematic general approach for codesigning, implementing, producing, and verifying objective forecasts at multiple time scales (Muñoz et al. 2019, 2020; WMO 2020). The approach starts with co-identifying with decision-makers and local experts their concrete demand, which defines the variable(s) to predict. A diagnostic analysis is then conducted to help identify the best observed and modeled predictor variables, including both climate and nonclimate factors. As part of the design and implementation of the NextGen forecast system, past model performance is assessed via a statistical and physical-process-based evaluation, helping inform how to best conduct model calibration and ensemble design. The set of predictions produced by the system includes the full range of possible outcomes of the variable (i.e., its entire probability density function, as opposed to just tercile-based predictions), such that decision-makers can obtain tailored forecasts for any particular threshold of interest, and thus trigger the precise set of actions required.

The NextGen methodology is both general and demand oriented, and has been applied to a wide variety of cases beyond forecasting climate variables such as rainfall or temperature. The range of applications include predictions of environmental suitability for transmission of Aedes-borne diseases such as dengue, Zika or chikungunya (Muñoz et al. 2017, 2020) acute undernutrition for children under 5 years old (Romero et al. 2020; White et al. 2022), coffee yield (Pons et al. 2021), and human migration (Muñoz et al. 2019).

To illustrate the approach with a concrete example using NMME model output, consider the NextGen system for Aedes-borne diseases' environmental suitability (*Ae*DES; Muñoz et al. 2020; see Fig. SB1), developed for a geographical domain encompassing North America, Central America, northern South America, and the Caribbean. Work led by the IRI and the Pan-American Health Organization (PAHO)/World Health Organization (WHO) helped identify environmental suitability for disease transmission as the key variable to monitor and forecast (Muñoz et al. 2020).

*Ae*DES follows a super-ensemble approach, involving four different environmental suitability models that require both entoepidemiological parameters (i.e., transmission-sensitive information involving both the mosquito and the diseases, such as mosquito biting rate, mosquito egg production rate, human susceptibility, and infectious rates; see Muñoz et al. 2020) and environmental variables (such as rainfall, temperature, and relative humidity; see Muñoz et al. 2020). *Ae*DES assimilates 2-m temperature forecasts from the NMME, forcing each one of the four environmental suitability models with 96 NMME climate realizations (i.e., a total of 384 realizations). Presently, 1982–2010 hindcasts for each season are used to calibrate each one of the realizations independently, before computing the ensemble; hence, a total of 4,608 (4 environmental suitability models times 12 initializations times 96 climate model members) 29-yr-long seasonal hindcasts are involved in the calibration and probabilistic forecast production process. The *Ae*DES system uses a pattern-based calibration approach that at the same time conducts a downscaling and corrects for magnitude and spatial pattern biases, such that the final forecasts reproduce the monitored environmental suitability. For additional technical details, see Muñoz et al. (2020).

This type of NextGen forecast system helps the decision-making process by directly providing information on the actual variable used to trigger an action. For example, the allocation of financial and human resources by PAHO/WHO to locations where a risk of disease transmission is expected to be higher (or lower) than normal, depends on the probability of exceeding (or not) a certain value in the risk-related variable, i.e., environmental suitability in the case of the *Ae*DES system.

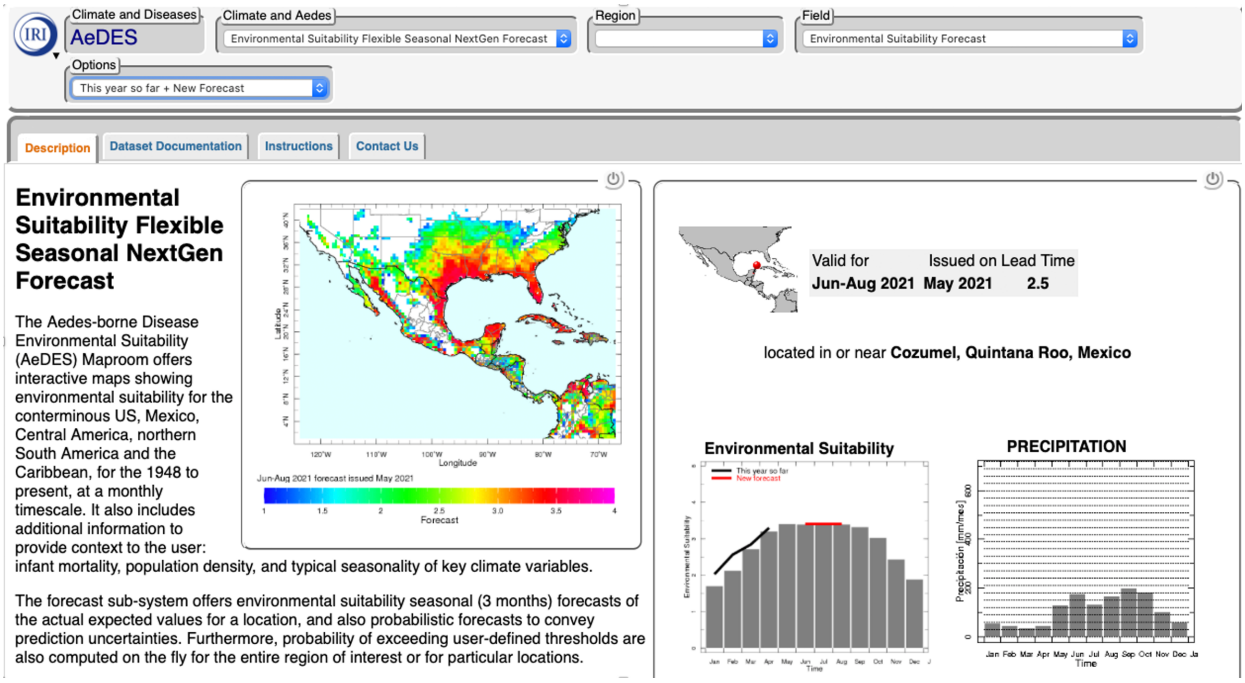

Fig. SB1. Example of the NextGen forecast system from the International Research Institute for Climate and Society's website.

model improvements were demonstrated by their developers (e.g., Lin et al. 2020; Molod et al. 2020; Wittenberg et al. 2018), while the assumption that improved models lead to increased multimodel prediction skill (Yoo and Kang 2005) was tested in Becker et al. (2020). It was found that each evolution of the model suite subsequently increased seasonal forecast skill of land and ocean surface temperature fields, while precipitation skill has shown only very small improvement overall. In Becker et al. (2020), it was suggested that more successful prediction of seasonal precipitation may require a different approach, such as higher-resolution models.

The data from all contributing models have been retained at the IRI Data Library (Blumenthal et al. 2014), creating an extensive database of at least 16 global coupled models, with at least 30 years of forecast data each, for research and development. A description of NMME data availability, access, and tools is provided in the appendix.

Multimodel ensembles. Multimodel ensembles were demonstrated to improve accuracy and quality over single-model ensemble forecasts by the international projects PROVOST (Doblas-Reyes et al. 2000), DEMETER (Palmer et al. 2004), and ENSEMBLES (Weisheimer et al. 2009). In brief, the value of MME stems from the enhancement of signal, cancellation of errors, and the improved ability to characterize the uncertainty of model forecasts (Doblas-Reyes et al. 2005; Hagedorn et al. 2005; Smith et al. 2013). Earlier studies of the NMME have confirmed that both the large total ensemble size—real-time NMME predictions include approximately 100 ensemble members—and the diversity of models results in improved forecast reliability and probabilistic skill scores when compared to smaller ensembles or single-model forecasts (Becker and van den Dool 2016; Tippett et al. 2019).

From a practical perspective, a multimodel ensemble provides continuity of operations if one or more models are unavailable or in error. While in an ideal world this situation would never occur, in this world, computer issues, government shutdowns, and other vagaries have occasionally—albeit infrequently—created obstacles for individual model data delivery. A few months in the past 10 years have found the NMME short of one model on the forecast deadline, but the system was still able to provide forecast data on time from the remaining models, never fewer than five. Forecast information from the NMME has become critical to seasonal prediction operations at NOAA, the U.S. Air Force, and many other centers, including private-sector entities, making reliable availability essential.

Another benefit of a closely monitored multimodel ensemble is the identification and diagnosis of potential model errors. One such error was identified in 2016 by users of NMME El Niño–Southern Oscillation (ENSO) predictions. NMME ENSO information has been available since 2011 in the form of Niño-3.4 "plumes," where the long-lead prediction of the average monthly or seasonal sea surface temperature anomaly (SSTA) in the ENSO monitoring Niño-3.4 region is shown for each ensemble member, the individual ensemble means, and the multimodel ensemble mean. NMME in ENSO prediction will be discussed in more detail in the next section.

Following the strong El Niño event of 2015/16, there was much interest in whether La Niña conditions would develop (L'Heureux et al. 2017). The NMME forecast initialized in early March 2016 showed that most models predicted a continued steep decline in Niño-3.4 SSTA, with ENSO-neutral or La Niña developing by the summer (Fig. 1). However, two models predicted only a slight decrease, followed by continued El Niño. El Niño rarely occurs in consecutive years (An et al. 2020; Choi et al. 2013), leading forecasters to question these two models' predictions. Key to the investigation was that both models, NCEP-CFSv2 and COLA-RSMAS-CCSM4, were initialized using the same reanalysis, CFSR (Saha et al. 2014), while the other models were not. Upon inspection, large erroneous cold subsurface anomalies in the tropical Atlantic were found in the CFSR, and NCEP's Environmental Modeling Center

devised a correction that returned the reanalysis, and consequently the Niño-3.4 prediction, to a realistic state (http://cfs.ncep.noaa.gov/pub/raid0/Briefing.pptx). This fix has been implemented on an ongoing, as-needed basis by NOAA's Environmental Modeling Center.

The NMME in practice

The most established application of NMME is the seasonal prediction of ENSO and North American temperature and precipitation, but NMME's global forecast and straightforward availability has led to the development of specific regional predictions, hybrid statistical–dynamical models, and many other applications. The NMME is frequently used as a baseline for prediction skill by teams developing new or improved prediction systems (e.g., Cohen et al. 2019; Ding et al. 2019; Dias et al. 2019).

ENSO prediction. NMME has become an integral component of NOAA's ENSO prediction (L'Heureux et al. 2019, 2020). In 2013, the NMME providers team agreed to modify the NMME protocol to support a forecast delivery date of close of business on the sixth of each month, in order to better support NOAA's ENSO outlook team. NMME predictions provided to the ENSO team include the SST anomaly in the Niño-3.4 region—the Niño-3.4 index—along with the probability that this index will be > 0.5°C (El Niño threshold), < -0.5°C (La Niña threshold), or between (neutral conditions). Other Niño indexes are of greater relevance for other regions; for example, the Instituto Geofisico del Peru creates NMME plume forecasts for the Niño-1 + 2 region and includes them in their monthly technical briefings, archived at https://repositorio.igp.gob.pe/.

