A Deficit of Seasonal Temperature Forecast Skill over West Coast Regions in NMME

ZHENHAI ZHANG, DAVID W. PIERCE, AND DANIEL R. CAYAN

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California

(Manuscript received 6 October 2018, in final form 26 April 2019)

ABSTRACT

This study investigates the forecast skill of seasonal-mean near-surface (2 m) air temperature in the North American Multimodel Ensemble (NMME) Phase 2, with a focus on the West Coast of the United States. Overall, 1-month lead time NMME forecasts exhibit skill superior or similar to persistence forecasts over many continental regions, and skill is generally higher over the ocean than the continent. However, forecast skill along most West Coast regions is markedly lower than in the adjacent ocean and interior, especially during the warm seasons. Results indicate that the poor forecast skill along the West Coast of the United States reflects deficiencies in their representation of multiple relevant physical processes. Analyses focusing on California find that summer forecast errors are spatially coherent over the coastal region and the inland region individually, but the correlation of forecast errors between the two regions is low. Variation in forecast performance over the coastal California region is associated with anomalous geopotential height over the lower middle latitudes and subtropics of the eastern Pacific, North America, and the western Atlantic. In contrast, variation in forecast performance over the inland California region is associated with the atmospheric circulation over the western United States. Further, it is found that forecast errors along the California coast are linked to anomalies of low cloudiness (stratus clouds) along the coastal region.

1. Introduction

Subseasonal to seasonal forecasts are long-term forecasts for 2 weeks to 12 months into the future ([National Academies of Sciences, Engineering, and Medicine 2016](#)). In contrast to weather forecasts, which rely mainly on atmospheric behavior predicated from initial atmospheric conditions, subseasonal to seasonal forecast skill derives partly from the initial atmospheric conditions and partly from the effects of slowly evolving boundary conditions, such as the sea surface temperature, sea ice, and soil moisture. Reliable forecasts at subseasonal to seasonal scale are urgently needed by decision makers in energy, agriculture, water management, public health, and disaster preparedness ([White et al. 2017](#)), especially in densely populated regions, such as some coastal areas.

There are many challenges to subseasonal to seasonal forecasting based on dynamical models. Previous studies demonstrate that the old generation of dynamical models can be less skillful than much simpler statistical models ([Winkler et al. 2001](#); [Newman et al. 2003](#);

[Rodrigues et al. 2014](#)). [Newman and Sardeshmukh \(2017\)](#) demonstrate that all eight individual dynamical models are less skillful than a linear inverse model in the prediction of tropical sea surface temperature, while the ensemble mean of the dynamical models has the forecast skill similar to the linear inverse model. On the other hand, in some cases dynamical models are found to outperform linear inverse statistical models, specifically when an ensemble prediction methodology is employed ([Pegion and Sardeshmukh 2011](#)). Recent research into seasonal climate predictability using dynamical prediction systems offers substantial evidence that subseasonal to seasonal forecast using dynamical models can be useful to the applications community ([Kirtman et al. 2014](#); [White et al. 2017](#)).

Progress in subseasonal to seasonal forecasting has been enhanced by multi-institutional international collaborations ([Brunet et al. 2010](#); [Kirtman et al. 2014](#)), which expand forecast ensembles to include different dynamical models. The North American Multimodel Ensemble (NMME; [Kirtman et al. 2014](#)) is one of the primary examples of these collaborations.

The NMME is a dynamical climate forecasting system using state-of-the-art coupled models from U.S. and

Corresponding author: Zhenhai Zhang, zhz422@ucsd.edu

Canadian modeling centers (Kirtman et al. 2014). It provides a multimodel framework for assessing the subseasonal to seasonal forecast skill in dynamical models. This multimodel forecast system demonstrates improved skill compared to individual models, such as the NOAA operational CFSv2 (Saha et al. 2014), in forecasting large-scale climate features (Kirtman et al. 2014; Becker et al. 2014). To this end, NMME benefits from both a large number of ensemble members and model diversity. There are currently two phases of NMME, Phase 1 and Phase 2. We use NMME Phase 2 in this work, as it represents models currently contributing to the real-time NMME forecasting system.

Many studies have evaluated the forecast skills of NMME for sea surface temperature, precipitation, and surface temperature at global scale (Becker et al. 2014; Mo and Lyon 2015; Becker and van den Dool 2016), and over North America and the continental United States (Chen et al. 2017; Hervieux et al. 2017; Infanti and Kirtman 2014; Slater et al. 2016; Wang 2014). These studies, most of which used NMME Phase 1 data exclusively, show that the NMME forecast skill varies significantly from season to season and from region to region. Jacox et al. (2017) investigated the NMME forecast skill of seasonal sea surface temperature in the California Current System and found that the NMME models were relatively more skillful for the forecasts of February–April than over the rest of the year. Shukla et al. (2015) found that the NMME forecast skill of surface land temperature in California was generally low, with relatively higher skill confined to interior regions in July–September.

Although subseasonal to seasonal forecast skill has large regional variability, it generally exhibits higher skill over the ocean than over the continent (e.g., Becker et al. 2014), largely because oceanic temperatures evolve more slowly than land temperatures due to the greater thermal inertia in the ocean (Frankignoul 1985; Goddard et al. 2001). Given higher forecast skill over the ocean, a similar boost in forecast skill over adjacent coastal areas that are influenced by the marine environment may seem plausible. However, some coastal areas have their own complicated dynamical processes such as land–sea breezes, coastal jets, coastal upwelling, and marine stratus decks, which could result in a more complicated and difficult to forecast region than the open ocean farther offshore. A more comprehensive evaluation of dynamical forecasts of seasonal-mean surface temperature over the coastal region, including both offshore ocean and adjacent continental land temperatures, has not been conducted. Furthermore, the fundamental physical processes influencing the spatially and temporally varying forecast skill in dynamical models are also

unexplored. These needs are especially acute in regions where the population is concentrated along the coast.

