SUBSEASONAL TO SEASONAL TEMPERATURE PREDICITON SKILL OVER THE CALIFORNIA REGION FROM GLOBAL DYNAMICAL FORECASTS

A Report for:

California's Fourth Climate Change Assessment

Prepared By:
Zhenhai Zhang[1], Daniel R. Cayan[1], David W. Pierce[1]

1 Scripps Institution of Oceanography, UC San Diego

Edmund G. Brown, Jr., *Governor*

August 2018
CCCA4-CEC-2018-010

ACKNOWLEDGEMENTS

Support for Zhenhai Zhang, Daniel Cayan, and David Pierce was provided by the California Energy Commission under Agreement Number PIR-15-005. We thank Rachel Clemesha and Sam Iacobellis at the Scripps Institution of Oceanography for providing the coastal low cloudiness (low-altitude coastal cloudiness) data.

PREFACE

California's Climate Change Assessments provide a scientific foundation for understanding climate-related vulnerability at the local scale and informing resilience actions. These assessments contribute to the advancement of science-based policies, plans, and programs to promote effective climate leadership in California. In 2006, California released its First Climate Change Assessment, which shed light on the impacts of climate change on specific sectors in California and was instrumental in supporting the passage of the landmark Assembly Bill 32 (Núñez, Chapter 488, Statutes of 2006), California's Global Warming Solutions Act. The Second Assessment concluded that adaptation is a crucial complement to reducing greenhouse gas emissions (2009), given that some changes to the climate are ongoing and inevitable, motivating and informing California's first Climate Adaptation Strategy released the same year. In 2012, California's Third Climate Change Assessment made substantial progress in projecting local impacts of climate change, investigating consequences to human and natural systems, and exploring barriers to adaptation.

Under the leadership of Governor Edmund G. Brown, Jr., a trio of state agencies jointly managed and supported California's Fourth Climate Change Assessment: California's Natural Resources Agency (CNRA), the Governor's Office of Planning and Research (OPR), and the California Energy Commission (Energy Commission). The Climate Action Team Research Working Group, through which more than 20 state agencies coordinate climate-related research, served as the Steering Committee, providing input for a multi-sector call for proposals, participating in selection of research teams, and offering technical guidance throughout the process.

California's Fourth Climate Change Assessment (Fourth Assessment) advances actionable science that serves the growing needs of state and local-level decision-makers from a variety of sectors. It includes research to develop rigorous, comprehensive climate change scenarios at a scale suitable for illuminating regional vulnerabilities and localized adaptation strategies in California; datasets and tools that improve integration of observed and projected knowledge about climate change into decision-making; and recommendations and information to directly inform vulnerability assessments and adaptation strategies for California's energy sector, water resources and management, oceans and coasts, forests, wildfires, agriculture, biodiversity and habitat, and public health.

The Fourth Assessment includes 44 technical reports to advance the scientific foundation for understanding climate-related risks and resilience options, nine regional reports plus an oceans and coast report to outline climate risks and adaptation options, reports on tribal and indigenous issues as well as climate justice, and a comprehensive statewide summary report. All research contributing to the Fourth Assessment was peer-reviewed to ensure scientific rigor and relevance to practitioners and stakeholders.

For the full suite of Fourth Assessment research products, please visit www.climateassessment.ca.gov. This report contributes to our understanding of energy sector resilience by improving our understanding of subseasonal and seasonal forecasts of climate-related variables of importance for managing California's natural gas system.

ABSTRACT

Prior research regarding the use of probabilistic seasonal forecasts for the energy system suggests the potential to substantially improve management of natural gas supply. For example, assessment of peak seasonal natural gas demand could be improved with reliable probabilistic forecasts of high summer temperatures, when natural gas is used to help meet peak electricity demand, as well as cold winter temperatures, when natural gas is used for space heating. This study investigates subseasonal-to-seasonal forecasts of mean surface temperature over the California region in the North American Multimodel Ensemble (NMME) at lead times from 0 to 6 months. The seasonal forecast skill is assessed using the anomaly correlation coefficient (ACC) based on a large number of forecasts (1982-2009) from six NMME models (a total of 76 ensemble members). Overall, the NMME forecasts exhibit positive skill superior or similar to persistence forecasts (a forecast that the future weather condition will be similar to the present condition) over the ocean and interior regions of California. NMME forecasts generally have higher skill over the ocean than over the continent. The forecast skill over the land adjacent to the California coast is markedly lower than skill of forecasts over the nearby ocean and also lower than those over the inland continent, especially during the warm seasons. Summer forecast errors are spatially coherent over the coastal region. Similarly, forecast errors are spatially coherent in the inland regions. However, the correlation of forecast errors between the two regions is very low. Summer atmospheric circulation appears to have a strong influence on forecast performance—the time history of forecast errors at two inland locations was strongly correlated (-0.7) with the 500 hPa (hectopascal) geopotential height anomaly (height above the surface where atmospheric pressure is 500 hPa) over the western U.S. In contrast, forecast errors at coastal locations are associated with geopotential anomalies over the tropics and subtropical high pressure system (correlation 0.2-0.5). Furthermore, the systematic model forecast errors in coastal regions are linked to errors in the simulation of low cloudiness (low-altitude cloudiness, stratus clouds) over the coastal area. These results suggest that the California coastal region presents complicated interacting physical processes that are poorly captured by the current generation of seasonal prediction models. More effort is needed to improve dynamical forecasts of temperature and other climate variables in this region to provide better information regarding energy demand, which is particularly important as climate changes.

Keywords: seasonal forecast, surface temperature, coastal low cloudiness, North American Multimodel Ensemble, persistence forecast, California coast

Please use the following citation for this paper:

Zhenhai, Zhang, Daniel R. Cayan, David W. Pierce. (Scripps Institution of Oceanography). 2018. ***Subseasonal to Seasonal Temperature Prediction Skill over the California Region from Global Dynamical Forecasts.*** California's Fourth Climate Change Assessment, California Energy Commission. Publication Number: CCCA4-CEC-2018-010.

HIGHLIGHTS

- Reliable forecasts of temperature over a time frame ranging from about one month out to several months are of interest to energy, agriculture, and public health considerations over the highly-populated California coastal region. For example, a reliable forecast for an anomalously warm summer at a lead time of one to a few months will have practical applications for natural gas as well as electric utilities because temperature anomalies drive energy demand. Skillful temperature forecasts will also elucidate conditions related to drought in California since anomalous temperature plays an important role in anomalous evaporation and snowpack loss.
- Prior research suggests that improved adaptation to current levels of climate variability improves our ability to respond to a changing climate (including potential increases in climate variability).
- North American Multimodel Ensemble (NMME) is a dynamical climate forecasting system using state-of-the-art models from U.S. and Canadian modeling centers.
- Persistence is a forecast that the future weather condition will be similar as the present condition. Persistence forecasts provide a baseline reference to assess the model forecast skill; a forecast needs to beat persistence in order to be useful.
- NMME temperature forecasts at 1 to 6 months, for winter, spring, summer, and fall, exhibit positive skill that is superior or similar to persistence forecasts over the ocean and interior regions of California.
- NMME seasonal surface temperature forecast skill over the California coastal region is very low, often less than that from persistence forecasts, especially during the warm season.
- NMME forecast error over the inland eastern California is influenced by particular patterns of large-scale atmospheric circulation, especially during the warm season.
- NMME forecast error over the California coastal region is highly correlated to anomalous coastal low cloudiness (low-altitude coastal cloudiness) during the warm season; periods with dense cloudiness tend to exhibit greatest NMME forecast errors.
- Better model representation of key processes associated with coastal low cloudiness can improve poorly performing seasonal surface temperature forecasts from NMME in the coastal region of California, which is densely-populated and has high energy demands. More research specific to California is needed to improve the forecast skill over this region and provide an improved basis for anticipating energy demand on time scales of one month to several months in advance.