ENSO is highly predictable (in the context of seasonal climate prediction) and NMME prediction of the Niño-3.4 index is overall skillful and reliable (Becker et al. 2014; Becker and van den Dool 2016; L'Heureux et al. 2019). However, failed ENSO forecasts are particularly notable, with a potentially large effect on the accuracy of long-lead seasonal outlooks (Tippett et al. 2020), and prediction of ENSO diversity and impacts is still challenging (e.g., Capotondi et al. 2015; Infanti and Kirtman 2016). Several studies have examined the NMME's ENSO prediction in detail, including both deterministic and probabilistic formats (Barnston et al. 2019; Tippett et al. 2019; respectively). Both of these studies find that the NMME is unable to overcome the well-known "spring predictability barrier," with forecasts for May–September targets substantially less skillful than for other times of the year (see

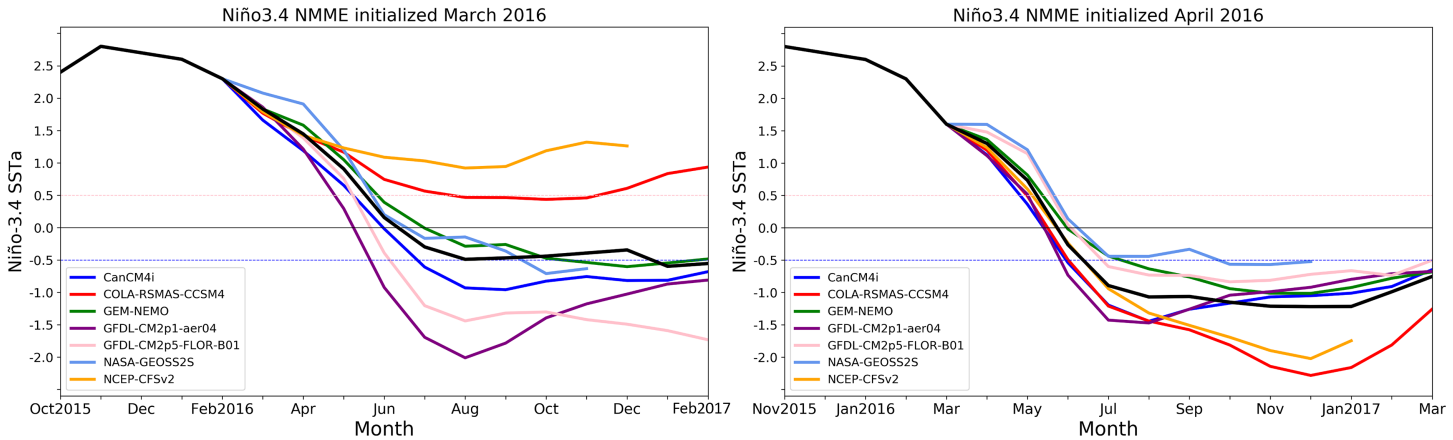

Fig. 1. NMME Niño-3.4 monthly sea surface temperature anomaly forecast initialized in (left) March 2016 and (right) April 2016, reconstructed from seven NMME models: COLA-RSMAS-CCSM4, CanCM4i, GEM-NEMO, GFDL-CM2.1, GFDL-FLORb01, NASA-GEOSS2S, and NCEP-CFSv2. Anomalies are based on a 1986–2015 climatology. Colored lines indicate individual model ensemble means. Black line in the forecast period indicates the equally weighted multimodel ensemble mean. Observations from NOAA's Optimal Interpolation SST (OISST; Reynolds et al. 2002).

also Larson and Kirtman 2017). However, Tippett et al. (2019) find that, using NMME, ENSO forecasts can be expanded from three categories (probabilities of El Niño, neutral, and La Niña using −0.5°C and +0.5°C cutoffs) to seven categories (−1.5°C, −1.0°C, −0.5°C, +0.5°C, +1.0°C, and +1.5°C), without a loss of skill and reliability. L'Heureux et al. (2019) leverage this finding, paving the way for an application that provides stakeholders the probability of Niño-3.4 index values of various amplitudes. This is useful because the flavors of El Niño are tied to the amplitude of Niño-3.4 (Capotondi et al. 2015), and, further, stronger ENSO events result in enhanced predictability (Chen and Kumar 2015; Hu et al. 2019). Furthermore, providing more categories or even the entire range of all possible outcomes (via the probability density function) has shown to be useful for a wide variety of decision-makers (e.g., WMO 2020; Muñoz et al. 2019, 2020).

Despite an overall substantial improvement in NMME prediction of SST over the past 10 years, the skill of long-lead (several months into the future) prediction of SST in the ENSO regions has decreased (Becker et al. 2020). A couple of studies have noted that the NMME does not represent the observed trend in SST in the central tropical Pacific, showing more warming than has been observed, an effect that intensifies with forecast lead (Barnston et al. 2019; Shin and Huang 2019). We have updated these studies with the more recent models, all of which have 39 years of retrospective and real-time data, with observed trend shown for ERSSTv5 (Huang et al. 2017) (Fig. 2). Trend is calculated for individual leads from all initial conditions. "Forecast lead" refers to the number of months into the future from initialization; by NMME convention, this is denoted by "lead-0.5, lead-1.5," etc. (e.g., Kirtman et al. 2014). Lead-0.5, the first forecast month, is the month in which the forecasts are initialized—the January lead-0.5 forecast is for the January average, the January lead-1.5 forecast is for the February average, and so on.

We see an overly strong tropical Pacific warming trend to varying degrees, especially in the central and eastern Pacific, in the longer leads of all models. (With the exception of CFSv2, and, to a lesser extent, CCSM4, the observed trend is represented well at the 0.5-month lead.) While this overly strong warming trend may play a part in the decrease in prediction skill at longer leads in newer models, more investigation would be required to fully diagnose the causes of this decrease.

Seasonal prediction. While the NMME's initial purpose was to inform NOAA's seasonal U.S. outlooks, the global fields have been used by many researchers and prediction centers across the world. A sample of these regional predictions include precipitation and/or temperature in northeast Brazil, Iran, Israel, the Sahel, and the Indian summer monsoon (da Rocha et al. 2021; Najafi et al. 2021; Givati et al. 2017; Giannini et al. 2020; Pillai et al. 2021; respectively). Statistically downscaled NMME predictions, and predictions aggregated to watershed or basin scales, have been explored for regions of the United States by several groups, potentially providing customized information beyond that available from a national or global map (Baker et al. 2019; Barbero et al. 2017; Bolinger et al. 2017).

Several years of operation provides an opportunity to assess various strengths and weaknesses of the system, as well as its performance in real time, and to explore the NMME representation of some observed features from the past four decades. The NMME team and seasonal forecasters, in paying close attention to the NMME forecasts every month for 10 years, have observed that the real-time NMME rarely accurately predicts below-average winter temperatures in North America (CPC staff 2019, personal communication). In fact, we suspected that the real-time NMME simply rarely predicts below-average North American winter temperatures at all. In a warming world, below-average winter months are rarer than above average, but do still occur, and can be very consequential for energy use, health, and many other factors (e.g., Trenary et al. 2015; Wolter et al. 1999).

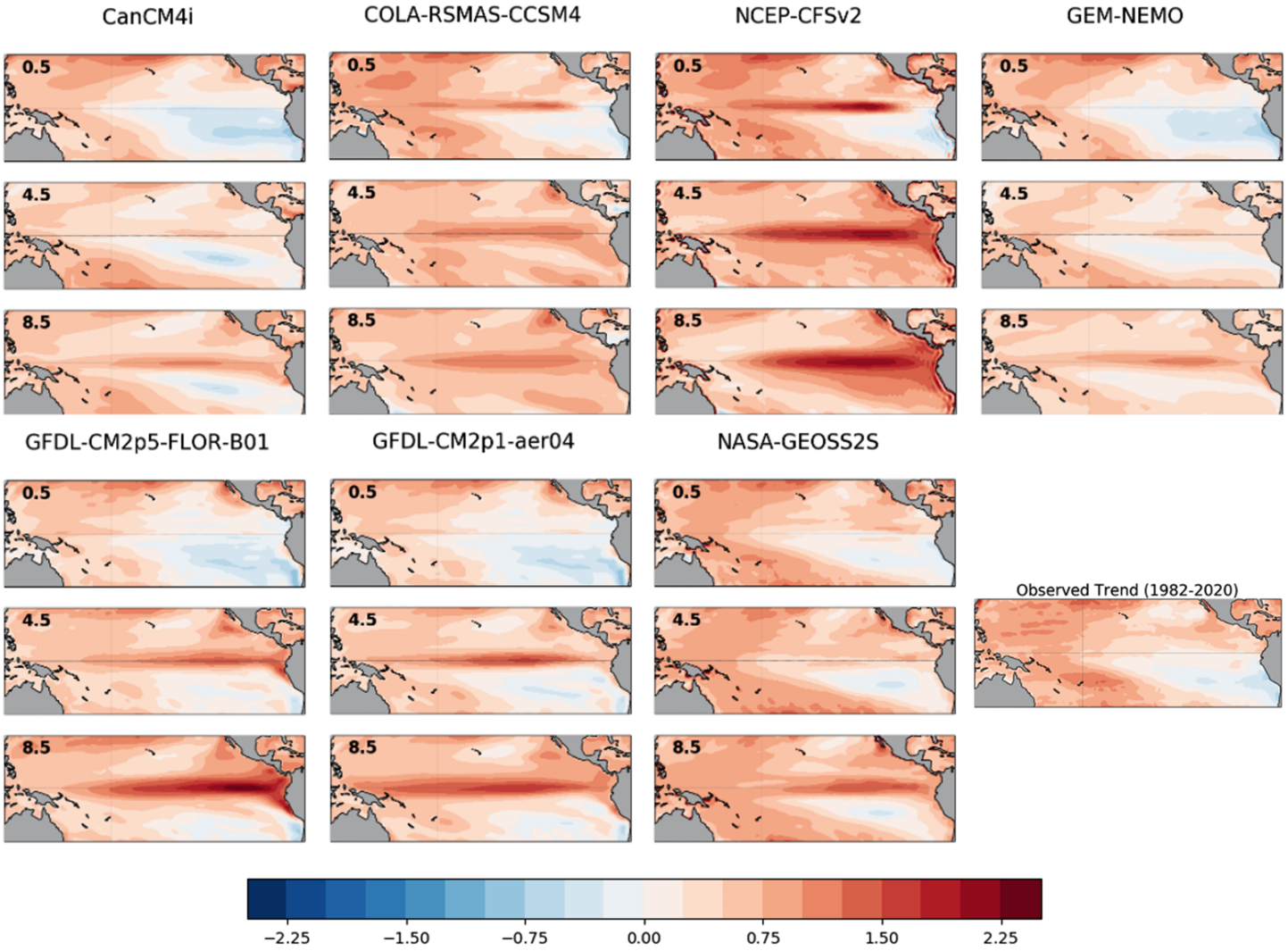

Fig. 2. The 1982–2020 trend (°C) in monthly mean sea surface temperature in the NMME models and observations. Models comprise COLA-RSMAS-CCSM4, CanCM4i, GEM-NEMO, GFDL-CM2.1, GFDL-FLORa06 and FLORb01, NASA-GEOSS2S, and NCEP-CFSv2; as the trends in GFDL-FLORa06 and FLORb01 are nearly identical, only FLORb01 is shown here. Observed trend is from ERSSTv5 data. The forecast lead is denoted in the upper-left corner of each panel: "0.5" indicates the first forecast month, "4.5" the fifth, and "8.5" the ninth.

To investigate this apparent tendency to underpredict below-average events, we examined the frequency of prediction of monthly 2-m temperature at a 1-month lead during the real-time period of 2011–20, using the multimodel ensemble mean of the eight models with complete 1982–2020 records (Figs. 3 and 4). Above average is here defined as the upper tercile: more than 0.431 standard deviation (sd) above the 1982–2010 mean. The below-average tercile is cooler than −0.431 sd. Ensemble-mean forecasts are calculated for each individual model using its own 1982–2010 climatology, thus bias corrected for systematic error in the mean and the standard deviation. "Lead-1.5" is the first month following the initial month, e.g., the February forecast made in early January. The NOAA Climate Prediction Center's monthly forecast at a 1-month lead is one of the operational outlooks that is informed by NMME output. While the instigation was temperature during winter months, we included all months in the analysis.

The GHCN+CAMS gridded observed dataset was used for comparison (Fan and van den Dool 2008). GHCN+CAMS, which combines two large individual datasets of station observations, is one of several verification records in use by NOAA's Climate Prediction Center. It has a native resolution of 0.5° longitude × 0.5° latitude and has been interpolated to match the NMME grid. The standard deviation of observed temperatures was found for each month

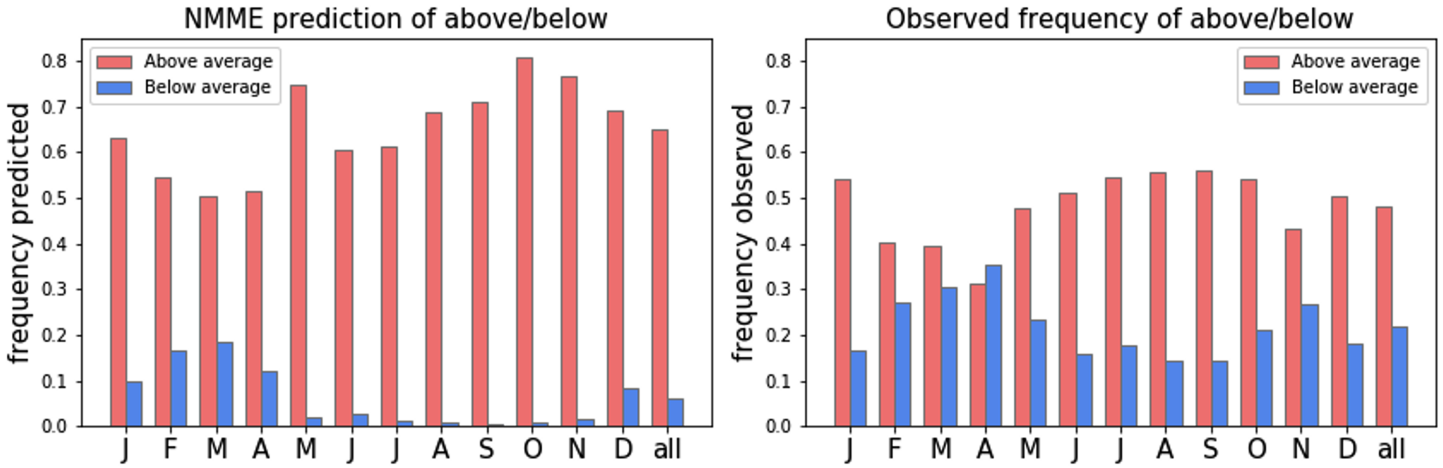

Fig. 3. (left) Frequency of prediction of above-average (upper tercile) and below-average (lower tercile) monthly mean land surface temperature anomaly in North America during the NMME real-time period of 2011–20; (right) observed frequency of the same from GHCN+CAMS. Above-average and below-average thresholds are based on 1982–2010. NMME prediction is shown for a 1.5-month lead and is the multimodel ensemble-mean anomaly of eight equally weighted models: COLA-RSMAS-CCSM4, CanCM4i, GEM-NEMO, GFDL-CM2.1, GFDL-FLORa06 and FLORb01, NASA-GEOSS2S, and NCEP-CFSv2. Each gridpoint forecast is treated as an individual forecast for the frequencies computed here.

using the 1982–2010 period, at each grid point, and used to determine the tercile thresholds as described above. The frequency of above average during the real-time forecast period of 2011–20 was then calculated as how many gridpoint observed temperatures exceeded the upper tercile threshold versus the total number of forecasts.