In this study, an evaluation of the performance of seasonal-mean temperature predictions is conducted over selected coastal areas, with an emphasis on western North America and a focus on the seasonal time scale, to improve our understanding of NMME seasonal forecast skill at lead times up to a few months. The forecast skill of global seasonal-mean near-surface (2 m) air temperature in NMME models is evaluated seasonally for March–May (MAM), June–August (JJA), September–November (SON), and December–February (DJF), and compared to the persistence forecast skill. We find that the forecast skill of seasonal-mean temperature over some west coast regions is markedly lower than the skill over either the nearby ocean or the inland continent, especially during the warm season. Moreover, coastal regions often contain a disproportionate fraction of the population, which suffers lower forecast skill than either offshore or inland. To explore the physical processes responsible for that low skill in a setting that affects a large number of people, we focus on the California coastal region. Associations between forecast errors and atmospheric circulation patterns, in the form of the 500-hPa geopotential height anomaly, are examined. The influence of local coastal low cloudiness on forecast skill, a key feature of the climate in Southern California coastal regions that is typically poorly simulated by dynamical models, is also investigated.

2. Data and methods

We use ensembles of retrospective seasonal-mean near-surface (2 m) air temperature forecasts (often called “hindcasts”) of seven models from the NMME Phase 2 (Table 1). The NMME Phase 2 models are currently contributing to the real-time forecasts and represent the skill of the current NMME real-time forecasting system. All the NMME models are coupled dynamical models and include key global climate fluctuations, such as El Niño–Southern Oscillation (ENSO) and the Madden–Julian oscillation (MJO), through initial conditions and internal model dynamics. However, it should be kept in mind that individual model quality in reproducing such natural climate fluctuations or their teleconnections can be poor, and may contribute to low model forecast skill in the region of interest. We downloaded the data from the NMME Phase 2 dataset (<https://www.earthsystemgrid.org/search.html?Project=NMME>).

All the models provide forecasts for lead times of at least 0–9 months. A total of 29 years of hindcasts (1982–2010) are available for all models. The hindcasts for the seasons in 2010 are not used in this study since the 2010

TABLE 1. The models used in this study from NMME Phase 2.

Model	Organization	Hindcast period	Lead time (months)	Ensemble size	Reference
CanCM3	Canadian Meteorological Center (CMC)	1981–2012	0–11	10	Merryfield et al. (2013)
CanCM4	Canadian Meteorological Center (CMC)	1981–2012	0–11	10	Merryfield et al. (2013)
CCSM4	National Center for Atmospheric Research (NCAR)	1982–2010	0–11	10	Infanti and Kirtman (2016)
CESM1	National Center for Atmospheric Research (NCAR)	1980–2010	0–11	10	Lawrence et al. (2012)
CFSv2	National Centers for Environmental Prediction (NCEP)	1982–2010	0–9	24 (28)	Saha et al. (2014)
FLORB01	Geophysical Fluid Dynamics Laboratory (GFDL)	1980–2014	0–11	12	Vecchi et al. (2014)
GEOS5	National Aeronautics and Space Administration (NASA)	1982–2012	0–8	10	Vernieres et al. (2012)

winter (from December 2010 to February 2011) is not entirely available. Five models (CanCM3, CanCM4, CCSM4, CESM1, and GEOS5) have 10 ensemble members, FLORB01 has 12 members, and CFSv2 has 24 (28) members (only 24 members are used, 28 members are available for November only). Hence, a total of 86 ensemble members are used in this study. CFSv2 is initialized every fifth day, with four members per day. The ensemble members of the other six models are all initialized at 0000 UTC on the first day of the month. In this study, “lead time 0 month” means a forecast made from initial conditions at the beginning of the first month for this season. For example, the forecast for seasonal-mean JJA temperature initialized at the beginning of June is defined as lead time 0 month, and the forecast initialized at the beginning of May is defined as “lead time 1 month.” Note that in this terminology, even a “lead-0 forecast” involves forecasted future conditions since it is the average of forecasted values from 1 day to 3 months into the future, and it is often desired by stakeholders (e.g., in the energy industry). Similarly, a “lead-1 forecast” is the average over forecasts from 1 to 4 months into the future. We calculate the seasonal-mean near-surface air temperature from the daily values. In total, 6 of the 7 models provide outputs at 1° latitude $\times$ 1° longitude, while FLORB01 provides outputs at 0.5° latitude $\times$ $\sim 0.6^\circ$ longitude. All data were interpolated onto a common $1^\circ \times 1^\circ$ latitude–longitude grid before analysis.

The NMME forecasts are compared to temperature observations, including the Climatic Research Unit (CRU) TS 3.24 ([Harris et al. 2014](#)) daily mean surface temperature ($0.5^\circ \times 0.5^\circ$) over land and the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; $1^\circ \times 1^\circ$) ([Rayner et al. 2003](#)) over the ocean. The observational data are interpolated to the same $1^\circ \times 1^\circ$ grid as the NMME data. The daily land temperature dataset is time averaged to months and seasons for the

forecast evaluations conducted here. The NMME forecasts are also compared to the observations from four stations (San Francisco International Airport, San Diego International Airport, Tahoe City, and Parker Dam), which represent both coastal and inland locations and have data covering the NMME hindcast period 1982–2010. These station (STN) data are obtained from the National Climatic Data Center (NCDC) Global Historical Climatology Network Daily dataset.