TABLE OF CONTENTS

1: Introduction

Sub-seasonal and seasonal forecasts are long-term forecasts that predict future weather conditions (i.e. temperature and precipitation) 2 weeks to 12 months out. In contrast to weather forecasts, which rely mainly on atmospheric behavior predicated from initial atmospheric conditions, sub-seasonal to seasonal forecast skill derives only 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. Sub-seasonal to seasonal forecasts are often cast in probabilistic terms, providing information about the likelihood of subsequent anomalous time average weather anomalies. Reliable forecasts at sub-seasonal 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 California.

There are many challenges to sub-seasonal to seasonal forecasting based on dynamical models. Some studies demonstrate that dynamical models are less skillful than much simpler statistical models, such as the linear inverse models (Winkler et al. 2001; Newman et al. 2003). In some recent cases, dynamical models have been found to out-perform linear inverse statistical models, specifically when an ensemble prediction methodology is employed (Pegion and Sardeshmukh 2011). Recent decades of development of dynamical prediction systems and research into seasonal climate predictability offers substantial evidence that sub-seasonal to seasonal forecast using dynamical models can be useful to the applications community (Kirtman et al. 2014; White et al. 2017). Progress in sub-seasonal to seasonal forecast has been enhanced by multi-institutional international collaborations (Kirtman et al. 2014). Primary examples of these collaborations include the North American Multimodel Ensemble (NMME; Kirtman et al., 2014), and the Sub-seasonal to Seasonal Prediction (S2S) Project proposed by the World Weather Research Programme and World Climate Research Programme (Vitart et al. 2017). The Working Group on Subseasonal to Interdecadal Prediction of the World Climate Research Programme aims to develop the numerical framework for predictions on timescales from subseasonal to interdecadal. Meanwhile, they also support three research projects advancing specific aspects of those forecasts: (1) how well climate forecast models represent global influences of tropical rainfall, (2) how snow predictably influences climate, and (3) how model drifts and biases develop and affect climate forecasts (Merryfield et al. 2017).

The North American Multimodel Ensemble (NMME) is a dynamical climate forecasting system using state-of-the-art coupled models from U.S. and Canadian modeling centers (Kirtman et al. 2014). It provides a multimodel framework for assessing the sub-seasonal to seasonal forecast skill in dynamical models. This multimodel forecast system demonstrates improved skill compared to individual models, such as the NOAA operational CFSv2, in forecasting large-scale climate features (Kirtman et al. 2014; Becker et al. 2014) since it benefits from a large number of ensemble members and model diversity. Many studies have evaluated the forecast skills of NMME for sea surface temperature, precipitation, and surface temperature at the global scale (Becker et al. 2014; Mo and Lyon 2015; Becker and van den Dool 2016) and over North America and the continental U.S. (Wang 2013; Infanti and Kirtman 2014; Chen et al. 2017; Slater et al. 2016; Hervieux et al. 2017). These studies showed that NMME forecast skill varies significantly from region to region, but only a few studies have focused on the California region. Jacox et al. (2017) investigated NMME forecast skill of seasonal sea surface temperature in the California Current System and found that NMME models were relatively skillful for the forecasts of February-April. Shukla et al. (2015) found that the NMME forecast skill of land surface temperature in California was generally low, with relatively higher correlation confined to interior regions in July-September. However, a more comprehensive evaluation of dynamical forecasts of seasonal averaged surface temperature over the California 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 is also unexplored.

In this study, a detailed evaluation of the performance of seasonal temperature predictions over the California region, which is urgently needed by decision makers, is conducted to improve our understanding of forecast skill at lead time of one to six months. The forecast skill of seasonal mean near-surface (2m) air temperature over the California region in NMME models is evaluated seasonally (March-May (MAM), June-August (JJA), September-November (SON), and December-February (DJF)), and compared to the persistence forecast skill. We find large and interesting differences in forecast skill over the eastern north Pacific Ocean, California coastal region, and inland eastern California. Associations between forecast errors and atmospheric circulation patterns, in the form of 500 hPa geopotential height anomaly, are examined. The influence of local coastal low cloudiness (low-altitude coastal 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 used ensembles of retrospective seasonal mean surface temperature forecasts (often called "hindcasts") of six models from the NMME phase 2 (Table 1). All the NMME models are coupled dynamical models and include key global climate fluctuations that affect California weather and climate, such as the 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 these natural climate fluctuations or their teleconnections can be poor, and may contribute to low model forecast skill in the region of interest.

All these models provide forecasts for lead times of at least 0-9 months. 28 years of hindcasts (i.e., retrospective forecasts over previous years, encompassing the period 1982-2009) are available for all models. Four models (CanCM3, CanCM4, CESM1, and GEOS5) have 10 ensemble members, FLORB01 has 12 members, and CFSv2 has 24 (28) members (only 24 members are used, 28 being available for November only). Hence a total of 76 ensemble members are used in this study. CFSv2 is initialized every fifth day, with four members per day. The ensemble members of the other five models are all initialized at 00 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 JJA 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". We calculate the seasonal mean near-surface air temperature from the daily values. Five of the six models provide outputs at 1°

latitude by 1° longitude, while FLORB01 provides outputs at 0.5° latitude by ~0.6° longitude. All data were interpolated onto a common 1°×1° 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 over land and the Hadley Center Sea Surface Temperature (HadISST) data (Rayner et al. 2003) over the ocean. The observational data are interpolated to the same 1°×1° grid as the NMME data. The daily land temperature dataset is time averaged to months and seasons for the forecast evaluations conducted here.

Persistence, the simplest forecast possible, is a forecast that persists the anomaly that exists at the time of the forecast. Persistence forecasts provide a baseline reference to assess the NMME model forecast skill; a forecast must beat persistence in order to be useful. 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, the averaged temperature anomaly of May and June is used as persistence forecast to JJA temperature anomaly at lead time 0 month. This provides a stringent test of the dynamical models' skill.

Table.1: The models in the North America Multimodel Ensemble (NMME) used in this study.

Model	Organization	Hindcast Period	Ensemble Size	Reference
CanCM3	CMC	1981-2012	10	Merryfield et al. 2013
CanCM4	CMC	1981-2012	10	Merryfield et al. 2013
CFSv2	NCEP	1982-2009	24(28)	Saha et al, 2014
CESM1	NCAR	1980-2010	10	Lawrence et al. 2012
FLORB01	GFDL	1980-2014	12	Vecchi et al. 2014
GEOS5	NASA	1982-2012	10	Vernieres et al. 2012

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 derived from a new high-resolution (4 km and half-hour) satellite-derived record of low clouds (Clemesha et al. 2016) from NASA/NOAA Geostationary Operational Environmental Satellite (GOES-9, GOES-10, GOES-11, and GOES-15). The data covers the ocean and west coast of U.S. (25°N to 50°N, 130°W to 113°W), over the period May-September 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. A number of different methods are used to evaluate forecast skill in this work. The anomaly correlation coefficient (ACC) 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. Additionally, contingency (conditional probability) tables are used to display the structure of model errors in categories, and so illustrate how frequently model forecasts are substantially different than what actually transpired. Finally, 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; these measures allow additional insight into the sources of model forecast error.