Counting each grid point's lead-1.5, 1-month-average, multimodel ensemble-mean forecast during 2011–21 as a single forecast, overall, 65% of forecasts predicted above-average temperature, and 5% below average. In contrast, above average was observed to have occurred 48% of the time and below average 22% during this same period (Fig. 3). Since average is defined as 1982–2010, it is expected that, in a warming world, above-average temperature would occur more often than below average during 2011–20, both in observations and predictions. However, the underprediction of below-average temperature by the NMME is found to be a systematic issue. This pattern is strongest for warm-season months, when the NMME virtually never predicts below average in the real-time period. There is some spatial similarity between the forecast and observed frequencies, with the southwest United States experiencing the highest frequency of both observed and forecast above-average temperature, but this relationship is not strong overall (Fig. 4).

A full diagnosis of the real-time overprediction of above-average 2-m temperature (t2m) is beyond the scope of this article, but one possible contributor to the forecast bias is an overly strong temperature trend in the NMME over much of the United States (Fig. 5). This tendency, especially over the eastern United States, was noted and explored by Meehl et al. (2012) using CCSM3, one of the earliest NMME models, and continues to be a difficulty for more recent models. North-central Asia is the only other region with a much stronger trend in NMME than observed; overall, NMME models tend to produce a slightly weaker global land t2m trend than observed (Krakauer 2019).

Extended applications. The free, real-time availability of global data from the NMME has led to many applications beyond seasonal temperature and precipitation prediction for North America. The NMME has become a valuable educational resource for the capacity-building efforts of NOAA's International Desks (W. Thiaw 2020, personal communication). The International Desks are a training organization with the goal of "preparing an international cadre of meteorologists who can face the challenges of a modern weather and climate forecast office" (ncep.noaa.gov/intldesk/). The International Desks hosted at the NOAA Climate Prediction

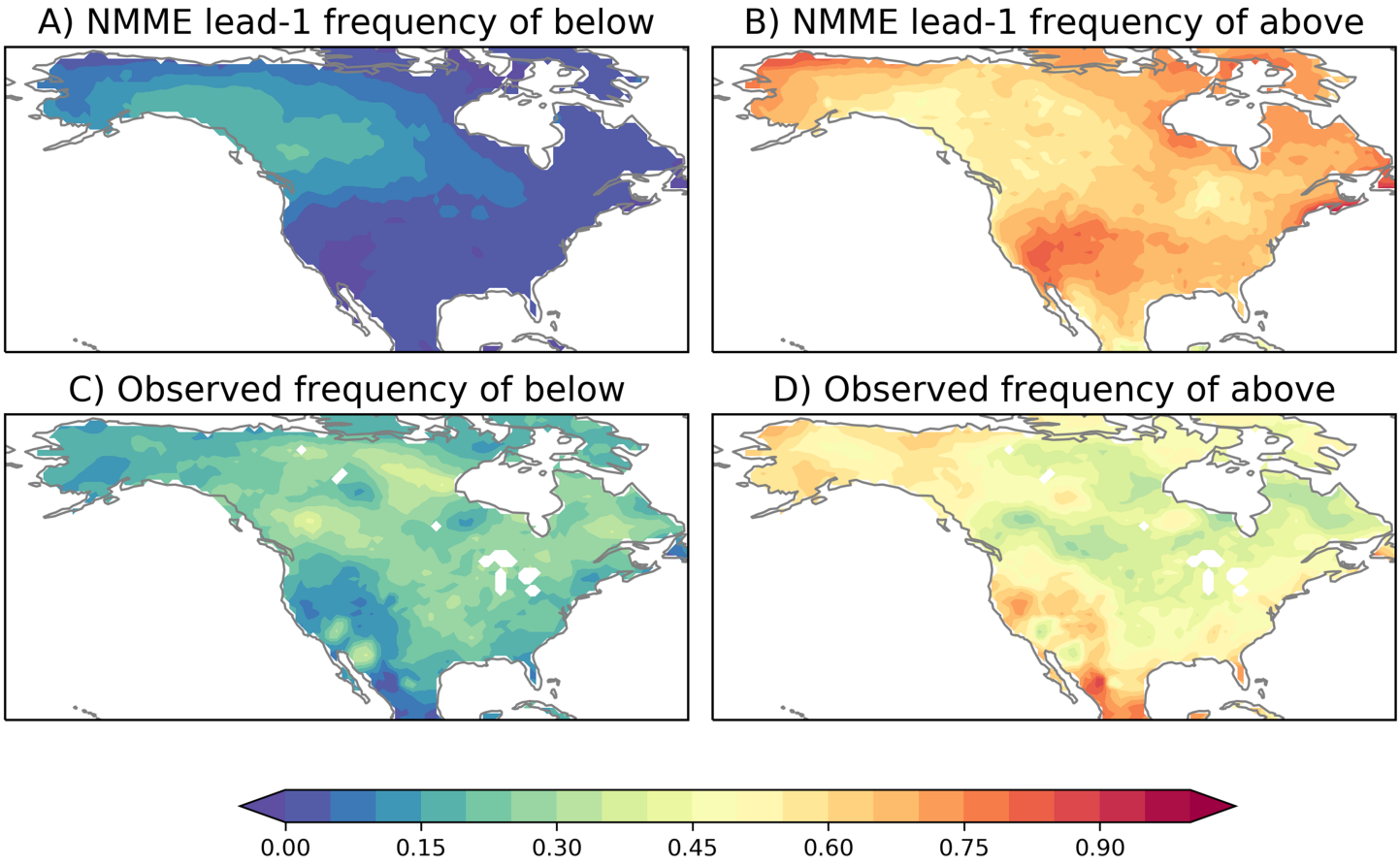

Fig. 4. Frequency of 1.5-month lead prediction of (a) below-average and (b) above-average monthly mean surface temperature anomaly by the NMME, aggregated over all 12 initial conditions, for the model suite described in Fig. 3. The (c),(d) corresponding observed frequency from GHCN+CAMS.

Center created NMME-based seasonal forecast tools for Africa (Fig. 6) and other regions of special interest (Thiaw and Kumar 2015), and publish real-time forecasts in an accessible format for use with the IRI's Climate Predictability Tool (CPT; Mason et al. 2021; see appendix). Visitors enrolled in the International Desks climate prediction training program are introduced to probabilistic ensemble prediction using the NMME.

Many teams have used NMME global monthly temperature, precipitation, and/or SST to force specialized models. Examples include prediction of streamflow in the Nile (Eldardiry and Hossain 2021); climatic suitability for invasive insects (Barker et al. 2020) and malaria in South Africa (Landman et al. 2020); mosquito-borne diseases in North America, Central America, and the Caribbean and South America (Muñoz et al. 2017, 2020; see also sidebar); coffee yield in Central America (Pons et al. 2021); hydropower planning (Koppa et al. 2019); model-analog ENSO prediction (Ding et al. 2019); and global seasonal fire activity (Turco et al. 2018).

Hybrid dynamical–statistical prediction systems combine dynamical model data with statistical relationships to predict specific quantities. An NMME-based hybrid model for the prediction of the seasonal number and accumulated cyclone energy of Atlantic hurricanes was developed by D. S. Harnos et al. (2019) and is currently in use to inform NOAA's seasonal Atlantic hurricane outlook. This system employs NMME tropical Atlantic SST and wind shear—wind shear is provided by a subset of the modeling centers to NOAA for this application. NMME models from GFDL are also used in hybrid systems to predict North Atlantic and North Pacific tropical cyclone activity (Villarini et al. 2019; Zhang and Villarini 2019).

NMME-based hybrid systems are used for several hydrological applications. Teams have designed drought prediction systems, including regional techniques for China, the United States,

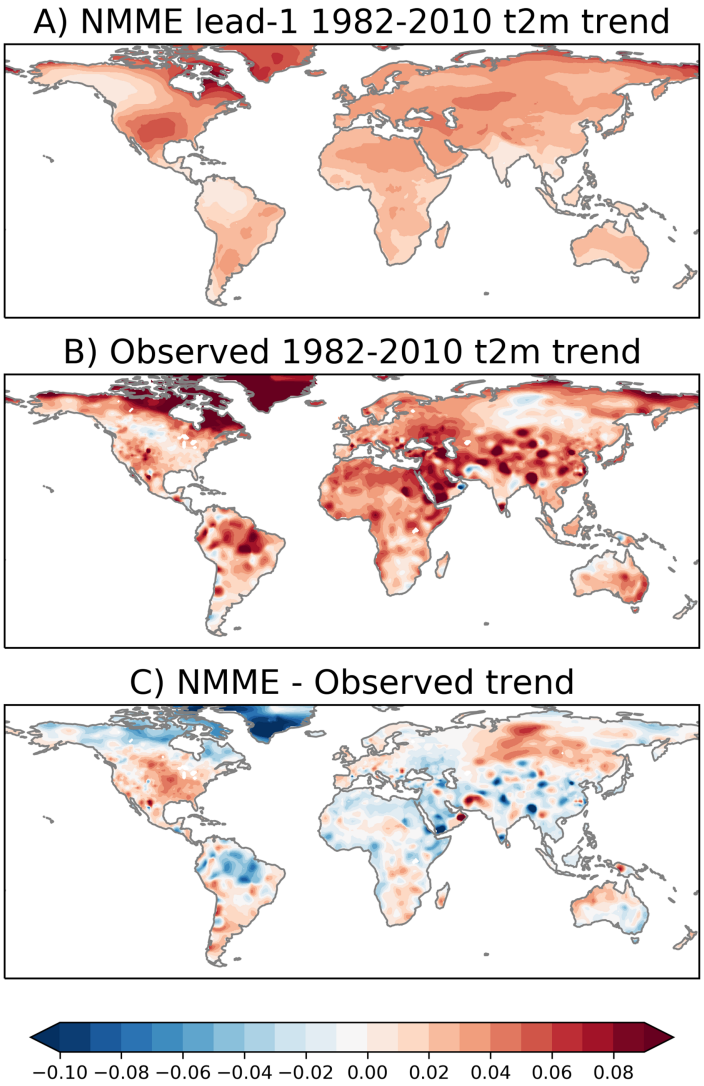

Fig. 5. The 1982–2010 linear trend (°C yr^{-1}) in 2-m temperature in the (a) NMME 1.5-month lead monthly mean and (b) GHCN+CAMS monthly mean, and (c) the difference. NMME is the multimodel mean as shown in Figs. 2 and 3.

and Europe (Ma et al. 2019; Madadgar et al. 2016; Thober et al. 2015), a global approach (Mo and Lyon 2015), and a technique for categorical drought prediction (Hao et al. 2017). Slater et al. (2019) and Slater and Villarini (2018) developed NMME-based hybrid streamflow predictions for the U.S. Midwest, and NASA's hydrological forecasting and analysis system (NHyFAS), developed to support the U.S. Agency for International Development's Famine Early Warning Systems network, employs NMME precipitation data as input (Arsenault et al. 2020) (Fig. 7).

Predictability and seasonal prediction research

Beyond prediction systems, teams using NMME data have made extensive contributions to our understanding of climate variability, predictability, and the interaction of climate variability on different time scales. New calibration and multimodel ensembling techniques provide valuable insight into model biases, potentially leading to improvements in future coupled models and, ultimately, better predictions.