Persistence, the simplest forecast possible, is a forecast that adopts the temperature anomaly that exists at the time of the forecast. Persistence forecasts provide a baseline reference to assess the NMME model forecast skill. In the persistence forecast, the mean temperature anomaly of two months (one month before and one month after the model initial time) is used to forecast the seasonal-mean temperature anomaly. For example, our 0-month lead time persistence forecast of JJA temperature anomaly is the average temperature anomaly of May and June. This persistence forecast is different from the persistence forecast in some previous studies, which use the anomaly from the month prior to the model initial time to conduct the forecast (e.g., [Stock et al. 2015](#); [Jacox et al. 2017](#)). However, the persistence forecast in this study (averaged temperature anomaly of one month before and one month after the model initial time) is the approximate temperature anomaly at the time of the model initialization. When compared to dynamical forecasts that are initialized at the start of the month, this version of the persisted anomaly provides a more difficult measure of forecast skill than using persisted anomalies from the prior month.

Monthly mean 500-hPa geopotential height from the Climate Forecast System Reanalysis (CFSR; [Saha et al. 2010](#)) is used to examine the relationship between atmospheric large-scale circulation and NMME forecast errors.

Coastal low cloudiness (stratus or stratocumulus clouds) data used in this study is provided by [Clemesha et al. \(2016\)](#), derived from a new high-resolution (4km and half-hour) satellite-derived record of low clouds from NASA/NOAA *Geostationary Operational Environmental Satellite-9, -10, -11, and -15* (*GOES-9, GOES-10, GOES-11, and GOES-15*). The data covers the ocean and the U.S. West Coast (25°–50°N, 130°–113°W), over the period of May–September, when the coastal low cloudiness is greatest, from 1996 to 2014. The coastal low cloudiness data is interpolated to the 1° × 1° common grid and aggregated to the monthly mean. Only the 14 years of overlap with the NMME forecasts (1996–2009) are used in this study.

In this study, we focus on dynamical forecasts at monthly to seasonal time leads, made from multiple model runs and often cast in probabilistic terms to provide information about the likelihood of anomalous seasonal-mean temperature anomalies. We make extensive use of the anomaly correlation coefficient (ACC), which is the correlation between the seasonal anomalies of the model forecast and observations. If the variation pattern of the forecast anomalies is perfectly coincident with that of observations (a perfect forecast), a maximum ACC value of 1 is obtained; if the forecast pattern is completely reversed from observations, a minimum ACC value of –1 is obtained. The ACC is well suited to this evaluation because it is a concise measure that captures the models' skill in forecasting departures from mean climatological conditions over many forecast seasons. In addition, the performance is also evaluated using correlations and composites of the NMME forecast error, expressed as the difference between the forecasted and observed seasonal-average temperature anomalies normalized by standard deviation; these measures allow additional insight into the sources of model forecast error.

The Brier skill score (BSS; [Brier 1950](#)) is utilized to verify the probability forecasts. The seasonal-mean temperature is categorized into upper (warm), normal, and lower (cold) terciles. The full multimodel ensemble (86 members) is used to forecast the likelihood of a season in warm or cold terciles. The BSS represents the degree of forecast improvements in reference to the climatological forecast. The BSS is equal to or lower than 1: BSS = 1 indicates a perfect forecast, BSS = 0 indicates the same skill with the climatological forecast, and BSS < 0 indicates lower skill than the climatological forecast.

3. Results

a. Skill of seasonal-mean temperature forecasts

The seasonal-mean surface temperature forecast of NMME ensemble mean exhibits positive forecast skill

(ACC) over the globe at lead time 0 month in each season ([Figs. 1a–d](#)). However, there are large regional variabilities in the forecast skill that are consistent across seasons. Over the ocean, the ACC ranges from 0.6 to 0.9 over most areas. The highest ACC (~0.9) is concentrated over the eastern tropical Pacific Ocean, where ENSO variability is high, while the relatively lower ACC at some area of southern Indian Ocean is <0.6. The forecast skill over the continent is generally lower than over the ocean, which is consistent with the results from [Becker et al. \(2014\)](#), who used hindcasts from NMME Phase 1. Although relatively high ACC (>0.7) can be found over land (e.g., some region of Africa), the forecast skill over most continental regions is around or lower than 0.6. Meanwhile, there are large seasonal variabilities. For example, the forecast skill over the western U.S. interior is ~0.7 in JJA ([Fig. 1b](#)) and only ~0.3 in DJF ([Fig. 1d](#)).

The NMME forecast skill is compared with the skill of the persistence forecast in [Figs. 1e–h](#). Red colors indicate locations where the NMME forecast skill is superior to persistence at the 90% confidence level; blue indicates the opposite. A Bootstrap approach ([Diaconis and Efron 1983](#)) is utilized to conduct the statistical confidence level. The NMME forecast skill is significantly lower over many regions in all seasons, especially over the southern Indian Ocean (ACC 0.1–0.7 lower), while it exceeds the persistence forecast over only some limited area in particular season ([Figs. 1e–h](#)). This result is different from [Jacox et al. \(2017\)](#), who found that the models of NMME Phase 1 exhibit significant skill above their persistence forecast (using temperature anomaly from the month prior to model initialization) in the California Current. This is mainly due to the different definitions of persistence forecast as described in the data and methods section and may also be due to different models used in their study.