3: Results

3.1 NMME Forecasted Temperature Climatology and Biases

The climatological mean surface temperature from the ensemble mean of NMME is compared with the observations for the four seasons in Figure 1. The model's bias in seasonal mean surface temperature varies in different regions and different seasons. Overall, the model bias over the Pacific Ocean is much smaller than the bias over the continent (although note that seasonal variability over the ocean is smaller than over land areas as well). Most regions in California exhibit an NMME ensemble mean warm bias in summer (1 to 6°C) and cold bias (-1 to -4°C) in spring and winter. Meanwhile, a warm bias over the coastal region from Los Angeles to northern Baja California in Mexico is found in all four seasons, with a maximum of 2-6 °C in summer and a minimum of 1-3 °C in winter. This warm bias expands offshore and northwards in summer and autumn.

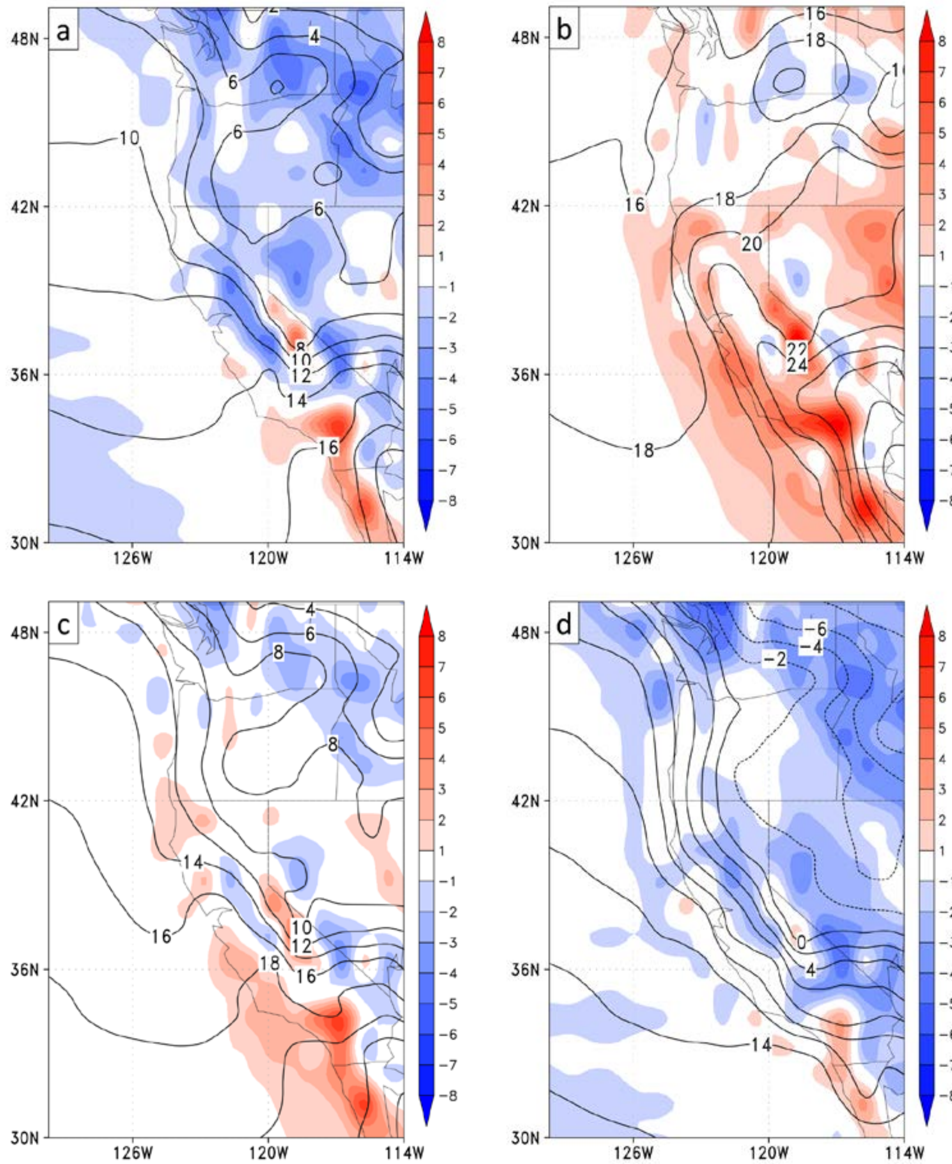

Figure 1. Bias of climatological mean surface temperature forecasts (model-minus-observations; °C) of NMME ensemble mean for the four seasons: (a) MAM (March-April-May), (b) JJA (June-July-August), (c) SON (September-October-November), and (d) DJF (December-January-February). The contours are climatological mean surface temperature (°C) in NMME ensemble mean.

These are the climatological biases in NMME models comparing to the observation. However, the seasonal temperature forecasts are used to identify the seasonal temperature anomalies. Hence, the climatological biases are not necessarily indicative of the NMME models' seasonal forecast skill in this study, because the ACC skill score gages the anomalies from the climatological mean of the forecasts. In particular, nothing prevents a model with mean biases from generating skillful forecasts; conversely, nothing guarantees that a model with small mean biases will generate skillful forecasts. However, these results demonstrate that bias correction is needed before using these temperature forecasts directly, for example, in a calculation that relates electrical load to actual temperatures. In the work performed here, we have circumvented the issue of model bias by focusing on skill measures that are affected little by the model biases, such as the ACC and anomaly composites.

3.2 Skill of Seasonal Mean Temperature Forecasts

Figure 2 shows the NMME seasonal mean surface temperature forecast skill (ACC) at a lead time of 1 month for each season. The colors depict regions where ACC statistical significance is equal or exceeds the 90% confidence level. A Bootstrap approach is utilized to conduct the statistical confidence level, using 1000 randomized seasonal temperature anomaly samples to estimate the probability that a given correlation threshold would be obtained simply by chance. Overall, forecast skill over the Pacific Ocean is higher than skill over the continent. Over the ocean, the ACC is about 0.3-0.8 and is relatively higher in MAM and DJF than in JJA and SON. Over the continent, which we include to put the California results into a wider-scale context, the forecast skill varies from 0 to 0.7 for different regions and different seasons. The NMME models have very high forecast skills over Washington state in spring (ACC = 0.5-0.7) and over Nevada in summer (ACC = 0.6-0.7). Meanwhile, there is little skill over the California coastal region in any season, with only modest skill (ACC < 0.4) at most in the northern California coastal region in MAM and DJF. In California, the forecast skill changes dramatically from the near-coast ocean to the adjacent land and then to the inland continent. This is particularly evident in JJA, when the ACCs offshore and over the inland continent are 0.4 and 0.6 respectively, while the ACC around the coast is near 0. This low skill exists in all seasons over the southern California coastal region, and in JJA and SON the area of low skill extends farther north.