Predictability, as opposed to prediction skill, is the extent to which a physical quantity in a chaotic system can be predicted, as small differences in initial states grow in time (Lorenz 1969, 1982). Predictability can be explored using ensembles, often through so-called perfect model experiments, in which a single model run is taken as the verification, and the remaining ensemble members are the prediction. Hence, the model is trying to predict an outcome

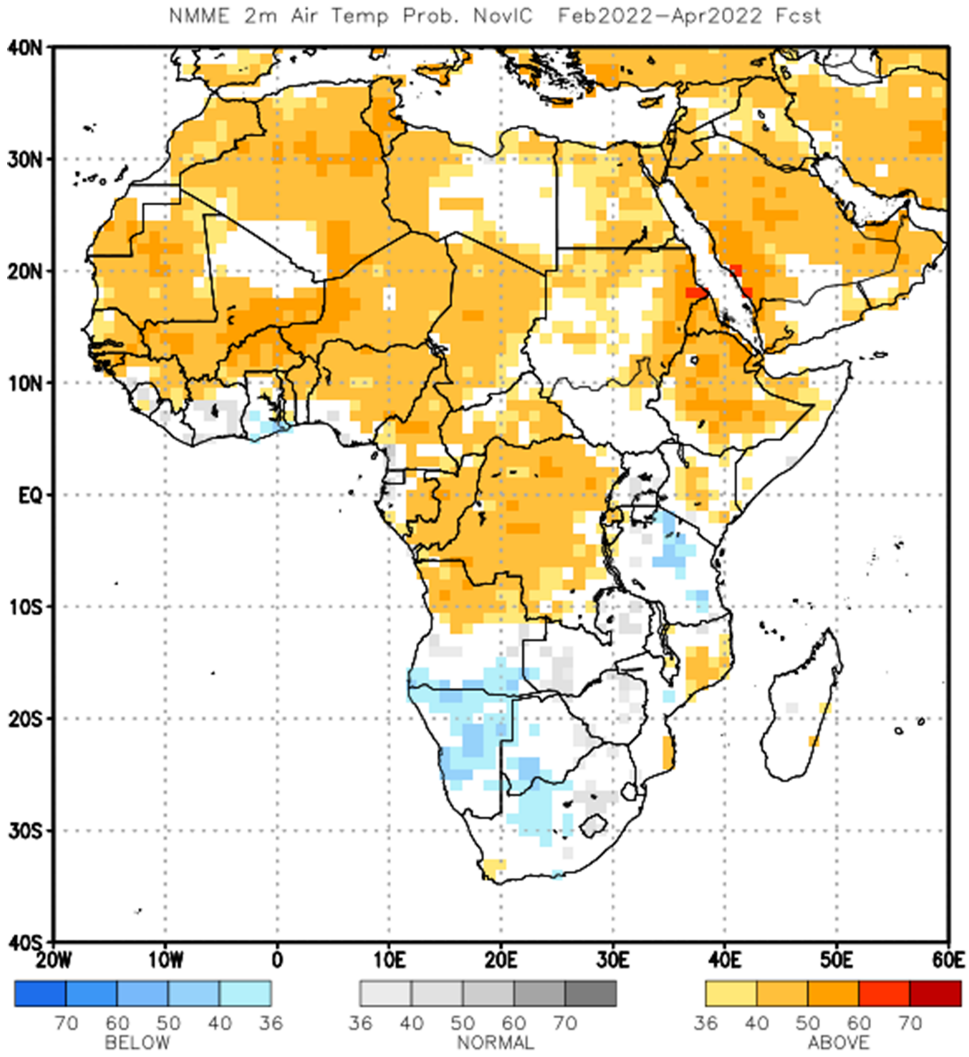

Fig. 6. Example map showing NMME probabilistic February–April-average 2-m temperature prediction for Africa from NOAA's International Desk. Colors indicate the probability of the most likely categorical outcome based on the six-model NMME from November 2021. See appendix for information on data download and other images. Details on forecast construction can be found in Becker and van den Dool (2016).

within its own world, negating the loss in skill that comes when the model's physics and assimilated data do not perfectly match the real world. NMME provides the opportunity to estimate predictability as represented in several different models and the relationship between the results, perhaps getting closer to the "true" predictability (Becker et al. 2014; Pegion et al. 2019). Newman and Sardeshmukh (2017), using NMME and the linear inverse model, find that our current forecast capabilities for tropical SST may be nearing the limits of predictability, although Pegion et al. (2019) argue that perfect model predictability estimates likely do not represent the upper limit of real-world predictability, due to model errors. A related method of estimating predictability that separates a field into its predictable and unpredictable components is employed with NMME to understand predictability of 200-hPa height (Z200) and ENSO (Jha et al. 2019; Kumar et al. 2017; respectively). Jha et al. (2019) use the NMME's larger ensemble and more advanced models to confirm the results of an earlier study (Kumar et al. 2007) that the predictable component of Z200 is highly dependent on ENSO, while Kumar et al. (2017) find ENSO predictability varies seasonally, and note that their results also vary with model.

Understanding the sources of predictability and their representation in models is another important area of study for climate prediction. Cash and Burls (2019) found that

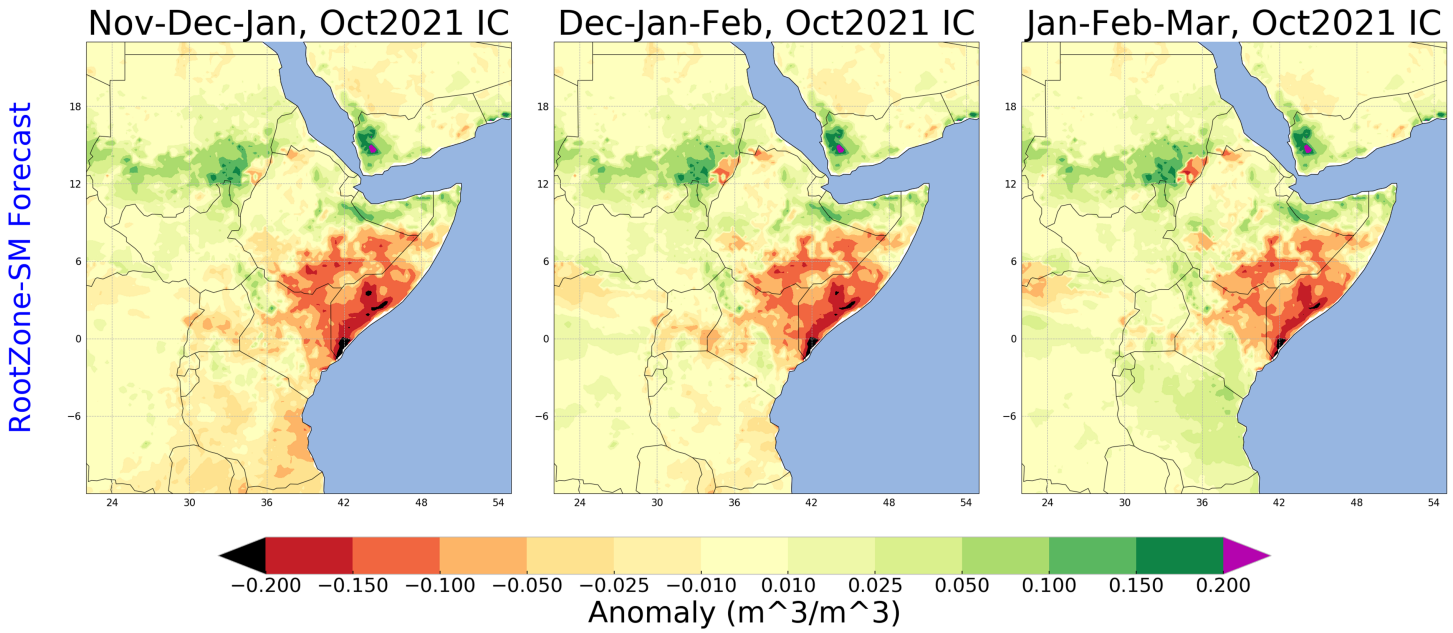

Fig. 7. Example map from NASA's NHyFAS website (https://ldas.gsfc.nasa.gov/fldas/models/forecast) showing the root zone soil moisture prediction for Africa based on October initial conditions. This hybrid system uses NMME precipitation forecast data as input. Details on NHyFAS can be found in Arsenault et al. (2020).

ENSO explained less than 30% of the variance of wintertime precipitation in California, a relationship that was reasonably well reproduced by the entire ensemble, while Kumar and Chen (2020) found a range of explained variance of about 9%–25%. The multimodel mean, however, reduces noise and substantially overestimates the ENSO–California precipitation relationship. Larson and Pegion (2020) examine the oceanic heat content along the equatorial Pacific in springtime as a source of ENSO predictability, finding that the signal of this predictor is stronger in NMME than in nature. This may be why ENSO forecasts appear overconfident when initialized in the spring, often providing a prediction based on the direction of the observed tendency (Tippett et al. 2020).

Some recent studies have focused on the sources of predictability for hydrological extremes. For example, U.S. drought prediction by the NMME has some useful skill, likely due to SST-linked predictability (Seager et al. 2020). On the other hand, U.S. flood prediction skill, which depends on low-predictability precipitation, is very limited (Neri et al. 2020). In Bangladesh, the NMME summer seasonal total rainfall actually has a stronger relationship to the observed number of dry and wet spells than to the observed seasonal total rainfall (Kelley et al. 2020).

NMME representation of persistent climate phenomena. The region of the North Atlantic Ocean to the south of Greenland has been exhibiting a surface temperature cooling trend, in stark contrast to the overall warming trend of most of the world's oceans (Hansen et al. 2010). The North Atlantic warming hole (NAWH) is associated with changes in the jet stream and circulation, with implications for global atmospheric circulation and weather (Gervais et al. 2019, 2020; Karnauskas et al. 2021; Woollings et al. 2012). The NAWH has been attributed to a combination of mechanisms, including a slowdown in the Atlantic meridional overturning circulation, changes in the subpolar gyre, cloud feedback, and teleconnections from the Indian Ocean (Caesar et al. 2018; Drijfhout et al. 2012; Hu and Fedorov 2020; Keil et al. 2020). NMME skillfully predicts SST in this region, with anomaly correlations above 0.7 even at longer leads, independent of target season (Becker et al. 2014; see also skill maps on www.cpc.ncep.noaa.gov/products/NMME/). Also, substantial skill increases have been noted with later NMME model suites (Becker et al. 2020).