At lead time 1 month, the spatial distribution of NMME forecast skill is similar to that at lead time 0 month, exhibiting relatively high skill over the ocean and low skill over the continent ([Figs. 2a–d](#)). However, the NMME forecast skill over the continents decreases quickly (by 0.1–0.5), while the skill over the tropical ocean remains high (0.7–0.9) in all seasons. The persistence forecast skill over the continents decreases even more dramatically, with values close to zero over many inland continental regions at middle and high latitudes (not shown). As a result, the NMME exhibits forecast skill superior or similar to persistence forecast over many continental regions ([Figs. 2e–h](#)). The NMME forecast skill over the southern Indian Ocean is still generally lower than persistence in all seasons.

FIG. 1. NMME ensemble mean surface temperature forecast skill measured by the ACC for lead time 0 month for (a) MAM, (b) JJA, (c) SON, and (d) DJF. The difference between NMME forecast skill and persistence forecast skill (NMME minus persistence; colors) for lead time 0 month for (e) MAM, (f) JJA, (g) SON, and (h) DJF, showing differences at or exceeding the 90% significant confidence level.

b. Forecast skill over western coasts

More skillful seasonal forecasts over coastal regions might be anticipated since the forecast of those regions may benefit from the high forecast skill over the nearby ocean. In fact, this does happen over some coastal regions, such as the Canadian west coast. The NMME

forecast skill (ACC) over the Canadian west coast is around 0.6, which is close to the ACC of the adjacent ocean (~ 0.7), and decreases gradually to <0.3 over the inland continent at lead time 1 month in MAM, JJA, and DJF (Figs. 2a–d). However, the NMME forecast skill is poor ($\text{ACC} < 0.3$) over the California coast at lead times of 0 and 1 month (Figs. 1a–d and 2a–d). This is markedly

FIG. 2. As in Fig. 1, but for lead time 1 month. The black boxes are the west coast regions for Fig. 3.

lower than the skill found in either the nearby offshore ocean or the inland continent, especially in warm seasons. For example, in JJA the ACC is nearly 0 along the California coast while it is around 0.4 over the nearby ocean and 0.6 over adjacent Nevada at lead time 1 month (the black box in Fig. 2b). This type of poor forecast skill in NMME is also found over several other west coast regions in the warm season (the black boxes

in Fig. 2), while it does not occur in the persistence forecast. The narrow regions of dark blue in the four black boxes in Figs. 2e–h indicate that the NMME forecast skill is significantly lower (0.3–0.7) than the persistence forecast in these narrow coastal strips. A closer view of the ACC over these regions is shown in Fig. 3, which focuses on the west coasts of North America, South America, Australia, and South Africa. For the west

FIG. 3. NMME ensemble mean surface temperature forecast skill (ACC) for lead time 1 month over the west coast regions (black boxes in Fig. 2), for (a) JJA in North America, (b) SON in South America, (c) DJF in Australia, and (d) DJF in South Africa. Colors depict regions where ACC statistical significance is equal or exceeds the 90% confidence level. The blue lines are the transects used in Figs. 5 and 6.

coast of South America, SON is selected as the warm season since in DJF the NMME exhibits forecast skill close to zero over the nearby ocean (an oceanic pattern, maybe related to the Peru Current), which is different from the pattern of forecast skill over the other west coast regions. The NMME 1-month lead time forecast ACC along these coasts is close to 0 and not significant, yet the ACC over the nearby ocean and adjacent inland continent is much higher (0.4–0.7) and significantly different than zero.

In addition to the gridded CRU data, the NMME forecast skill is also evaluated using the observation data from four stations: San Francisco International Airport, San Diego International Airport, Tahoe City, and Parker Dam (the blue dots in Fig. 7). The reason for doing this is to verify that our results are not a result of gridding errors in the CRU data near the coasts. For each location, the ACC is calculated between the time series from the observation at the station and the NMME forecast at the nearest grid cell. The results based on the station data are consistent with the results based on the

gridded CRU data. Table 2 shows the ACC between NMME forecasts at lead time 1 month and observations for the four locations. At San Francisco, the ACC between NMME forecast and station observation (STN) is very close to the ACC between NMME and CRU with the lowest skill (~ 0.06) in JJA and highest skill (0.36–0.38) in DJF. At Tahoe City, both of the ACCs based on STN and CRU are relatively higher in MAM and JJA, and relatively lower in SON and DJF. Overall, in JJA the ACC at the coastal location (San Francisco, ~ 0.06) is significantly lower than the ACC at the nearby inland location (Tahoe City, 0.68–0.78). The comparison between San Diego and Parker Dam is similar. The ACC between the gridded CRU data and the station observation are significantly high, around or above 0.75, which increases our confidence to use gridded CRU data to evaluate the NMME forecasts.

The BSS of NMME ensemble at lead time 1 month for the upper (warm) and lower (cold) terciles over those

TABLE 2. The ACC for seasonal-mean temperature between the NMME forecast at lead time 1 month and CRU/STN, and between CRU and STN for San Francisco, San Diego, Tahoe City, and Parker Dam. The bold values are statistically significant ACCs that equal or exceed the 90% confidence level.

	MAM	JJA	SON	DJF
San Francisco				
NMME-CRU	0.12	0.06	0.12	0.36
NMME-STN	0.14	0.06	0.09	0.38
CRU-STN	0.98	0.91	0.95	0.94
San Diego				
NMME-CRU	-0.11	-0.03	-0.13	-0.07
NMME-STN	-0.11	-0.06	0.02	-0.06
CRU-STN	0.82	0.86	0.78	0.76
Tahoe City				
NMME-CRU	0.45	0.78	0.25	0.22
NMME-STN	0.30	0.68	-0.06	0.12
CRU-STN	0.86	0.83	0.79	0.79
Parker Dam				
NMME-CRU	0.52	0.77	0.50	0.29
NMME-STN	0.28	0.53	0.34	0.16
CRU-STN	0.80	0.72	0.72	0.74

four west coast regions exhibits a similar pattern with the ACC along the coast (Fig. 4). The BSS for both warm and cold terciles is close to or lower than 0 along the coastal region of North America, which indicates that the NMME probability forecast skill is close to or lower than the climatological forecast. The probability forecast skill of the NMME ensemble cannot surpass the climatological forecast along the other three west coasts either. Meanwhile, the BSS varies from 0.1 to 0.4 over most areas of the nearby inland continent and the

adjacent ocean, suggesting higher skill than climatological forecast. The BSS for the other lead times also has relatively lower skill along those coasts (not shown).