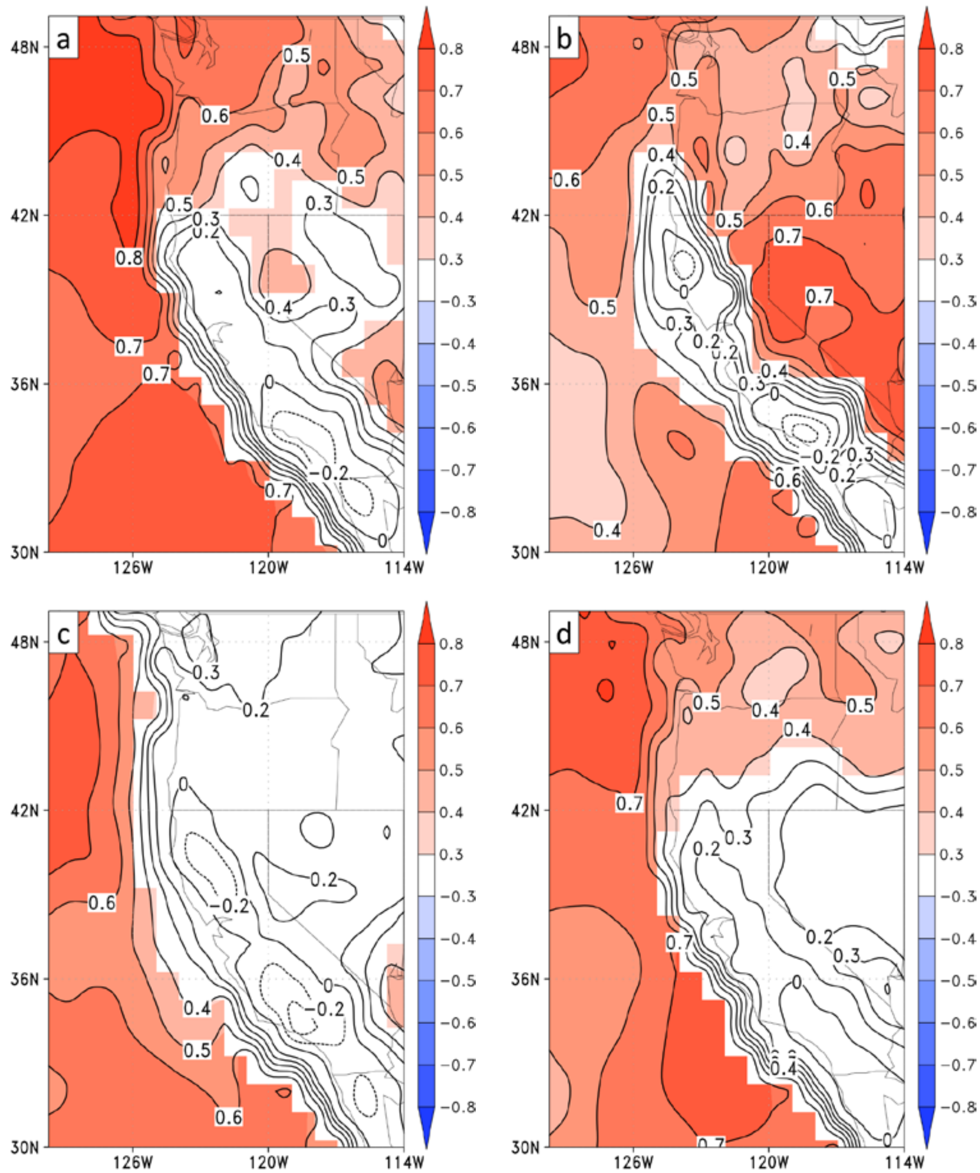

Figure 2. NMME ensemble mean surface temperature forecast skill measured by the anomaly correlation coefficient (ACC) for lead time 1 month, for (a) MAM, (b) JJA, (c) SON, and (d) DJF. Colors depict regions where ACC statistical significance is equal or exceeds the 90% confidence level.

The skill of the persistence forecast is compared with the skill of the NMME ensemble mean in Figure 3, for lead times of 0 (upper row) and 1 (lower row) month. Red colors indicate locations where the NMME skill significantly exceeds persistence; blue indicates the opposite. The persistence forecast has higher skill over the ocean than over the continent in all seasons and at different lead times, which is similar to the NMME ensemble mean. However, unlike the NMME results, the persistence forecast does not show near-zero skill over the California coastal land region. Thus, at both of lead time 0 and 1 month, NMME has lower skill than persistence over the California coastal region in JJA and SON, and over the southern part of this region in MAM and DJF (Figure 3). At lead time 0 month over most of the other regions, the NMME ensemble mean forecasts yield similar forecast skill (ACC difference < 0.1) in MAM, JJA, and

DJF. However, at lead time 1 month for these three seasons, NMME forecasts yield greater skill than persistence in some inland areas and the relatively higher latitudes. For SON, NMME 0-month lead forecast skill falls below persistence over almost all of California and northern Nevada, and NMME 1-month lead forecast skill is about equal to persistence over interior California and Nevada.

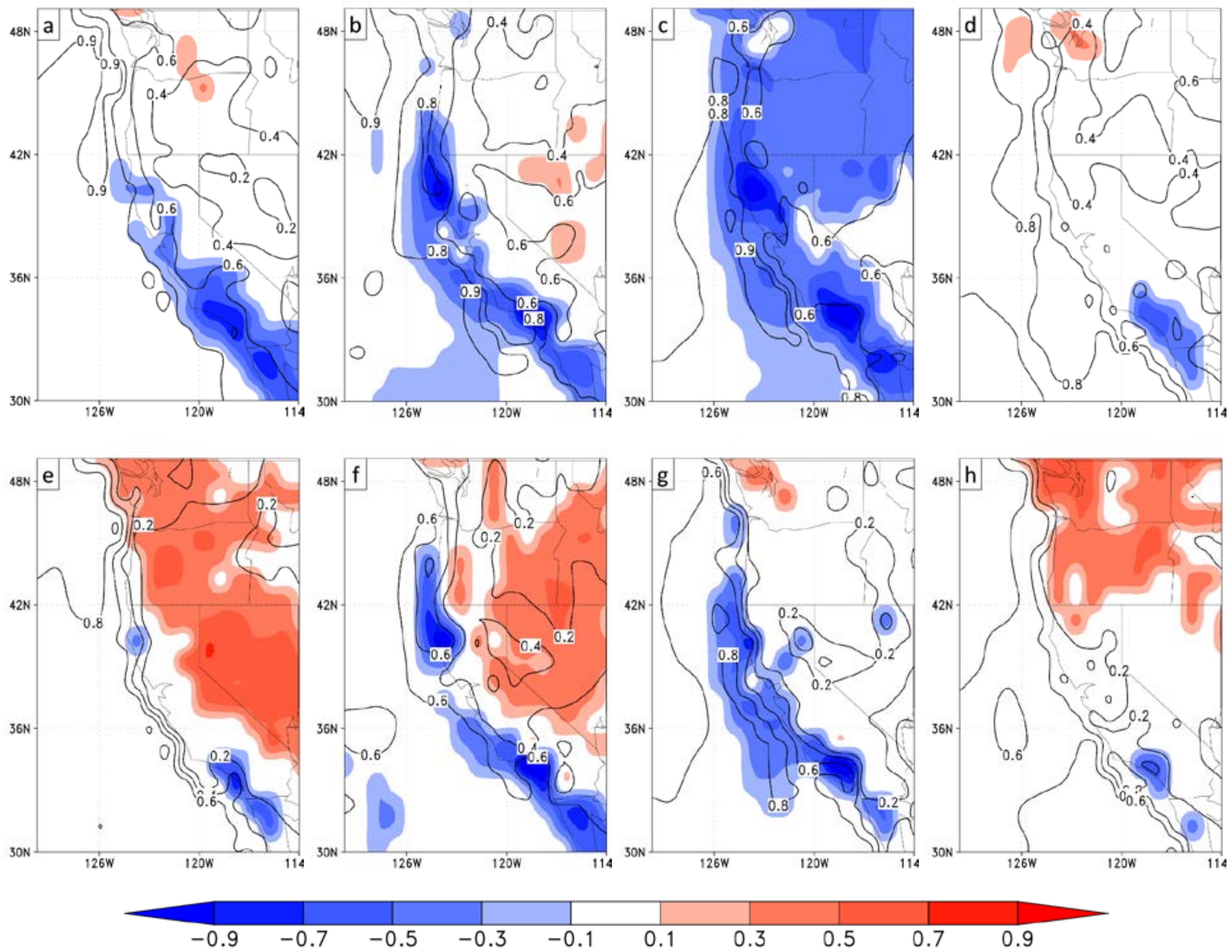

Figure 3. Difference between NMME temperature forecast skill and persistence forecast skill (NMME minus persistence, colors), showing differences at or exceeding the 90% significant confidence level. Forecast skill is the anomaly correlation coefficient (ACC) for lead time 0 month (upper) and lead time 1 month (lower) for (a, e) MAM, (b, f) JJA, (c, g) SON, and (d, h) DJF. Persistence forecast skill is shown by contours.