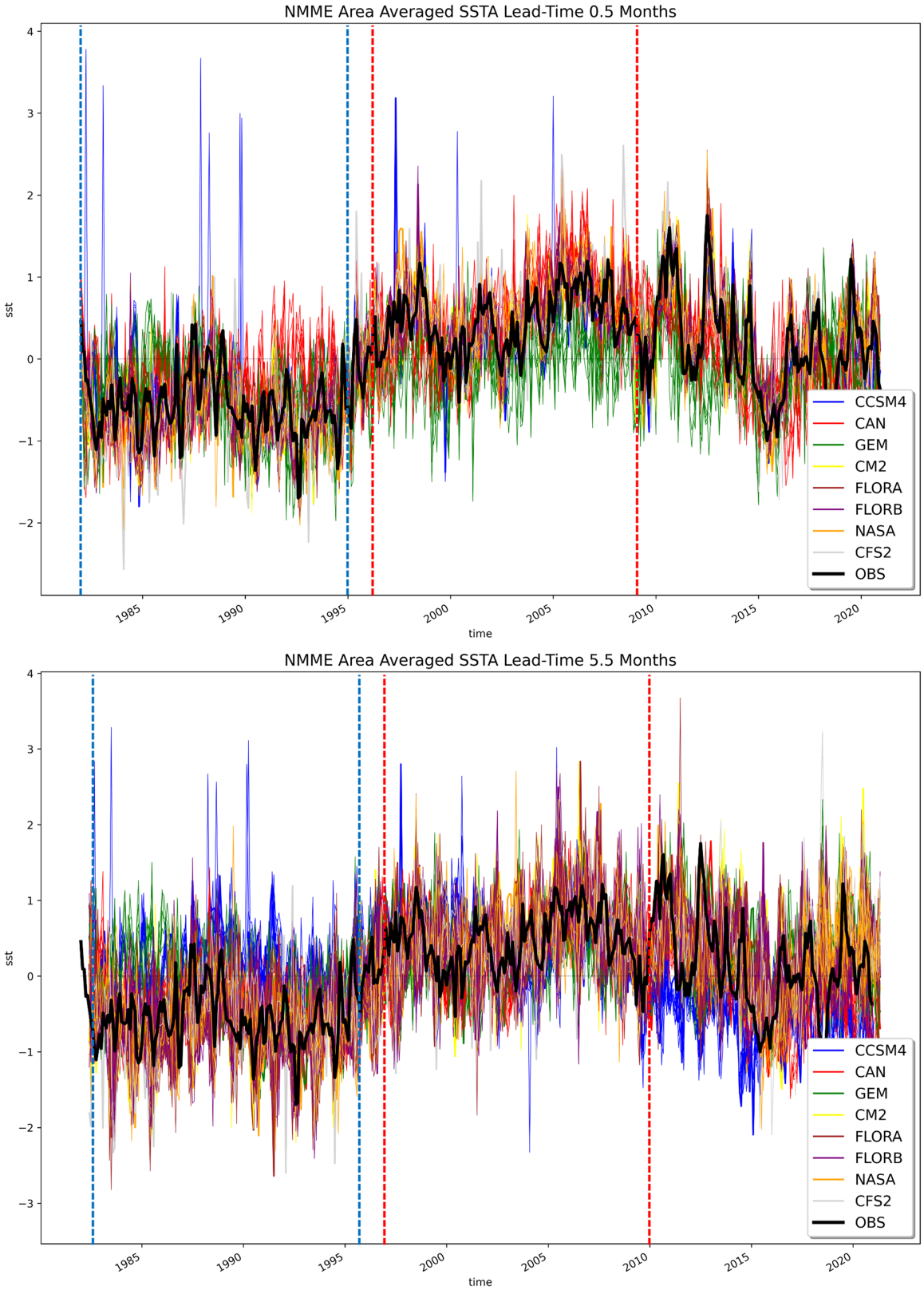

Fig. 8. Time series for area-averaged (50°–60°N, 55°–10°W; see Fig. 9) monthly mean SSTA from all model forecasts and all ensemble members, for (top) 0.5-month lead and (bottom) 5.5-month lead. The thin colored curve in the top panel indicate the forecast and the black curve is the observational estimates from OISST. Anomalies are calculated against the 1982–2020 mean; model suite comprises COLA-RSMAS-CCSM4, CanCM4i, GEM-NEMO, GFDL-CM2.1, GFDL-FLORa06 and FLORb01, NASA-GEOSS2S, and NCEP-CFSv2. Red and blue vertical dashed lines indicate averaging periods shown in Fig. 9.

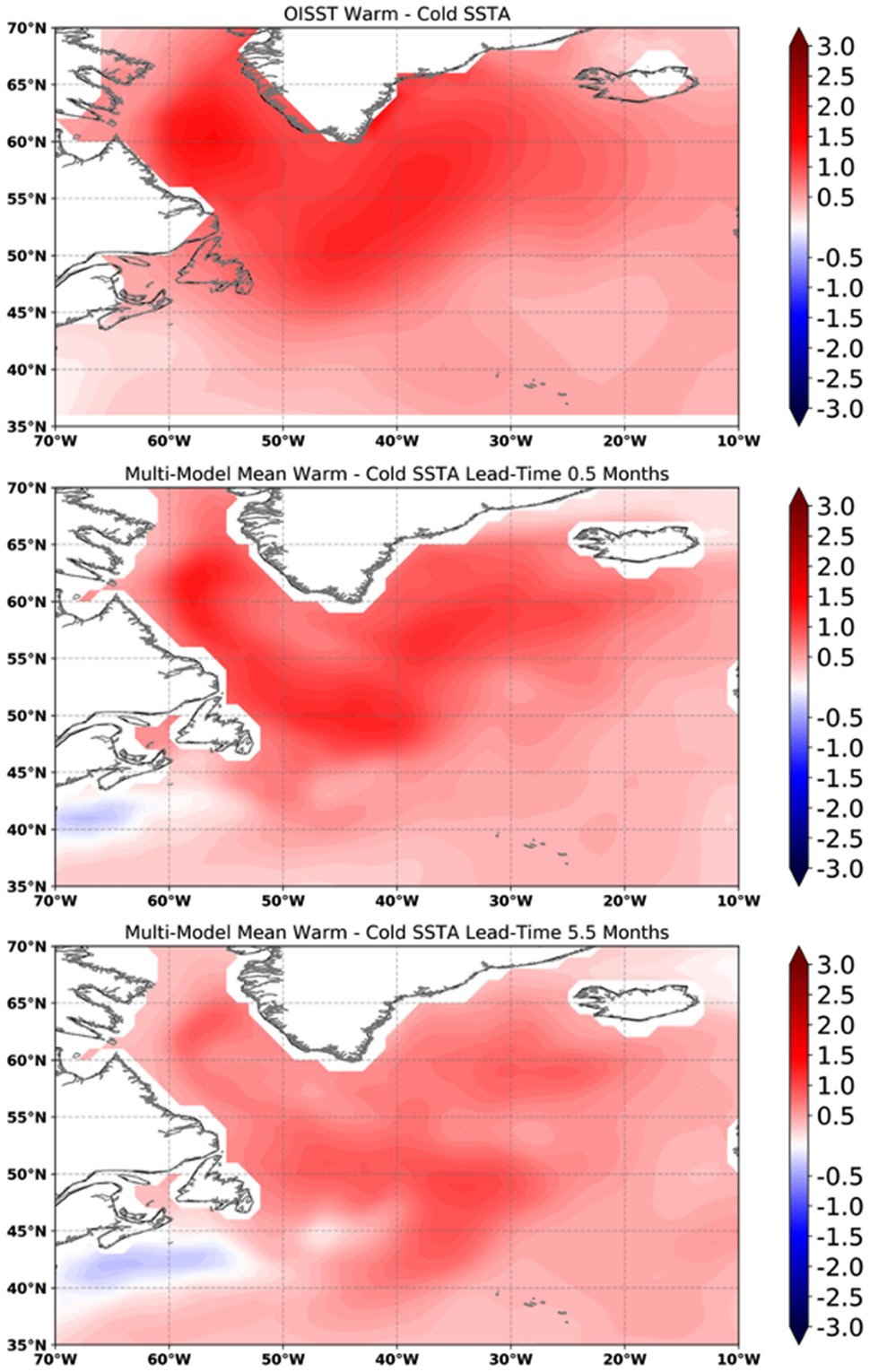

Fig. 9. (top) North Atlantic SST warm period (1996–2009) minus cold period (1982–95) OISST observational estimates, (middle) NMME 0.5-month lead monthly mean prediction, and (bottom) NMME 5.5-month lead. NMME is the multimodel ensemble mean of the eight model ensemble means shown in Fig. 8, with equal weight per model. Anomalies are based on 1982–2020.

NMME's overall successful prediction of NAWH, along with the NAWH's interaction with storm tracks and weather, invites study of more specific characteristics, including prediction of lower- and higher-frequency variability (Figs. 8, 9). Observed SSTA here is provided by NOAA's Optimum Interpolation SST version 2 (OISST; Reynolds et al. 2002). Interannual

variability is not well captured overall, with some models, notably CCSM4, worse than others, even at the shortest lead (Fig. 8). CFSv2, which uses the same initial condition data as CCSM4 (CFSR), does not exhibit the same occasional large swings in monthly SSTA. Low-frequency variability, represented here by averages of monthly mean SSTA over 1982–95 (cooler period) and 1996–2009 (warmer period), is captured reasonably well, especially at the shortest lead, but with some success still at the 5.5-month lead (Figs. 8, 9). The challenge of retaining decadal signals in the context of seasonal prediction was previously addressed by Barnston and Lyon (2016), who found that NMME captured an observed decadal shift in rainfall patterns in the U.S. Southwest and was able to maintain the decadal shifts without regressing to the overall mean in short-lead seasonal forecasts. At longer leads, the amplitude of the decadal shift was still present, but substantially reduced, similar to the behavior of the NMME's prediction of NAWH.

Bias correction and multimodel techniques. Basic systematic errors in a model's mean prediction are usually removed by subtracting the model's lead-dependent long-term mean from the forecast (Mason 2008; Becker et al. 2014). The resulting anomaly can be used as is, or the observed climatology can be added to the anomaly for a bias-corrected full field forecast. However, models may contain many biases beyond a shift in the mean climate, including inaccurate representation of the variance, skewness, and teleconnection patterns. Many studies have identified calibration and other postprocessing methods to correct for systematic errors, improve reliability, and explore optimal model consolidation.

Established bias correction methods such as canonical correlation analysis (CCA; Tippett et al. 2008), when applied to NMME rainfall forecasts in Africa, resulted in a substantial improvement in forecast skill across most of the individual models and in the multimodel mean (Thiaw and Kumar 2015). On the other hand, Barnston and Tippett (2017) found mixed results when applying CCA to 15 different regions. Ensemble regression, developed by Unger et al. (2009), was found to result in skill increases in South American precipitation forecasts (Osman et al. 2021). Chen et al. (2017) assesses NMME representation of ENSO teleconnection patterns in North America, and Strazzo et al. (2019) deploy a calibration, bridging, and merging (CBaM) method to statistically correct the teleconnection patterns to improve seasonal t2m and precipitation prediction.

Many researchers have used the NMME to explore more effective multimodel combination methods, challenging the conventional wisdom that effective weighting strategies are elusive (e.g., Hagedorn et al. 2005; Tebaldi and Knutti 2007). Bayesian updating methods are applied by Zhang et al. (2017) to NMME prediction of Niño-3.4, resulting in improved forecast skill over the equally weighted ensemble mean at almost all lead and target months, and by Slater et al. (2017) to European temperature and precipitation, resulting in a mix of improvements and loss of information. Khajehei et al. (2018) applies Bayesian postprocessing to NMME precipitation forecasts, again finding improved prediction. An optimal weighting system developed by Chen and van den Dool (2017) tests different methods of improving the performance of ridge regression, finding overall some improvement in precipitation forecasts from a non-equal-weight multimodel ensemble. Optimal NMME model weighting was also explored by Wanders and Wood (2016).

Novel bias correction and calibration methods with indications of improved prediction skill, reliability, or other assessment statistics, have been developed using NMME. These include a bias correction for precipitation based on dynamically linking it with temperature (Narapusetty et al. 2018) and experimental calibration methods applied to probabilistic ENSO prediction (Graziani et al. 2021). Van den Dool et al. (2017) discusses the probability anomaly correlation calibration technique, which is employed by the NMME probabilistic forecasts

shown on NOAA's NMME website. These NMME forecasts, and those produced through IRI's NextGen approach (see sidebar), calibrate each model's forecast independently, before equally weighting the results in the multimodel ensemble.

Emerging and future directions

Daily resolution of the long-lead NMME hindcasts (1982–2010) is available for 2-m temperature, precipitation, and SST, as well as several additional fields for a limited set of models in the NMME Phase II database hosted at NCAR (see appendix). Real-time forecast data in daily resolution for five models are available at NOAA's National Center for Environmental Information (NCEI; see appendix). To our knowledge, the real-time daily data are not widely employed by forecast centers, but several research studies using daily resolution from the Phase-II database have found potentially useful prediction capabilities, especially beyond mean temperature and precipitation. Examples include sudden stratospheric warmings, sea ice, and atmospheric rivers (Furtado et al. 2021; K. J. Harnos et al. 2019; Zhou and Kim 2018; respectively). The 10-m winds from one NMME model, CCSM4, were used to force wave models by Bell and Kirtman (2019). Carrillo et al. (2018) find promise in the prediction of the beginning of the spring season using daily maximum and minimum temperature, but only after the application of a sophisticated postprocessing. They point out that the length of training period and ensemble size are both important for developing a skillful forecast for spring onset, a common theme among studies that use the Phase II database.

With the exception of ENSO prediction, most widely used NMME prediction applications are land based. However, recent years have seen an increase in the potential for oceanic applications, especially for fisheries-relevant fields such as marine ecosystems and marine heatwaves (Hervieux et al. 2019; Jacox et al. 2019). Sea level height prediction, relevant for coastal and island flooding, is also a topic of substantial current interest. Successful model prediction for tropical Pacific islands has been demonstrated by Widlansky et al. (2017), with potential application of NMME sea level height information, were it available. Khouakhi et al. (2019) finds high sea level along the U.S. West Coast is predicted up to 6 months in advance using ENSO forecasts from NMME.

Prediction systems developed using single seasonal climate models can be expanded, and potentially improved, by employing multimodel ensembles. Some existing systems are currently investigating expanding to use NMME inputs, including the Seasonal Probabilistic Outlook for Tornadoes (SPOTter; Lee et al. 2021), which was developed using CFSv2. These systems often employ nonstandard NMME fields which are furnished to researchers by the individual model providers, similar to how the wind shear information is provided by some of the models for the NMME-based Atlantic hurricane prediction. The NMME team is currently discussing adding a limited number of variables, including some ocean and pressure-level data, to the monthly prediction to facilitate this type of project. The list of mandatory variables was increased once before, in 2013, when 200-hPa geopotential height, maximum and minimum 2-m temperature, soil moisture, and runoff were added.