Results from the seven individual models contributing to NMME were examined to see how consistent this deficit of coastal predictability is. The poor forecasts over those west coast regions in warm season are seen in nearly all models. Figure 5 shows the ACC at grid cells along the ocean-land transect shown as blue lines in Fig. 3, in the warm season at lead time 1 month. The individual NMME models exhibit a large variety of forecast skills. For example, the ACC at grid W3 (the third grid cell over the ocean/west from the coastline) is 0.58 for CESM1 and is only -0.35 for CanCM3 for the west coast of South America in SON (Fig. 5b). The spread of forecast skill over the west coast of North America (California coast) is relatively smaller than over the other west coasts (Fig. 5). However, in all four regions the ACC profiles of the NMME ensemble mean and the majority of the individual models form a “V” shape from the ocean to inland continent, exhibiting high skill over the oceanic region, low skill over the immediate coastal area, and high skill over the interior land region. For example, along the transect of Australia from W6 (the sixth grid cell over the ocean/west from the coastline) to E1 (the first grid cell over the continent/east from the coastline) to E6, the ACC of the NMME ensemble mean changes from 0.72 to 0.01 to 0.41. The ACC profile of the South Africa west coast has the lowest point slightly offshore, which is different from the other west coast regions with the lowest point at the immediate coastal land area. The NMME ensemble

FIG. 4. As in Fig. 3, but for the BSS for the (top) upper (warm) tercile and (bottom) lower (cold) tercile.

FIG. 5. Forecast skill (ACC) of lead time 1 month seasonal temperature forecasts along the transects (blue lines in Fig. 3) for the persistence, NMME ensemble mean, and the individual models for (a) JJA in North America, (b) SON in South America, (c) DJF in Australia, and (d) DJF in South Africa. W1–W6 (E1–E6) on the x axis indicate the first through sixth grid cells (1° resolution) over the ocean/west (continent/east) from the coastline.

mean forecast skill is higher than the individual models over the offshore ocean of North America and Australia (Figs. 5a and 5c). However, over the other regions there are some individual models with notably poor skill, so the NMME ensemble mean skill is lower than the relatively good individual models. Some of those relatively good individual NMME models and the NMME ensemble mean have similar or greater forecast skill than persistence over the inland continent, while the models have lower skill than persistence over the oceanic and coastal regions (Fig. 5). The ACC profiles along transects at different latitudes within those west coast boxes exhibit similar results (not shown). The consistency of these results across different regions and models suggests that the poor forecast skill over those west coast regions is caused by systematic model errors in the simulation of common western coastal processes.

The profile of forecast skill along the ocean–land transect is shown for different lead times in Fig. 6. The NMME and persistence forecast skills at a lead time 3 months are generally lower than the skills at a lead time 1 month, but have similar patterns. The NMME ensemble mean exhibits a “V” shape from the ocean to

inland continent while the persistence forecast does not have such a shape. The NMME ensemble mean has similar or higher forecast skill than persistence over the interior in all of the four west coast regions at both lead time 1 month and 3 months. Meanwhile, the NMME has lower forecast skills than persistence over the coastal region at different lead times.

c. Factors related to temperature forecast errors over the California coast

The results in section 3b show that the NMME forecast skill of seasonal-mean surface temperature is near zero over some west coast regions, in contrast to higher skill to the west over the nearby ocean and east over the adjacent interior. This contrast is especially strong in the warm season. In this section, we focus on the California coastal region to investigate the factors related to the poor forecast skill. The NMME forecast error and some related meteorological factors are examined to explore potential causes of the coastal deficit in temperature predictability. The forecast error is defined as the NMME forecasted seasonal temperature anomaly minus the observational temperature anomaly. Then the forecast error

FIG. 6. As in Fig. 5, but for the NMME ensemble mean (solid lines) and the persistence forecast (dashed lines) at lead time 1 (red) and 3 (blue) months.

is related to local and regional meteorological conditions. We focus on the overlapping summer seasons [May–July (MJJ), JJA, and July–September (JAS)], because the forecast skill deficit is large in that season and the coastal low cloudiness data is available for this time of the year. Three overlapping summer seasons were used to increase the sample size of forecast errors and the related factors, especially for the coastal low cloudiness (available for 1996–2009), although the nonindependence of these periods must be kept in mind.

To investigate the spatial coherence of forecast errors, the correlation between the time series of NMME temperature forecast error at a specific location and that at all locations on the map at forecast lead time 1 month are calculated for two California coastal locations (San Francisco and San Diego, blue dots in Figs. 7a and 7c) and two eastern California inland locations (Tahoe City and Parker Dam, blue dots in Figs. 7b and 7d). Note that a 95% confidence level is used here and in the subsequent figures since they describe climatological connections between different regions, unlike the 90% confidence level used in previous figures describing forecast skill. This is because the current state of GCMs is that they do a better job reproducing the connections

between related climate fields (temperature, cloudiness, geopotential height) than they do in producing skillful seasonal forecasts.