The previous results indicate that there is a large difference in the NMME forecast skills between the coastal and inland regions of California. To examine this further, we selected for analysis two locations in the California coastal region, near San Francisco and near San Diego, and two locations in the inland eastern California region, near Tahoe City and near Parker Dam (stars in Figure 4). Figure 5 shows the NMME ensemble mean forecast skill of seasonal mean surface temperature at these four locations for different seasons and lead times. There is no significant (at or exceeding the 90% confidence level) forecast skill at the two coastal locations. At San Diego, the ACC is close to 0 for all seasons and lead times. Meanwhile, the forecast skill at the two inland locations are much higher, especially in summer (~0.7 ACC at lead time 0 month), illustrating the difference of NMME forecast skill between the California coastal land region and the inland eastern California region. These results show that although there are relatively skillful seasonal forecasts for the inland eastern California region, the California coastal regions are poorly served by the seasonal temperature forecast.

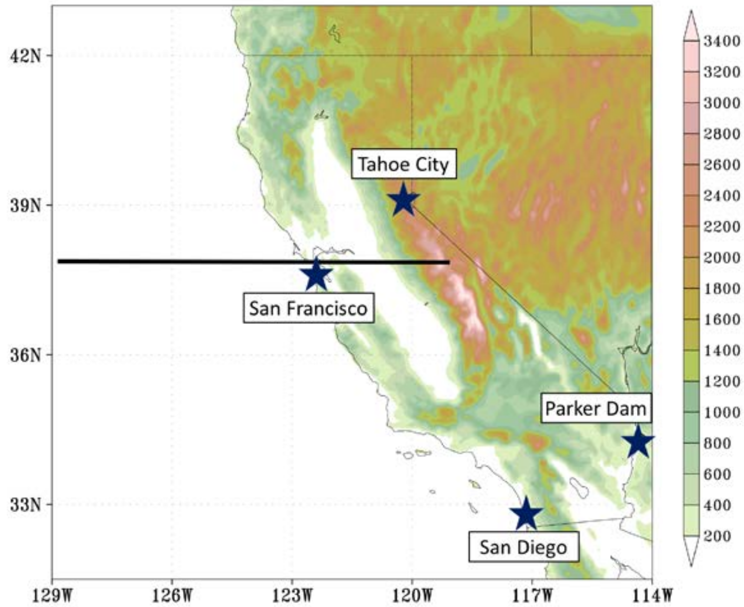

Figure 4. Four locations (stars) and 38°N transect (black line) for which forecast skill is evaluated (Figure 6, 7) in this study. Locations are San Francisco, Tahoe City, San Diego, and Parker Dam. Colors depict elevation (m).

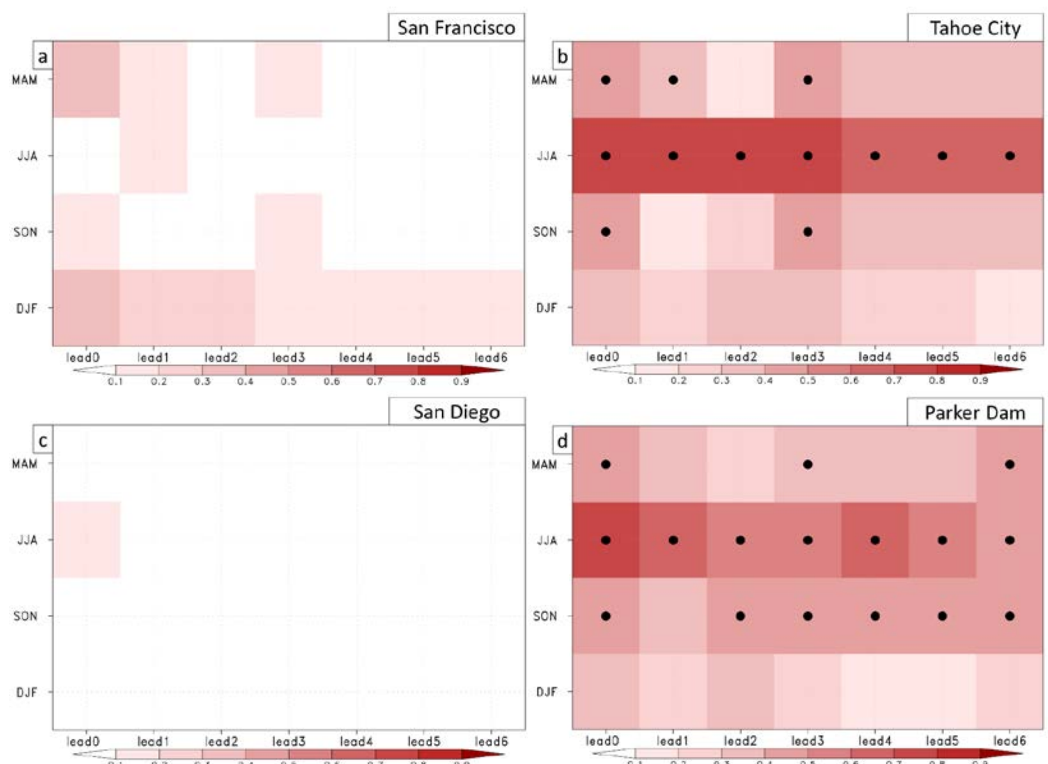

Figure 5. Forecast skill (ACC) of NMME ensemble mean surface temperature at grid cells near: (a) San Francisco, (b) Tahoe City, (c) San Diego, and (d) Parker Dam (locations shown in Figure 4). Each season shown on y-axis, and lead time, from 0 to 6 months, shown on x-axis. Black dots indicate forecasts whose ACC achieves statistical significance reaches or exceeds 90% confidence level.

To better illustrate the NMME forecasts of seasonal mean temperature for the four specific locations, a contingency table analysis is conducted for the four locations, presented in Table 2. Temperature anomalies for the 27 individual seasons (1983-2009) from NMME forecasts (1-month lead time) and the observation are separately categorized into three catalogs: warm (9 seasons with the warmest temperature anomaly), cold (9 seasons with the coldest temperature anomaly), and neutral (9 seasons with the temperature closest to the climatological mean). The contingency table shows the number of seasons in each combination of forecast and observation. For example, the top right panel in Table 2 shows the contingency table for JJA in Tahoe City: for the 9 observed warm JJA, the NMME ensemble mean forecasted 6 of the 9 correctly, 3 of the 9 as the neutral, and none of the 9 as the cold (opposite). The result of Parker Dam is similar. The strongly distinct result is consistent with the highly significant JJA ACC feature that includes these two locations in Figure 5, confirming that the NMME is able to provide practical implications in terms of informing decision or policy making for the California inland locations for summer. In San Francisco (top left panel in Table 2), the NMME forecasted 5 of the 9 observed cold DJF correctly and 2 of them as the warm (opposite), while the ACC is about 0.26. This implies that, even though the ACC is low, the NMME forecasts may have some potential implications for some specific locations and seasons, such as the winter in San Francisco.