The three brief investigations into NMME predictions included earlier in this paper (underprediction of below-average temperature, trends in tropical Pacific SST, and representation of the North Atlantic warming hole) all contain elements of trend analysis. The latter two illustrate that model trends can differ at longer leads from the initial state, although a thorough diagnosis of the origins of these trends is beyond the scope of this article. Some of the oldest NMME models did not capture greenhouse gas–related temperature trends at all, and prediction skill has substantially improved as these models were replaced (Becker et al. 2014, 2020). However, the impact on forecast quality from model trend biases is still unclear. While no trend-based calibration is currently applied to operational NMME predictions, Shao et al. (2021a,b), using ECMWF's SEAS5 model, propose a calibration method for

model representation of temperature trends that leads to improved forecast skill and reliability. Shao et al. (2021c) expands this methodology to trends in precipitation, finding substantial improvement to forecasts in some regions.

The discussion included here is by no means intended to be comprehensive. Rather, we hope that we have illustrated the variety of questions that could be explored in the NMME and other seasonal forecast systems. A clearer understanding of the strengths and weaknesses of the system could potentially contribute to forecasts of opportunity, periods of increased forecast confidence (e.g., Mariotti et al. 2020). Also, while the substantial number of bias correction, calibration, and weighting techniques developed using NMME hindcasts are an important contribution to understanding model biases—and potentially correcting them—it is unclear to what extent these methods can improve truly independent forecasts, that is, real-time forecasts for which we do not know the outcome ahead of time. The majority of them have been tested on the same sample used to develop them, a necessity given the limitations of climate prediction time scales, and even the most careful cross validation cannot create a truly independent forecast set over which to assess skill impacts. This concern was identified decades ago (e.g., Barnston et al. 1994) and the value of such hindcast skill assessments has recently been called into question by Risbey et al. (2021). An alternative is the use of retroactive approaches (e.g., Mason and Stephenson 2008), which mimic the "real-time setting" mentioned by Barnston et al. (1994) but require long time records. As the NMME continues, the database of independent forecasts grows, potentially providing an opportunity to assess the true impact of postprocessing methods on forecast skill.

Finally, new modeling and computational techniques are emerging that will be relevant for the NMME in its second decade. Higher resolution models have been tested in the NMME protocol, with some promising results using a 0.5° atmosphere and 0.1° ocean suggesting improved prediction of precipitation (Infanti and Kirtman 2019; Siqueira et al. 2021). Some recent studies have explored machine learning applications of NMME, including convolutional Gaussian process (Wang et al. 2021), neural networks (Pakdaman et al. 2020), and wavelet methods (Xu et al. 2019). As the number of recently published studies cited in this paper illustrates, NMME research and development is only accelerating, and we look forward to new applications and new discoveries in the second decade.

Acknowledgments. The NMME project and data dissemination is supported by NOAA, NSF, NASA, and DOE. The authors appreciate the extensive work of NOAA, IRI, and NCAR personnel in creating, updating, and maintaining the NMME archive, and the many NMME team members at modeling centers who develop and maintain the models and have delivered their forecasts on time every month for more than a decade. Á. G. M. was partially supported by the NOAA Grants NA18OAR4310275, NA18OAR4310339, the FORMAS Arbo-Prevent Project, and the Columbia World Project "ACToday."

Data availability statement. All NMME and verification data for this manuscript can be freely obtained from the database hosted at IRI (http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/).

Appendix: NMME data access

NMME monthly mean data resources are available at the Climate Prediction Center:

- NMME real-time monthly and seasonal forecast images for temperature, precipitation, and sea surface temperature: www.cpc.ncep.noaa.gov/products/NMME/
- Real-time forecast anomaly and probability data in netCDF format: ftp://ftp.cpc.ncep.noaa.gov/NMME/
- Regionalized NMME forecasts images, archive, and data downloads, including binary and CPT-compatible (text) formats: www.cpc.ncep.noaa.gov/products/international/nmme/nmme.shtml

NMME monthly mean data resources are available at the IRI:

- Archived hindcast and real-time data in netCDF format at the IRI Data Library: http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/. Includes years available, links to model documentation, spatial and temporal subsetting capability, and viewing options.
- The Climate Predictability Tool: a software package for constructing a seasonal climate forecast model, performing model validation, and producing forecasts given updated data. Provides seamless access to NMME: https://iri.columbia.edu/our-expertise/climate/tools/cpt/

NMME monthly and daily data available for an expanded set of variables (1982–2010 only) are available at the NCAR Climate Data Gateway:

- www.earthsystemgrid.org/search.html?Project5NMME

NMME daily data (2011–present only) are available at NOAA National Centers for Environmental Information (NCEI) for CCSM4, CFSv2, CanCM3/4, GEOS-5, and GEM-NEMO (2019–present only):

- www.ncei.noaa.gov/products/weather-climate-models/north-american-multi-model

References

An, S.-I., E. Tziperman, Y. M. Okumura, and T. Li, 2020: ENSO irregularity and asymmetry. *El Niño Southern Oscillation in a Changing Climate*, *Geophys. Monogr.*, Vol. 253, Amer. Geophys. Union, 153–172, https://doi.org/10.1002/9781119548164.ch7.

Arsenault, K. R., and Coauthors, 2020: The NASA hydrological forecast system for food and water security applications. *Bull. Amer. Meteor. Soc.*, **101**, E1007–E1025, https://doi.org/10.1175/BAMS-D-18-0264.1.

Baker, S. A., A. W. Wood, and B. Rajagopalan, 2019: Developing subseasonal to seasonal climate forecast products for hydrology and water management. *J. Amer. Water Resour. Assoc.*, **55**, 1024–1037, https://doi.org/10.1111/1752-1688.12746.

Barbero, R., J. T. Abatzoglou, and K. C. Hegewisch, 2017: Evaluation of statistical downscaling of North American Multimodel Ensemble forecasts over the western United States. *Wea. Forecasting*, **32**, 327–341, https://doi.org/10.1175/WAF-D-16-0117.1.

Barker, B. S., L. Coop, T. Wepprich, F. Grevstad, and G. Cook, 2020: DDRP: Real-time phenology and climatic suitability modeling of invasive insects. *PLOS ONE*, **15**, e0244005, https://doi.org/10.1371/journal.pone.0244005.

Barnston, A. G., and B. Lyon, 2016: Does the NMME capture a recent decadal shift toward increasing drought occurrence in the southwestern United States? *J. Climate*, **29**, 561–581, https://doi.org/10.1175/JCLI-D-15-0311.1.

——, and M. K. Tippett, 2017: Do statistical pattern corrections improve seasonal climate predictions in the North American Multimodel Ensemble models? *J. Climate*, **30**, 8335–8355, https://doi.org/10.1175/JCLI-D-17-0054.1.

——, and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.*, **75**, 2097–2114, https://doi.org/10.1175/1520-0477(1994)075<2097:LLSFDW>2.0.CO;2.

——, M. K. Tippett, M. Ranganathan, and M. L. L'Heureux, 2019: Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dyn.*, **53**, 7215–7234, https://doi.org/10.1007/s00382-017-3603-3.

Becker, E., and H. van den Dool, 2016: Probabilistic seasonal forecasts in the North American Multimodel Ensemble: A baseline skill assessment. *J. Climate*, **29**, 3015–3026, https://doi.org/10.1175/JCLI-D-14-00862.1.

——, ——, 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.

——, B. P. Kirtman, and K. Pegion, 2020: Evolution of the North American Multi-Model Ensemble. *Geophys. Res. Lett.*, **47**, e2020GL087408, https://doi.org/10.1029/2020GL087408.

Bell, R., and B. Kirtman, 2019: Seasonal forecasting of wind and waves in the North Atlantic using a grand multimodel ensemble. *Wea. Forecasting*, **34**, 31–59, https://doi.org/10.1175/WAF-D-18-0099.1.

Blumenthal, M. B., M. Bell, J. del Corral, R. Cousin, and I. Khomyakov, 2014: IRI data library: Enhancing accessibility of climate knowledge. *Earth Perspect.*, **1**, 19, https://doi.org/10.1186/2194-6434-1-19.

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.

Borovikov, A., R. Cullather, R. Kovach, J. Marshak, G. Vernieres, Y. Vikhliaev, B. Zhao, and Z. Li, 2019: GEOS-5 seasonal forecast system. *Climate Dyn.*, **53**, 7335–7361, https://doi.org/10.1007/s00382-017-3835-2.

Caesar, L., S. Rahmstorf, A. Robinson, G. Feulner, and V. Saba, 2018: Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, **556**, 191–196, https://doi.org/10.1038/s41586-018-0006-5.

Capotondi, A., and Coauthors, 2015: Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, **96**, 921–938, https://doi.org/10.1175/BAMS-D-13-00117.1.

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.

Cash, B. A., and N. J. Burls, 2019: Predictable and unpredictable aspects of U.S. West Coast rainfall and El Niño: Understanding the 2015/16 event. *J. Climate*, **32**, 2843–2868, https://doi.org/10.1175/JCLI-D-18-0181.1.

Chen, L.-C. G., and H. van den Dool, 2017: Combination of multimodel probabilistic forecasts using an optimal weighting system. *Wea. Forecasting*, **32**, 1967–1987, https://doi.org/10.1175/WAF-D-17-0074.1.

——, ——, E. Becker, and Q. Zhang, 2017: ENSO precipitation and temperature forecasts in the North American Multimodel Ensemble: Composite analysis and validation. *J. Climate*, **30**, 1103–1125, https://doi.org/10.1175/JCLI-D-15-0903.1.

Chen, M., and A. Kumar, 2015: Influence of ENSO SSTs on the spread of the probability density function for precipitation and land surface temperature. *Climate Dyn.*, **45**, 965–974, https://doi.org/10.1007/s00382-014-2336-9.

Choi, K.-Y., G. A. Vecchi, and A. T. Wittenberg, 2013: ENSO transition, duration, and amplitude asymmetries: Role of the nonlinear wind stress coupling in a conceptual model. *J. Climate*, **26**, 9462–9476, https://doi.org/10.1175/JCLI-D-13-00045.1.

Cohen, J., D. Coumou, J. Hwang, L. Mackey, P. Orenstein, S. Totz, and E. Tziperman, 2019: S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdiscip. Rev.: Climate Change*, **10**, e00567, https://doi.org/10.1002/wcc.567.

da Rocha, R. L., Jr., and Coauthors, 2021: An empirical seasonal rainfall forecasting model for the northeast region of Brazil. *Water*, **13**, 1613, https://doi.org/10.3390/w13121613.

Delworth, T. L., A. J. Broccoli, A. Rosati, R. J. Stouffer, V. Balaji, J. A. Beesley, W. F. Cooke, and K. W. Dixon, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, https://doi.org/10.1175/JCLI3629.1.

——, and Coauthors, 2020: SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001895, https://doi.org/10.1029/2019MS001895.

DeWitt, D. G., 2005: Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere–ocean general circulation model. *Mon. Wea. Rev.*, **133**, 2972–2995, https://doi.org/10.1175/MWR3016.1.

Dias, D. F., A. Subramanian, L. Zanna, and A. J. Miller, 2019: Remote and local influences in forecasting Pacific SST: A linear inverse model and a multimodel ensemble study. *Climate Dyn.*, **52**, 3183–3201, https://doi.org/10.1007/s00382-018-4323-z.

Ding, H., M. Newman, M. A. Alexander, and A. T. Wittenberg, 2019: Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophys. Res. Lett.*, **46**, 1721–1730, https://doi.org/10.1029/2018GL080598.

Doblas-Reyes, F. J., M. Déqué, and J.-P. Piedelievre, 2000: Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.*, **126**, 2069–2087, https://doi.org/10.1002/qj.49712656705.

——, R. Hagedorn, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus*, **57A**, 234–252, https://doi.org/10.3402/tellusa.v57i3.14658.

Drijfhout, S., G. J. van Oldenborgh, and A. Cimatoribus, 2012: Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns? *J. Climate*, **25**, 8373–8379, https://doi.org/10.1175/JCLI-D-12-00490.1.

Eldardiry, H., and F. Hossain, 2021: The value of long-term streamflow forecasts in adaptive reservoir operation: The case of the high Aswan Dam in the transboundary Nile River basin. *J. Hydrometeor.*, **22**, 1099–1115, https://doi.org/10.1175/JHM-D-20-0241.1.

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

Furtado, J. C., J. Cohen, E. J. Becker, and D. C. Collins, 2021: Evaluating the relationship between sudden stratospheric warmings and tropospheric weather regimes in the NMME phase-2 models. *Climate Dyn.*, **56**, 2321–2338, https://doi.org/10.1007/s00382-020-05591-x.