The forecast error of San Francisco correlates highly ($r > 0.6$) with forecast errors over the California coastal region and the adjacent ocean, but does not correlate significantly with the forecast errors over inland California (Fig. 7a). The pattern of correlations for San Diego is similar to that of San Francisco, although displaced to the south. Meanwhile, forecast errors at the inland Tahoe City and Parker Dam correlate highly with errors over a broad swath of the inland region, but do not correlate significantly with the errors over the coastal region. These results suggest that a set of common drivers may be at play in causing coastal temperature forecast errors, separate from those that cause forecast errors over the inland region.

To investigate associations of regional forecast error with the larger-scale atmospheric circulation, Fig. 8 shows how temperature forecast errors at those four individual locations correlate with observed 500-hPa geopotential height anomalies over an extensive region from the central North Pacific eastward to the western North Atlantic Ocean. The 500-hPa observations during

FIG. 7. Correlation (contours) of NMME forecast error time series at a specific location and that at all grid points on the map at lead time 1 month for three summer seasons (MJJ, JJA and JAS). Locations are (a) San Francisco, (b) Tahoe City, (c) San Diego, and (d) Parker Dam (blue dot in each panel). Colors indicate locations whose correlation is significant at 95% confidence level or higher.

the three months of the forecast are used. The forecast error at Tahoe City has a strong negative correlation ($r = -0.8$) with the 500-hPa geopotential height anomaly over the U.S. West Coast and a weak positive correlation ($r = 0.3$) at the upstream region over central North Pacific Ocean (Fig. 8b). That is, negative 500-hPa height anomalies over the West Coast result in positive NMME forecast errors (forecast is warmer than observed) and vice versa for positive 500-hPa height anomalies. A similar pattern of 500-hPa correlation occurs in association with forecast errors at Parker Dam (Fig. 8d), which is not surprising since the forecast errors at the inland eastern California region are highly

coherent. However, the patterns of 500-hPa correlation associated with forecast errors at the two California coastal locations are quite different. Forecast errors at San Francisco and San Diego correlate positively ($r = 0.3$ – 0.5) with the 500-hPa anomaly at the subtropical/tropical region and the eastern Pacific Ocean (Figs. 8a and 8c). It indicates that a strengthened subtropical/tropical high and the eastern Pacific ridge produce positive forecast errors along the California coast, wherein positive errors mean that NMME forecasts tend to produce anomalies that are warmer than those of the observations, and vice versa for a weakened subtropical ridge.

FIG. 8. Correlations (contours) between the 500-hPa geopotential height anomaly time series at all grid points and the time series of NMME forecast error at lead time 1 month at (a) San Francisco, (b) Tahoe City, (c) San Diego, and (d) Parker Dam (black dot in each panel) for three summer seasons (MJJ, JJA and JAS). Colors designate correlations significant at or above the 95% confidence level.

In addition to the associations between forecast performance and the large-scale atmospheric circulation, the impact of regional coastal low cloudiness on NMME forecast error is also investigated. Figure 9 shows the local (grid cell by grid cell) correlations between coastal low cloudiness and the NMME forecast error at lead time 1 month for 1996–2009 summer seasons (MJJ, JJA, and JAS), when the cloud data are available. Correlations consider the cloudiness averaged over the same 3-month period as the 3-month temperature forecast and are computed independently for each grid cell. Significant (95% confidence level) positive correlations ($r = 0.3$ – 0.6) occur over most of the coastal region. Around San Francisco the correlation is about $+0.6$, indicating that when there is a positive coastal low cloudiness anomaly around San Francisco, the NMME forecasted temperature tends to be warmer than was actually observed in May–September. Conversely, when the skies are unusually clear in May–September, NMME forecasts tend to be cooler than observed temperatures. Since the coastal low cloudiness can modulate the coastal summer

temperature by reducing the daytime maximum temperature (Jacobellis and Cayan, 2013), higher coastal low cloudiness may lower the summer temperature, and vice versa for situations with clear skies. Meanwhile, the coarse-resolution dynamical models are known to have significant limitations in resolving the low clouds (Randall et al. 2003; Schneider et al. 2017). Together with these previous results, our findings suggest that the models' poor skill in predicting coastal California summer temperatures is a direct result of the models' poor skill in forecasting coastal cloud conditions.

The temperature anomalies in summer seasons at San Francisco and an adjacent area over the ocean (same latitude but 3° to the west of San Francisco) are extracted to examine the NMME forecast errors in more detail. At a lead time of 1 month, the difference between the temperature anomalies in the NMME and observations at San Francisco (land; Fig. 10a) is much larger than the difference over the ocean (Fig. 10b). About 50% of these seasons have NMME anomalies with opposite signs from the observed temperature anomalies.

FIG. 9. Correlation (colors) between coastal low cloudiness anomalies and the NMME forecast errors at lead time 1 month at each grid point in 1996–2009 summer seasons (MJJ, JJA and JAS). Black dots indicate correlations that are statistically significant at the 95% confidence level.

Over the land, the correlation between the NMME and observed temperature anomalies is only about 0.03 and not significant. Meanwhile, over the adjacent ocean, the correlation between the NMME temperature anomalies and those of observations is significant and much higher, about 0.58.