Table.2 Contingency table for diagnosis of seasonal mean temperature anomaly in observation and NMME forecast for DJF in San Francisco (top left), JJA in Tahoe City (top right), DJF in San Diego (bottom left), and JJA in Parker Dam (bottom right)

San Francisco (DJF)		**observation**		
		warm	neutral	cold
NMME	warm	5	2	2
	neutral	2	5	2
	cold	2	2	5

Tahoe City (JJA)		**observation**		
		warm	neutral	cold
NMME	warm	6	3	0
	neutral	3	3	3
	cold	0	3	6

San Diego (DJF)		**observation**		
		warm	neutral	cold
NMME	warm	4	2	3
	neutral	1	5	3
	cold	4	2	3

Parker Dam (JJA)		**observation**		
		warm	neutral	cold
NMME	warm	6	3	0
	neutral	3	3	3
	cold	0	3	6

It is also interesting to examine the forecast skill of individual NMME models. Figure 6 shows the ACC at grid cells along the 38°N ocean-land transect shown as the heavy line in Figure 4, at lead time 1 month. The individual NMME models exhibit a large variety of forecast skills. In MAM, the spread of ACC across models is much smaller over the ocean than over the inland region. For example, over the continent at about 120°W, the ACC of the CFSv2 model is 0.50 while the ACC of the CanCM4 model is 0.02. In JJA, the ACC profiles of all six NMME models form a "V" shape from the ocean to inland continent, exhibiting high skill over the oceanic region, low skill over the immediate coastal land area, and higher skill over the interior land region. For example, in JJA from ocean (129°W) to the coastline (123°W) and then to the inland continent (119°W), the ACC of NMME ensemble mean changes from 0.51 to 0.08 to 0.73, and the ACC of CanCM4 changes from 0.42 to -0.15 to 0.66. In DJF, the NMME ensemble mean forecast skill is higher than the individual models over the oceanic region. However, in MAM, JJA, SON, and DJF over the continental region, 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 in MAM and DJF, while in JJA and SON the models have lower skill than persistence over the oceanic and coastal region. For the lower latitudes, forecast skills along similar transects exhibit similar results, wherein the "V" shaped ACC profile with high skill over the oceanic region, low skill over the coastal land, and high skill over the interior land also appears in MAM and SON in the models (not shown). These results indicate that the poor forecast skill over the California coastal region is likely caused by some systematic error in these models. More research is needed to understand this peculiar reduction in seasonal forecast skill over this densely-populated region.

The profile of forecast skills along the 38°N transect is also examined at different lead times for the NMME ensemble mean and the persistence forecast in Figure 7. The forecast skills at lead time 5 months is lower than the skills at lead time 1 month, but still have similar patterns. The NMME forecast skills around the coast at different lead times are very low (ACC < 0.2) in JJA and SON. The NMME ensemble mean exhibits similar or higher forecast skills than persistence at both of lead time 1 month and 5 months in DJF and over the interior land region in all seasons. Meanwhile, the NMME has similar or lower forecast skills than persistence over the coastal land region in MAM, JJA, and SON at different lead times.

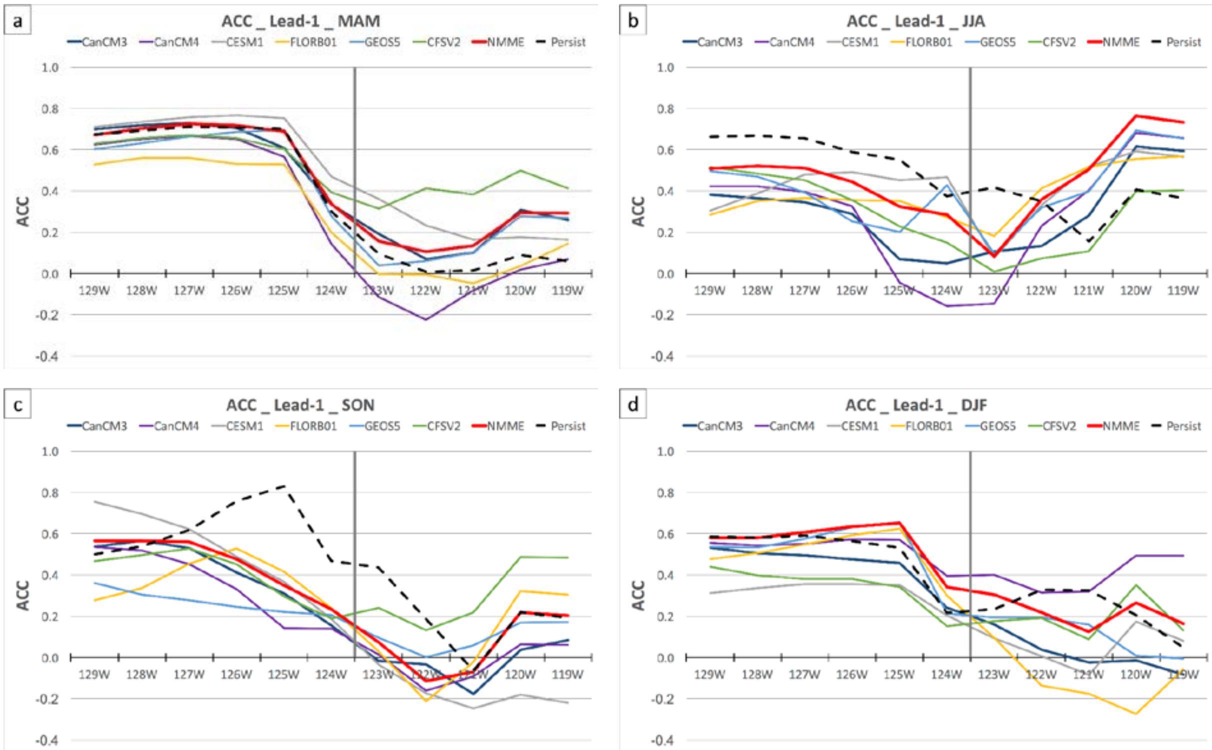

Figure 6. Forecast skill (ACC) of lead 1 month seasonal temperature forecasts along 38°N transect (black line in Figure 4) for the persistence, NMME ensemble mean, and the individual models for (a) MAM, (b) JJA, (c) SON, and (d) DJF. Vertical gray line is the position of the coast line.