Gervais, M., J. Shaman, and Y. Kushnir, 2019: Impacts of the North Atlantic warming hole in future climate projections: Mean atmospheric circulation and the North Atlantic jet. *J. Climate*, **32**, 2673–2689, https://doi.org/10.1175/JCLI-D-18-0647.1.

——, ——, and ——, 2020: Impact of the North Atlantic warming hole on sensible weather. *J. Climate*, **33**, 4255–4271, https://doi.org/10.1175/JCLI-D-19-0636.1.

Giannini, A., A. Ali, C. P. Kelley, B. L. Lamptey, B. Minoungou, and O. Ndiaye, 2020: Advances in the lead time of Sahel rainfall prediction with the North American Multimodel Ensemble. *Geophys. Res. Lett.*, **47**, e2020GL087341, https://doi.org/10.1029/2020GL087341.

Givati, A., M. Housh, Y. Levi, D. Paz, I. Carmona, and E. Becker, 2017: The advantage of using International Multimodel Ensemble for seasonal precipitation forecast over Israel. *Adv. Meteor.*, **2017**, e9204081, https://doi.org/10.1155/2017/9204081.

Graziani, C., R. Rosner, J. M. Adams, and R. L. Machete, 2021: Probabilistic recalibration of forecasts. *Int. J. Forecasting*, **37**, 1–27, https://doi.org/10.1016/j.ijforecast.2019.04.019.

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/10.3402/tellusa.v57i3.14657.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, https://doi.org/10.1029/2010RG000345.

Hao, Z., Y. Xia, L. Luo, V. P. Singh, W. Ouyang, and F. Hao, 2017: Toward a categorical drought prediction system based on U.S. Drought Monitor (USDM) and climate forecast. *J. Hydrol.*, **551**, 300–305, https://doi.org/10.1016/j.jhydrol.2017.06.005.

Harnos, D. S., J.-K. E. Schemm, H. Wang, and C. A. Finan, 2019: NMME-based hybrid prediction of Atlantic hurricane season activity. *Climate Dyn.*, **53**, 7267–7285, https://doi.org/10.1007/s00382-017-3891-7.

Harnos, K. J., M. L'Heureux, Q. Ding, and Q. Zhang, 2019: Skill of seasonal Arctic sea ice extent predictions using the North American Multimodel Ensemble. *J. Climate*, **32**, 623–638, https://doi.org/10.1175/JCLI-D-17-0766.1.

Hervieux, G., M. A. Alexander, C. A. Stock, M. G. Jacox, K. Pegion, E. Becker, F. Castruccio, and D. Tommasi, 2019: More reliable coastal SST forecasts from the North American Multimodel Ensemble. *Climate Dyn.*, **53**, 7153–7168, https://doi.org/10.1007/s00382-017-3652-7.

Hu, S., and A. V. Fedorov, 2020: Indian Ocean warming as a driver of the North Atlantic warming hole. *Nat. Commun.*, **11**, 4785, https://doi.org/10.1038/s41467-020-18522-5.

Hu, Z., A. Kumar, J. Zhu, P. Peng, and B. Huang, 2019: On the challenge for ENSO cycle prediction: An example from NCEP Climate Forecast System, version 2. *J. Climate*, **32**, 183–194, https://doi.org/10.1175/JCLI-D-18-0285.1.

Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Infanti, J. M., and B. P. Kirtman, 2016: North American rainfall and temperature prediction response to the diversity of ENSO. *Climate Dyn.*, **46**, 3007–3023, https://doi.org/10.1007/s00382-015-2749-0.

——, and ——, 2019: A comparison of CCSM4 high-resolution and low-resolution predictions for south Florida and southeast United States drought. *Climate Dyn.*, **52**, 6877–6892, https://doi.org/10.1007/s00382-018-4553-0.

Jacox, M. G., D. Tommasi, M. A. Alexander, G. Hervieux, and C. A. Stock, 2019: Predicting the evolution of the 2014–2016 California Current system marine heatwave from an ensemble of coupled global climate forecasts. *Front. Mar. Sci.*, **6**, 497, https://doi.org/10.3389/fmars.2019.00497.

Jha, B., A. Kumar, and Z.-Z. Hu, 2019: An update on the estimate of predictability of seasonal mean atmospheric variability using North American Multi-Model Ensemble. *Climate Dyn.*, **53**, 7397–7409, https://doi.org/10.1007/s00382-016-3217-1.

Karnauskas, K. B., L. Zhang, and D. J. Amaya, 2021: The atmospheric response to North Atlantic SST trends, 1870–2019. *Geophys. Res. Lett.*, **48**, e2020GL090677, https://doi.org/10.1029/2020GL090677.

Keil, P., T. Mauritsen, J. Jungclaus, C. Hedemann, D. Olonscheck, and R. Ghosh, 2020: Multiple drivers of the North Atlantic warming hole. *Nat. Climate Change*, **10**, 667–671, https://doi.org/10.1038/s41558-020-0819-8.

Kelley, C., N. Acharya, C. Montes, T. J. Krupnik, and M. Mannan, 2020: Exploring the predictability of within-season rainfall statistics of the Bangladesh monsoon using North American Multimodel Ensemble outputs. *Theor. Appl. Climatol.*, **141**, 495–508, https://doi.org/10.1007/s00704-020-03202-7.

Khajehei, S., A. Ahmadalipour, and H. Moradkhani, 2018: An effective post-processing of the North American Multi-Model Ensemble (NMME) precipitation forecasts over the continental US. *Climate Dyn.*, **51**, 457–472, https://doi.org/10.1007/s00382-017-3934-0.

Khouakhi, A., G. Villarini, W. Zhang, and L. J. Slater, 2019: Seasonal predictability of high sea level frequency using ENSO patterns along the U.S. West Coast. *Adv. Water Resour.*, **131**, 103377, https://doi.org/10.1016/j.advwatres.2019.07.007.

Kirtman, B. P., and D. Min, 2009: Multimodel ensemble ENSO prediction with CCSM and CFS. *Mon. Wea. Rev.*, **137**, 2908–2930, https://doi.org/10.1175/2009MWR2672.1.

——, 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.

Koppa, A., M. Gebremichael, R. C. Zambon, W. W.-G. Yeh, and T. M. Hopson, 2019: Seasonal hydropower planning for data-scarce regions using multimodel ensemble forecasts, remote sensing data, and stochastic programming. *Water Resour. Res.*, **55**, 8583–8607, https://doi.org/10.1029/2019WR025228.

Krakauer, N. Y., 2019: Temperature trends and prediction skill in NMME seasonal forecasts. *Climate Dyn.*, **53**, 7201–7213, https://doi.org/10.1007/s00382-017-3657-2.

Kumar, A., and M. Chen, 2020: Understanding skill of seasonal mean precipitation prediction over California during boreal winter and role of predictability limits. *J. Climate*, **33**, 6141–6163, https://doi.org/10.1175/JCLI-D-19-0275.1.

——, B. Jha, Q. Zhang, and L. Bounoua, 2007: A new methodology for estimating the unpredictable component of seasonal atmospheric variability. *J. Climate*, **20**, 3888–3901, https://doi.org/10.1175/JCLI4216.1.

——, Z.-Z. Hu, B. Jha, and P. Peng, 2017: Estimating ENSO predictability based on multi-model hindcasts. *Climate Dyn.*, **48**, 39–51, https://doi.org/10.1007/s00382-016-3060-4.

Landman, W. A., N. Sweijd, N. Masedi, and N. Minakawa, 2020: The development and prudent application of climate-based forecasts of seasonal malaria in the Limpopo province in South Africa. *Environ. Dev.*, **35**, 100522, https://doi.org/10.1016/j.envdev.2020.100522.

Larson, S. M., and B. P. Kirtman, 2017: Drivers of coupled model ENSO error dynamics and the spring predictability barrier. *Climate Dyn.*, **48**, 3631–3644, https://doi.org/10.1007/s00382-016-3290-5.

——, and K. Pegion, 2020: Do asymmetries in ENSO predictability arise from different recharged states? *Climate Dyn.*, **54**, 1507–1522, https://doi.org/10.1007/s00382-019-05069-5.

Lee, S.-K., H. Lopez, D. Kim, A. T. Wittenberg, and A. Kumar, 2021: A Seasonal Probabilistic Outlook for Tornadoes (SPOTter) in the contiguous United States based on the leading patterns of large-scale atmospheric anomalies. *Mon. Wea. Rev.*, **149**, 901–919, https://doi.org/10.1175/MWR-D-20-0223.1.

L'Heureux, M. L., M. K. Tippett, A. Kumar, A. H. Butler, L. M. Ciasto, Q. Ding, K. J. Harnos, and N. C. Johnson, 2017: Strong relations between ENSO and the Arctic Oscillation in the North American Multimodel Ensemble. *Geophys. Res. Lett.*, **44**, 11 654–11 662, https://doi.org/10.1002/2017GL074854.

——, and Coauthors, 2019: Strength outlooks for the El Niño–Southern Oscillation. *Wea. Forecasting*, **34**, 165–175, https://doi.org/10.1175/WAF-D-18-0126.1.

——, A. F. Z. Levine, M. Newman, C. Ganter, J.-J. Luo, M. K. Tippett, and T. N. Stockdale, 2020: ENSO prediction. *El Niño Southern Oscillation in a Changing Climate*, *Geophys. Monogr.*, Vol. 253, Amer. Geophys. Union, 227–246, https://doi.org/10.1002/9781119548164.ch10.

Lin, H., and Coauthors, 2020: The Canadian Seasonal to Interannual Prediction System version 2 (CanSIPSv2). *Wea. Forecasting*, **35**, 1317–1343, https://doi.org/10.1175/WAF-D-19-0259.1.

Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc.*, **50**, 345–349, https://doi.org/10.1175/1520-0477-50.5.345.

——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513, https://doi.org/10.3402/tellusa.v34i6.10836.

Ma, F., A. Ye, and Q. Duan, 2019: Seasonal drought ensemble predictions based on multiple climate models in the upper Han River basin, China. *Climate Dyn.*, **53**, 7447–7460, https://doi.org/10.1007/s00382-017-3577-1.

Madadgar, S., A. AghaKouchak, S. Shukla, A. W. Wood, L. Cheng, K.-L. Hsu, and M. Svoboda, 2016: A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States. *Water Resour. Res.*, **52**, 5095–5110, https://doi.org/10.1002/2015WR018547.

Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101**, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1.

Mason, S. J., 2008: From dynamical model predictions to seasonal climate forecasts. *Seasonal Climate: Forecasting and Managing Risk*, A. Troccoli et al., Eds., 205–234, Springer, https://doi.org/10.1007/978-1-4020-6992-5_8.

——, and D. B. Stephenson, 2008: How do we know whether seasonal climate forecasts are any good? *Seasonal Climate: Forecasting and Managing Risk*, A. Troccoli et al., Eds., Springer, 259–289.

——, M. K. Tippet, L. Song, and Á. G. Muñoz, 2021: Climate Predictability Tool, version 17.5.2. Columbia University Academic Commons, https://doi.org/10.7916/d8-em2t-k781.

Meehl, G. A., J. M. Arblaster, and G. Branstator, 2012: Mechanisms contributing to the warming hole and the consequent U.S. east–west differential of heat extremes. *J. Climate*, **25**, 6394–6408, https://doi.org/10.1175/JCLI-D-11-00655.1.

Merryfield, W. J., and Coauthors, 2013: The Canadian Seasonal to Interannual Prediction System. Part I: Models and initialization. *Mon. Wea. Rev.*, **141**, 2910–2945, https://doi.org/10.1175/MWR-D-12-00216.1.

Mo, K. C., and B. Lyon, 2015: Global meteorological drought prediction using the North American Multi-Model Ensemble. *J. Hydrometeor.*, **16**, 1409–1424, https://doi.org/10.1175/JHM-D-14-0192.1.

Molod, A., and Coauthors, 2020: GEOS-S2S version 2: The GMAO high-resolution coupled model and assimilation system for seasonal prediction. *J. Geophys. Res. Atmos.*, **125**, e2019JD031767, https://doi.org/10.1029/2019JD031767.

Muñoz, Á. G., M. C. Thomson, A. M. Stewart-Ibarra, G. A. Vecchi, X. Chourio, P. Najera, Z. Moran, and X. Yang, 2017: Could the recent Zika epidemic have been predicted? *Front. Microbiol.*, **8**, 1291, https://doi.org/10.3389/fmicb.2017.01291.

——, and Coauthors, 2019: NextGen: A next-generation system for calibrating, ensembling and verifying regional seasonal and subseasonal forecasts. *2019 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract A23U-3024.