Based on the forecast errors at San Francisco shown in Fig. 10a, the 10 seasons with the largest positive forecast errors (model forecast too warm) and 10 seasons with the smallest absolute forecast errors (anomalies in the model and observation very close) were selected for the period 1996–2009, when the cloud data are available. In the 10 seasons with the largest positive errors (forecast too warm), observations show large positive coastal low cloudiness anomalies (0.5–0.6 standard deviations greater than the mean) in the California central coast (Fig. 11b). In comparison, for the 10 seasons with the lowest errors at San Francisco, the coastal low cloudiness is very close to the climatological mean (Fig. 11a). In other words, forecast skill degrades when cloud conditions depart from the climatological norm. The same analysis is also conducted for San Diego, with similar results. When the model forecast is too warm around San Diego, there are large positive coastal low cloudiness

anomalies around this region (not shown). This indicates that compared to the observations, the NMME ensemble mean overestimate the surface temperature anomalies in May–September around San Francisco and San Diego during summers with positive coastal low cloudiness anomaly. In addition, all of the individual models have the same bias as the NMME ensemble mean.

4. Discussion and conclusions

This study investigates the skill of retrospective seasonal forecasts (hindcasts) of near-surface air temperature using forecasts from seven coupled dynamical models participating in the North American Multimodel Ensemble (NMME) Phase 2 over the period 1982–2009. While NMME seasonal forecasts exhibit skill that exceeds persistence forecast in many regions of the globe at 1-month lead time, the NMME forecasts exhibit a consistent pattern of poor forecast skill, compared to persistence forecasts, over many coastal regions of western North America, especially during warm seasons. The societal impacts of this coastal deficit in forecast skill can be large since in many regions coasts are more densely populated than the adjacent interior.

Possible causes of this NMME forecast error were developed by exploring associations between NMME forecast errors and selected atmospheric fields (500-hPa geopotential height and coastal low cloudiness). The major conclusions are as follows:

- 1) At lead time 0 month, NMME ensemble mean seasonal-average surface temperature skill is positive but lower than or at most equal to persistence skill over many regions globally. At lead time 1 month, the NMME models produce positive skill that exceeds that of persistence forecast over many continental regions.
- 2) Over many west coast regions temperature forecast skill in NMME is poor ($ACC < 0.3$) and markedly lower than the skill over the nearby offshore ocean and inland continent in the warm season. This pattern of poor forecast skill over those west coast regions occurs in the individual models as well as the NMME ensemble average, and at different lead times, but is not seen in the persistence forecast. These results indicate that the poor forecast skill over those west coast regions is likely caused by systematic errors in these models in simulating applicable western coastal weather processes.
- 3) Summer NMME forecast errors are spatially coherent along the California coastal region. Importantly, although forecast errors in the adjacent inland region are also spatially coherent, the correlation of forecast errors between the two regions is low,

FIG. 10. Time series of the surface temperature anomaly (standard deviation) for (a) San Francisco and (b) adjacent location over the ocean, in the 1983–2009 summer seasons (MJJ, JJA and JAS) for the NMME ensemble mean at lead time 1 month (red) and the observed temperature (blue). The red vertical bars are the spread (± 1 standard deviation) of the temperature anomalies in the models.

indicating that they are driven by different factors. Forecast errors over the coastal region are associated with the pattern of atmospheric circulation anomalies over lower middle latitudes and subtropics of the eastern Pacific, North America and western Atlantic. In contrast, forecast errors at inland locations are related to anomalous atmospheric circulation over the western United States.

- 4) For summer seasons (MJJ, JJA, and JAS), when NMME seasonal forecasts over the California coastal region are notably poor, the NMME forecast errors correlate significantly with variations in stratus clouds along the California coast. The temperature forecast error correlates positively (from +0.2 to +0.6) with local low cloudiness over the coastal and offshore areas during summer seasons. NMME forecasts tend to be warmer than observed when coastal low cloudiness is anomalously high over the California coast.

Although poor forecast performance of seasonal-mean surface temperature along the narrow west coastal

regions may be ignored in studies considering the global or North American domain, there are important regional consequences. Skillful forecasts of temperature at time leads from the subsequent month out to several months in advance would be very useful for applications involving energy, agriculture, and public health in highly populated coastal regions, such as California coast. The NMME temperature forecasts offer usable skill at time leads of one month and beyond for some regions, but their skill is notably poor along much of the California coast and some other west coastal regions globally.

Results in this study indicate that NMME's ability to forecast certain patterns of regional atmospheric circulation and its poor representation of coastal low cloudiness are key factors contributing to deficient warm season temperature forecast skill along the California coast. Locally, the error associated with coastal cloudiness is as important in affecting warm season forecast skill as other more widely recognized factors, such as the sea surface temperature, ENSO, and the Madden–Julian oscillation.

FIG. 11. Composite coastal low cloudiness anomaly of the 10 summer seasons having (a) the lowest forecast error and (b) the largest positive forecast error at lead time 1 month, at San Francisco during 1996–2009. Units are standard deviations indicated by colors and contours in percent.

Given similar skill deficits in other regions on the west coast of continents and the high occurrence of coastal low cloudiness over those regions in warm season, we hypothesize that the poor temperature forecast skill over the collective of west coast regions may be associated with problems in the models in representing low clouds along their off-shore and coastal regions.

Acknowledgments. This study was supported by the California Energy Commission under Agreement PIR-15-005. The authors thank Rachel Clemesha and Sam Iacobellis at the Scripps Institution of Oceanography, UC San Diego for providing the coastal low cloudiness data. We thank the climate modeling groups of NMME Phase 2 for producing and making available their model output. NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME Phase 2 system.