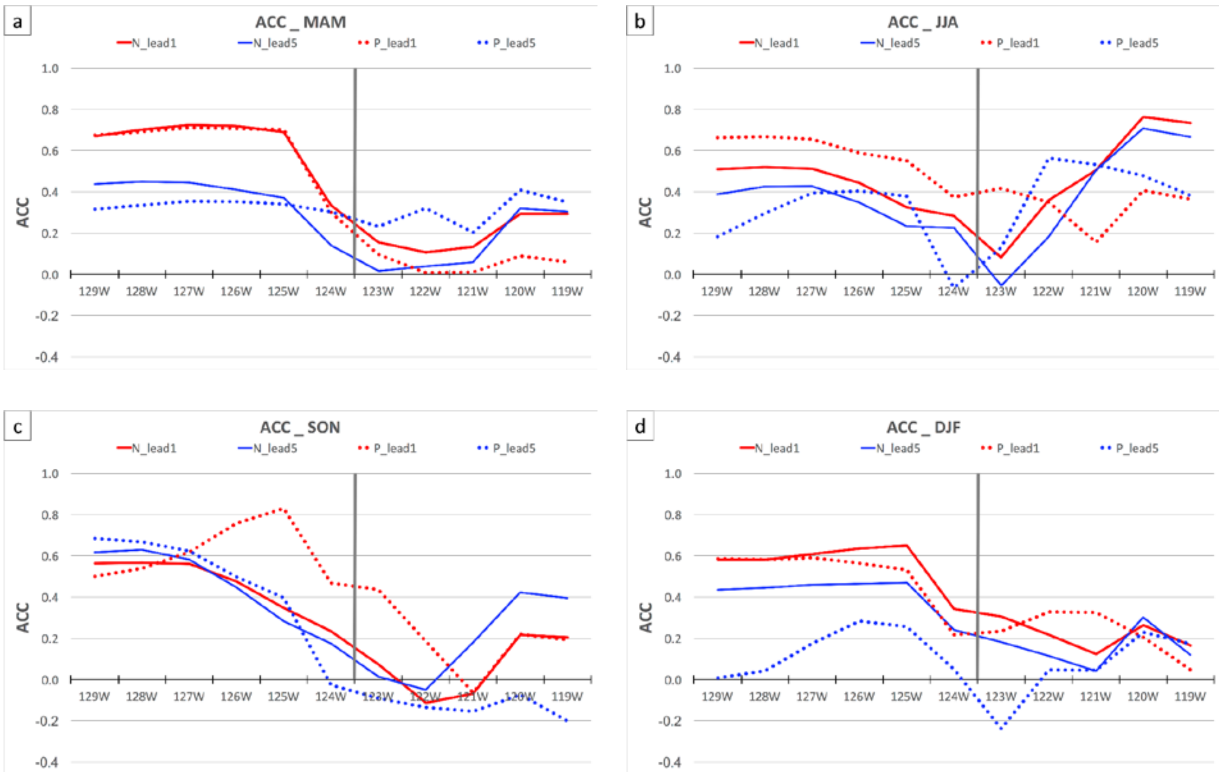

Figure 7. Same as Figure 6 but for the NMME ensemble mean (solid lines) and the persistence forecast (dotted lines) at lead time 1 (red) and 5 (blue) months.

3.3 Factors related to NMME Forecast Error

The results in Section 3.2 show that the NMME forecast skill of seasonal mean surface temperature is very low over the California coastal land region, in contrast to the higher skill exhibited in the west over the coastal ocean and east over the interior. This contrast is especially strong in summer months (JJA). In this section, the NMME forecast error and some related factors are examined to explore potential impacts of different factors that might affect seasonal temperature forecasts, with particular attention paid to providing an understanding of the poor forecast skill over the California coastal region. The model forecast error is defined as the NMME forecasted seasonal temperature anomaly minus the observational temperature anomaly in the overlapping summer seasons (MJJ, JJA, and JAS). These forecast errors vary in magnitude and sign (sometimes too warm, sometimes too cool) between years within the 28-year (1982-2009) time history of NMME forecasts.

To investigate the spatial coherence of forecast errors, the correlation between the time series of NMME forecast error at a specific location and that at all locations on the map at lead time 1 month are calculated for the two California coastal locations and the two eastern California inland locations (Figure 8). Note that a 95% confidence level is used here and in the subsequent two figures since they describe climatological connections between different regions, unlike the 90% confidence level used in previous figures describing forecast still. This is because current

GCMs do a better job reproducing the connections between related climate fields (temperature, cloudiness, geopotential height) than they do in producing skillful 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 (Figure 8). The pattern of correlations for San Diego is similar with that of San Francisco. 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 forecast errors, separate from those that cause forecast errors over the inland region.

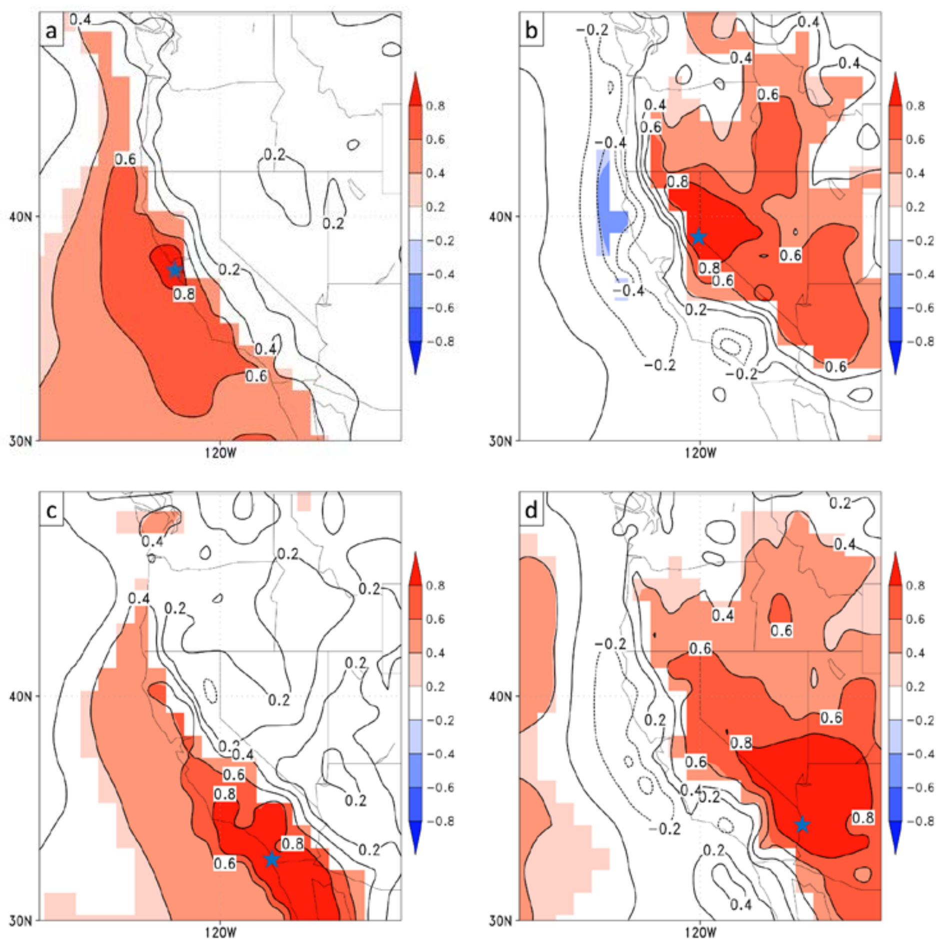

Figure 8. Correlation (contours) of NMME forecast error time series at a specific location and at all grid points on the map at lead time 1 month for three summer seasons (MJJ, JJA and JAS). Locations are (a) San Francisco, (b) Tahoe City, (c) San Diego, and (d) Parker Dam (star in each panel). Colors indicate locations whose correlation is significant at 95% confidence level or higher.