——, X. Chourio, A. Rivière-Cinnamond, M. A. Diuk-Wasser, P. A. Kache, E. A. Mordecai, L. Harrington, and M. C. Thomson, 2020: AeDES: A next-generation monitoring and forecasting system for environmental suitability of *Aedes*-borne disease transmission. *Sci. Rep.*, **10**, 12640, https://doi.org/10.1038/s41598-020-69625-4.

Najafi, H., A. W. Robertson, A. R. Massah Bavani, P. Irannejad, N. Wanders, and E. F. Wood, 2021: Improved multi-model ensemble forecasts of Iran's precipitation and temperature using a hybrid dynamical-statistical approach during fall and winter seasons. *Int. J. Climatol.*, **41**, 5698–5725, https://doi.org/10.1002/joc.7148.

Narapusetty, B., D. C. Collins, R. Murtugudde, J. Gottschalck, and C. Peters-Lidard, 2018: Bias correction to improve the skill of summer precipitation forecasts over the contiguous United States by the North American Multi-Model Ensemble system. *Atmos. Sci. Lett.*, **19**, e818, https://doi.org/10.1002/asl.818.

National Research Council, 2010: *Assessment of Intraseasonal to Interannual Climate Prediction and Predictability.* National Academies Press, 193 pp., https://doi.org/10.17226/12878.

Neri, A., G. Villarini, and F. Napolitano, 2020: Intraseasonal predictability of the duration of flooding above National Weather Service flood warning levels across the U.S. Midwest. *Hydrol. Processes*, **34**, 4505–4511, https://doi.org/10.1002/hyp.13902.

Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.*, **44**, 8520–8529, https://doi.org/10.1002/2017GL074088.

Osman, M., C. A. S. Coelho, and C. S. Vera, 2021: Calibration and combination of seasonal precipitation forecasts over South America using ensemble regression. *Climate Dyn.*, **57**, 2889–2904, https://doi.org/10.1007/s00382-021-05845-2.

Pakdaman, M., Y. Falamarzi, I. Babaeian, and Z. Javanshiri, 2020: Post-processing of the North American Multi-Model Ensemble for monthly forecast of precipitation based on neural network models. *Theor. Appl. Climatol.*, **141**, 405–417, https://doi.org/10.1007/s00704-020-03211-6.

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.

Pegion, K., T. DelSole, E. Becker, and T. Cicerone, 2019: Assessing the fidelity of predictability estimates. *Climate Dyn.*, **53**, 7251–7265, https://doi.org/10.1007/s00382-017-3903-7.

Pillai, P. A., S. A. Rao, A. Srivastava, D. A. Ramu, M. Pradhan, and R. S. Das, 2021: Impact of the tropical Pacific SST biases on the simulation and prediction of Indian summer monsoon rainfall in CFSv2, ECMWF-System4, and NMME models. *Climate Dyn.*, **56**, 1699–1715, https://doi.org/10.1007/s00382-020-05555-1.

Pons, D., Á. G. Muñoz, L. M. Meléndez, M. Chocooj, R. Gómez, X. Chourio, and C. González Romero, 2021: A coffee yield next-generation forecast system for rain-fed plantations: The case of the Samalá watershed in Guatemala. *Wea. Forecasting*, **36**, 2021–2038, https://doi.org/10.1175/WAF-D-20-0133.1.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Risbey, J. S., and Coauthors, 2021: Standard assessments of climate forecast skill can be misleading. *Nat. Commun.*, **12**, 4346, https://doi.org/10.1038/s41467-021-23771-z.

Romero, C. G., A. Munoz, A. M. G. Solorzano, S. J. Mason, X. M. Chourio, and D. Pons, 2020: When rainfall meets hunger: Towards an early-action system for acute undernutrition in Guatemala. *2020 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract GC051-0012.

Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517, https://doi.org/10.1175/JCLI3812.1.

——, 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.

Seager, R., J. Nakamura, and M. Ting, 2020: Prediction of seasonal meteorological drought onset and termination over the southern Great Plains in the North American Multimodel Ensemble. *J. Hydrometeor.*, **21**, 2237–2255, https://doi.org/10.1175/JHM-D-20-0023.1.

Shao, Y., Q. J. Wang, A. Schepen, and D. Ryu, 2021a: Embedding trend into seasonal temperature forecasts through statistical calibration of GCM outputs. *Int. J. Climatol.*, **41**, E1553–E1565, https://doi.org/10.1002/joc.6788.

——, ——, ——, and ——, 2021b: Going with the trend: Forecasting seasonal climate conditions under climate change. *Mon. Wea. Rev.*, **149**, 2513–2522, https://doi.org/10.1175/MWR-D-20-0318.1.

——, ——, ——, ——, and F. Pappenberger, 2021c: Improved trend-aware post-processing of GCM seasonal precipitation forecasts. *J. Hydrometeor.*, **23**, 25–37, https://doi.org/10.1175/JHM-D-21-0099.1.

Shin, C.-S., and B. Huang, 2019: A spurious warming trend in the NMME equatorial Pacific SST hindcasts. *Climate Dyn.*, **53**, 7287–7303, https://doi.org/10.1007/s00382-017-3777-8.

Siqueira, L., B. P. Kirtman, and L. C. Laurindo, 2021: Forecasting remote atmospheric responses to decadal Kuroshio stability transitions. *J. Climate*, **34**, 379–395, https://doi.org/10.1175/JCLI-D-20-0139.1.

Slater, L. J., and G. Villarini, 2018: Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach. *Geophys. Res. Lett.*, **45**, 6504–6513, https://doi.org/10.1029/2018GL077945.

——, ——, and A. A. Bradley, 2017: Weighting of NMME temperature and precipitation forecasts across Europe. *J. Hydrol.*, **552**, 646–659, https://doi.org/10.1016/j.jhydrol.2017.07.029.

——, ——, ——, and G. A. Vecchi, 2019: A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed. *Climate Dyn.*, **53**, 7429–7445, https://doi.org/10.1007/s00382-017-3794-7.

Small, R. J., and Coauthors, 2014: A new synoptic scale resolving global climate simulation using the Community Earth System Model. *J. Adv. Model. Earth Syst.*, **6**, 1065–1094, https://doi.org/10.1002/2014MS000363.

Smith, D. M., and Coauthors, 2013: Real-time multi-model decadal climate predictions. *Climate Dyn.*, **41**, 2875–2888, https://doi.org/10.1007/s00382-012-1600-0.

Strazzo, S., D. C. Collins, A. Schepen, Q. J. Wang, E. Becker, and L. Jia, 2019: Application of a hybrid statistical–dynamical system to seasonal prediction of North American temperature and precipitation. *Mon. Wea. Rev.*, **147**, 607–625, https://doi.org/10.1175/MWR-D-18-0156.1.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc.*, **A365**, 2053–2075, https://doi.org/10.1098/rsta.2007.2076.

Thiaw, W. M., and V. B. Kumar, 2015: NOAA's African desk: Twenty years of developing capacity in weather and climate forecasting in Africa. *Bull. Amer. Meteor. Soc.*, **96**, 737–753, https://doi.org/10.1175/BAMS-D-13-00274.1.

Thober, S., R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego, 2015: Seasonal soil moisture drought prediction over Europe using the North American Multi-Model Ensemble (NMME). *J. Hydrometeor.*, **16**, 2329–2344, https://doi.org/10.1175/JHM-D-15-0053.1.

Tippett, M. K., T. DelSole, S. J. Mason, and A. G. Barnston, 2008: Regression-based methods for finding coupled patterns. *J. Climate*, **21**, 4384–4398, https://doi.org/10.1175/2008JCLI2150.1.

——, M. Ranganathan, M. L'Heureux, A. G. Barnston, and T. DelSole, 2019: Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Climate Dyn.*, **53**, 7497–7518, https://doi.org/10.1007/s00382-017-3721-y.

——, M. L. L'Heureux, E. J. Becker, and A. Kumar, 2020: Excessive momentum and false alarms in late-spring ENSO forecasts. *Geophys. Res. Lett.*, **47**, e2020GL087008, https://doi.org/10.1029/2020GL087008.

Trenary, L., T. DelSole, M. K. Tippett, and B. Doty, 2015: Was the cold eastern US winter of 2014 due to increased variability [in "Explaining Extreme Events of 2014 from a Climate Perspective"]? *Bull. Amer. Meteor. Soc.*, **96** (12), S15–S19, https://doi.org/10.1175/BAMS-D-15-00138.1.

Turco, M., S. Jerez, F. J. Doblas-Reyes, A. AghaKouchak, M. C. Llasat, and A. Provenzale, 2018: Skillful forecasting of global fire activity using seasonal climate predictions. *Nat. Commun.*, **9**, 2718, https://doi.org/10.1038/s41467-018-05250-0.

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

van den Dool, H., E. Becker, L.-C. Chen, and Q. Zhang, 2017: The probability anomaly correlation and calibration of probabilistic forecasts. *Wea. Forecasting*, **32**, 199–206, https://doi.org/10.1175/WAF-D-16-0115.1.

Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, https://doi.org/10.1175/JCLI-D-14-00158.1.

Villarini, G., B. Luitel, G. A. Vecchi, and J. Ghosh, 2019: Multi-model ensemble forecasting of North Atlantic tropical cyclone activity. *Climate Dyn.*, **53**, 7461–7477, https://doi.org/10.1007/s00382-016-3369-z.

Wanders, N., and E. F. Wood, 2016: Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environ. Res. Lett.*, **11**, 094007, https://doi.org/10.1088/1748-9326/11/9/094007.

Wang, C., W. Zhang, and G. Villarini, 2021: On the use of convolutional Gaussian processes to improve the seasonal forecasting of precipitation and temperature. *J. Hydrol.*, **593**, 125862, https://doi.org/10.1016/j.jhydrol.2020.125862.

Weisheimer, A., and Coauthors, 2009: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett.*, **36**, L21711, https://doi.org/10.1029/2009GL040896.

White, C. J., and Coauthors, 2022: Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-20-0224.1, in press.

Widlansky, M. J., and Coauthors, 2017: Multimodel ensemble sea level forecasts for tropical Pacific islands. *J. Appl. Meteor. Climatol.*, **56**, 849–862, https://doi.org/10.1175/JAMC-D-16-0284.1.

Wittenberg, A. T., and Coauthors, 2018: Improved simulations of tropical Pacific annual-mean climate in the GFDL FLOR and HiFLOR coupled GCMs. *J. Adv. Model. Earth Syst.*, **10**, 3176–3220, https://doi.org/10.1029/2018MS001372.

WMO, 2020: Guidance on operational practices for objective seasonal forecasting. WMO Tech. Rep. 1246, 106 pp., https://library.wmo.int/index.php?lvl=notice_display&id=21741#.Yg1NFejMKUk.

Wolter, K., R. M. Dole, and C. A. Smith, 1999: Short-term climate extremes over the continental United States and ENSO. Part I: Seasonal temperatures. *J. Climate*, **12**, 3255–3272, https://doi.org/10.1175/1520-0442(1999)012<3255:STCEOT>2.0.CO;2.

Woollings, T., J. M. Gregory, J. G. Pinto, M. Reyers, and D. J. Brayshaw, 2012: Response of the North Atlantic storm track to climate change shaped by ocean–atmosphere coupling. *Nat. Geosci.*, **5**, 313–317, https://doi.org/10.1038/ngeo1438.

Xu, L., N. Chen, X. Zhang, Z. Chen, C. Hu, and C. Wang, 2019: Improving the North American Multi-Model Ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. *Climate Dyn.*, **53**, 601–615, https://doi.org/10.1007/s00382-018-04605-z.

Yoo, J. H., and I.-S. Kang, 2005: Theoretical examination of a multi-model composite for seasonal prediction. *Geophys. Res. Lett.*, **32**, L18707, https://doi.org/10.1029/2005GL023513.

Zhang, W., and G. Villarini, 2019: Seasonal forecasting of western North Pacific tropical cyclone frequency using the North American Multi-Model Ensemble. *Climate Dyn.*, **52**, 5985–5997, https://doi.org/10.1007/s00382-018-4490-y.

——, ——, L. Slater, G. A. Vecchi, and A. A. Bradley, 2017: Improved ENSO forecasting using Bayesian updating and the North American Multimodel Ensemble (NMME). *J. Climate*, **30**, 9007–9025, https://doi.org/10.1175/JCLI-D-17-0073.1.

Zhou, Y., and H.-M. Kim, 2018: Prediction of atmospheric rivers over the North Pacific and its connection to ENSO in the North American Multi-Model Ensemble (NMME). *Climate Dyn.*, **51**, 1623–1637, https://doi.org/10.1007/s00382-017-3973-6.