REFERENCES

- 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>.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Brunet, G., and Coauthors, 2010: Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bull. Amer. Meteor. Soc.*, **91**, 1397–1406, <https://doi.org/10.1175/2010BAMS3013.1>.
- Chen, L. C., H. van den Dool, 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>.
- Clemesha, R. E., A. Gershunov, S. F. Iacobellis, A. P. Williams, and D. R. Cayan, 2016: The northward march of summer low cloudiness along the California coast. *Geophys. Res. Lett.*, **43**, 1287–1295, <https://doi.org/10.1002/2015GL067081>.
- Diaconis, P., and B. Efron, 1983: Computer-intensive methods in statistics. *Sci. Amer.*, **248**, 116–131, <https://doi.org/10.1038/scientificamerican0583-116>.
- Frankignoul, C., 1985: Sea surface temperature anomalies, planetary waves, and air-sea feedback in the middle latitudes. *Rev. Geophys.*, **23**, 357–390, <https://doi.org/10.1029/RG023i004p00357>.
- Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152, <https://doi.org/10.1002/joc.636>.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, <https://doi.org/10.1002/joc.3711>.
- Hervieux, G., M. A. Alexander, C. A. Stock, M. G. Jacox, K. Pegion, E. Becker, F. Castruccio, and D. Tommasi, 2017:

- More reliable coastal SST forecasts from the North American multimodel ensemble. *Climate Dyn.*, <https://doi.org/10.1007/s00382-017-3652-7>, in press.
- Iacobellis, S. F., and D. R. Cayan, 2013: The variability of California summertime marine stratus: Impacts on surface air temperatures. *J. Geophys. Res. Atmos.*, **118**, 9105–9122, <https://doi.org/10.1002/jgrd.50652>.
- Infanti, J. M., and B. P. Kirtman, 2014: Southeastern U.S. rainfall prediction in the North American Multimodel Ensemble. *J. Hydrometeor.*, **15**, 529–550, <https://doi.org/10.1175/JHM-D-13-072.1>.
- , and —, 2016: Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America. *J. Geophys. Res. Atmos.*, **121**, 12 690–12 701, <https://doi.org/10.1002/2016JD024932>.
- Jacox, M. G., M. A. Alexander, C. A. Stock, and G. Hervieux, 2017: On the skill of seasonal sea surface temperature forecasts in the California Current System and its connection to ENSO variability. *Climate Dyn.*, <https://doi.org/10.1007/s00382-017-3608-y>, in press.
- 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>.
- Lawrence, D. M., K. W. Oleson, M. G. Flanner, C. G. Fletcher, P. J. Lawrence, S. Levis, S. C. Swenson, and G. B. Bonan, 2012: The CCSM4 land simulation, 1850–2005: Assessment of surface climate and new capabilities. *J. Climate*, **25**, 2240–2260, <https://doi.org/10.1175/JCLI-D-11-00103.1>.
- 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 Multimodel Ensemble. *J. Hydrometeor.*, **16**, 1409–1424, <https://doi.org/10.1175/JHM-D-14-0192.1>.
- National Academies of Sciences, Engineering, and Medicine, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. National Academies Press, 350 pp., <https://doi.org/10.17226/21873>.
- 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>.
- , —, C. R. Winkler, and J. S. Whitaker, 2003: A study of subseasonal predictability. *Mon. Wea. Rev.*, **131**, 1715–1732, <https://doi.org/10.1175/2558.1>.
- Pegion, K., and P. D. Sardeshmukh, 2011: Prospects for improving subseasonal predictions. *Mon. Wea. Rev.*, **139**, 3648–3666, <https://doi.org/10.1175/MWR-D-11-00004.1>.
- Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547–1564, <https://doi.org/10.1175/BAMS-84-11-1547>.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Rodrigues, L. R. L., J. García-Serrano, and F. Doblas-Reyes, 2014: Seasonal forecast quality of the West African monsoon rainfall regimes by multiple forecast systems. *J. Geophys. Res. Atmos.*, **119**, 7908–7930, <https://doi.org/10.1002/2013JD021316>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- , 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>.
- Schneider, T., J. Teixeira, C. S. Bretherton, F. Briant, K. G. Pressel, C. Schär, and A. P. Siebesma, 2017: Climate goals and computing the future of clouds. *Nat. Climate Change*, **7**, 3–5, <https://doi.org/10.1038/nclimate3190>.
- Shukla, S., M. Safeeq, A. AghaKouchak, K. Guan, and C. Funk, 2015: Temperature impacts on the water year 2014 drought in California. *Geophys. Res. Lett.*, **42**, 4384–4393, <https://doi.org/10.1002/2015GL063666>.
- Slater, L. J., G. Villarini, and A. A. Bradley, 2016: Evaluation of the skill of North-American Multi-Model Ensemble (NMME) global climate models in predicting average and extreme precipitation and temperature over the continental USA. *Climate Dyn.*, <https://doi.org/10.1007/s00382-016-3286-1>, in press.
- Stock, C. A., and Coauthors, 2015: Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Prog. Oceanogr.*, **137**, 219–236, <https://doi.org/10.1016/j.pcean.2015.06.007>.
- 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>.
- Vernieres, G., M. M. Rienecker, R. Kovach, and C. L. Keppenne, 2012: The GEOS-iODAS: Description and evaluation. NASA Tech. Rep. NASA/TM-2012-104606, Vol. 30, 73 pp., <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20140011278.pdf>.
- Wang, H., 2014: Evaluation of monthly precipitation forecasting skill of the National Multi-Model Ensemble in the summer season. *Hydrol. Processes*, **28**, 4472–4486, <https://doi.org/10.1002/hyp.9957>.
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- Winkler, C. R., M. Newman, and P. D. Sardeshmukh, 2001: A linear model of wintertime low-frequency variability. Part I: Formulation and forecast skill. *J. Climate*, **14**, 4474–4494, [https://doi.org/10.1175/1520-0442\(2001\)014<4474:ALMOWL>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<4474:ALMOWL>2.0.CO;2).