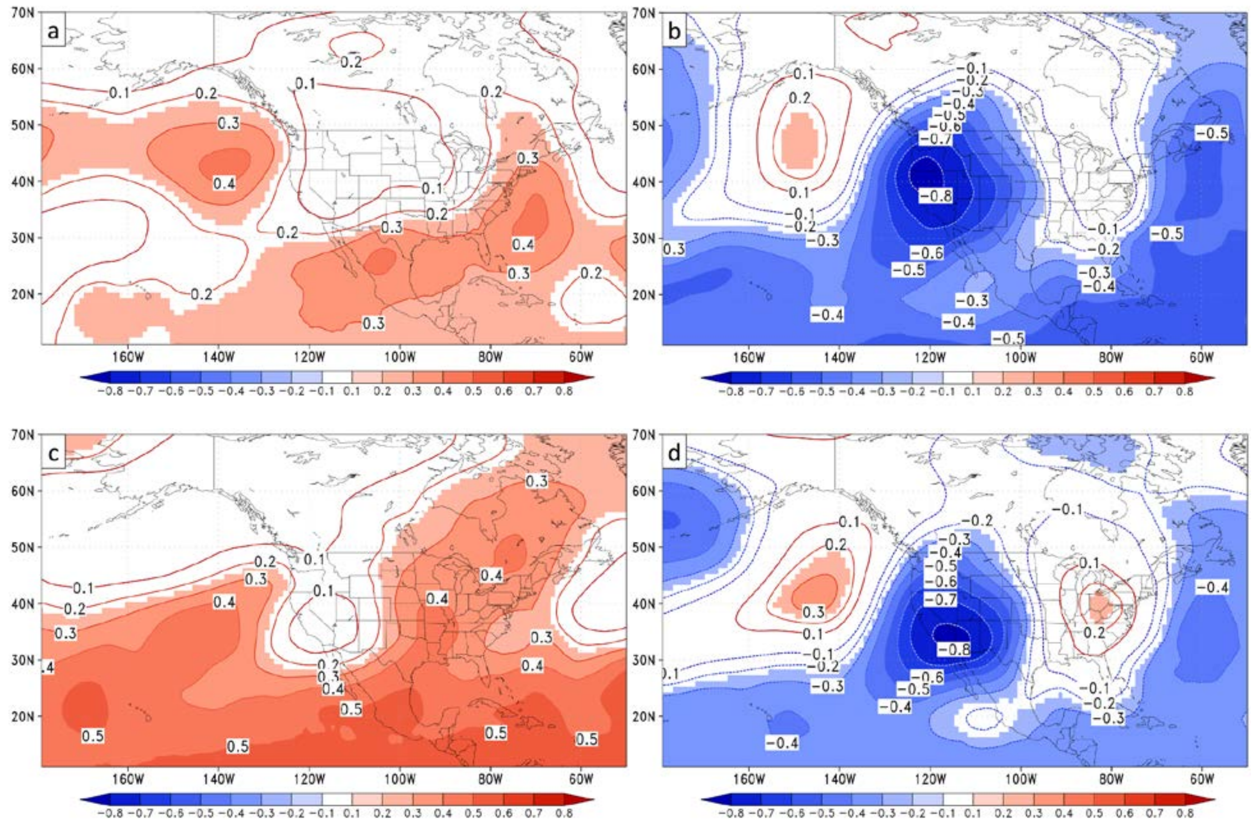

Figure 9. Correlations (contours) between the 500 hPa geopotential height anomaly time series at all grid points and the time series of NMME forecast error at lead time 1 month at (a) San Francisco, (b) Tahoe City, (c) San Diego, and (d) Parker Dam (stars in Figure 4) for three summer seasons (MJJ, JJA and JAS). Colors designate correlations significant at or above the 95% confidence level.

To investigate the associations of regional forecast error with larger scale atmospheric circulation, Figure 9 shows how forecast errors at the four individual locations of Figure 4 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 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. 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, which is not surprising since the forecast errors at the inland eastern California region are highly coherent. Meanwhile, 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 500hPa anomaly at the subtropical/tropical region and the eastern Pacific Ocean. 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 observations, and vice versa for a weakened subtropical ridge.

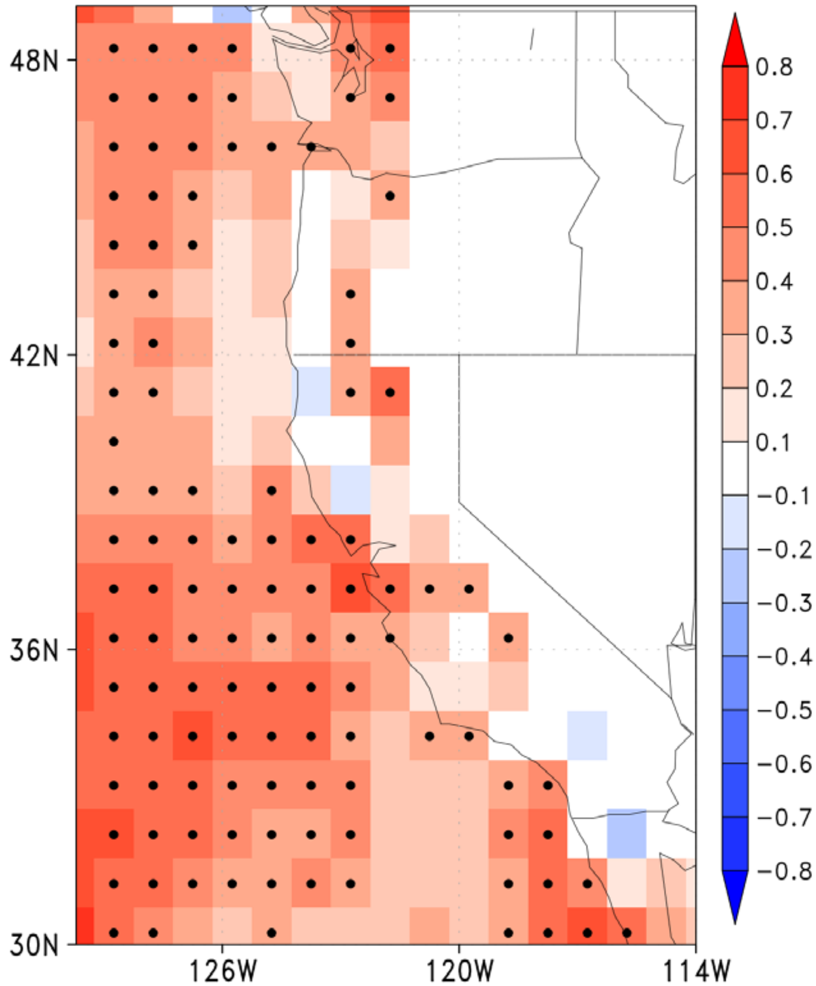

Figure 10. Correlation (colors) between coastal low cloudiness anomalies and the NMME forecast errors at lead time 1 month at each grid point in 1996-2009 summer seasons (MJJ, JJA and JAS). Black dots indicate correlations that are statistically significant at the 95% confidence level.

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 10 shows correlations between the field of 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 is available. Correlations consider the cloudiness 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 it is anomalous foggy and/or cloudy 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 (Iacobellis and Cayan, 2013), more 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 2° west of San Francisco) are extracted to examine the NMME forecast errors in more detail. At a lead time of one month, the difference between the temperature anomalies in the NMME and observations at San Francisco (land; Figure 11, upper) is much larger than the corresponding values over the ocean (Figure 11, lower). About 50% of these seasons have NMME anomalies with opposite signs from the observed temperature anomalies. Over land, the correlation between the NMME and observed temperature anomalies is only about 0.1. Meanwhile, over the adjacent ocean, the correlation between the NMME temperature anomalies and those of observations is much higher, about 0.7.

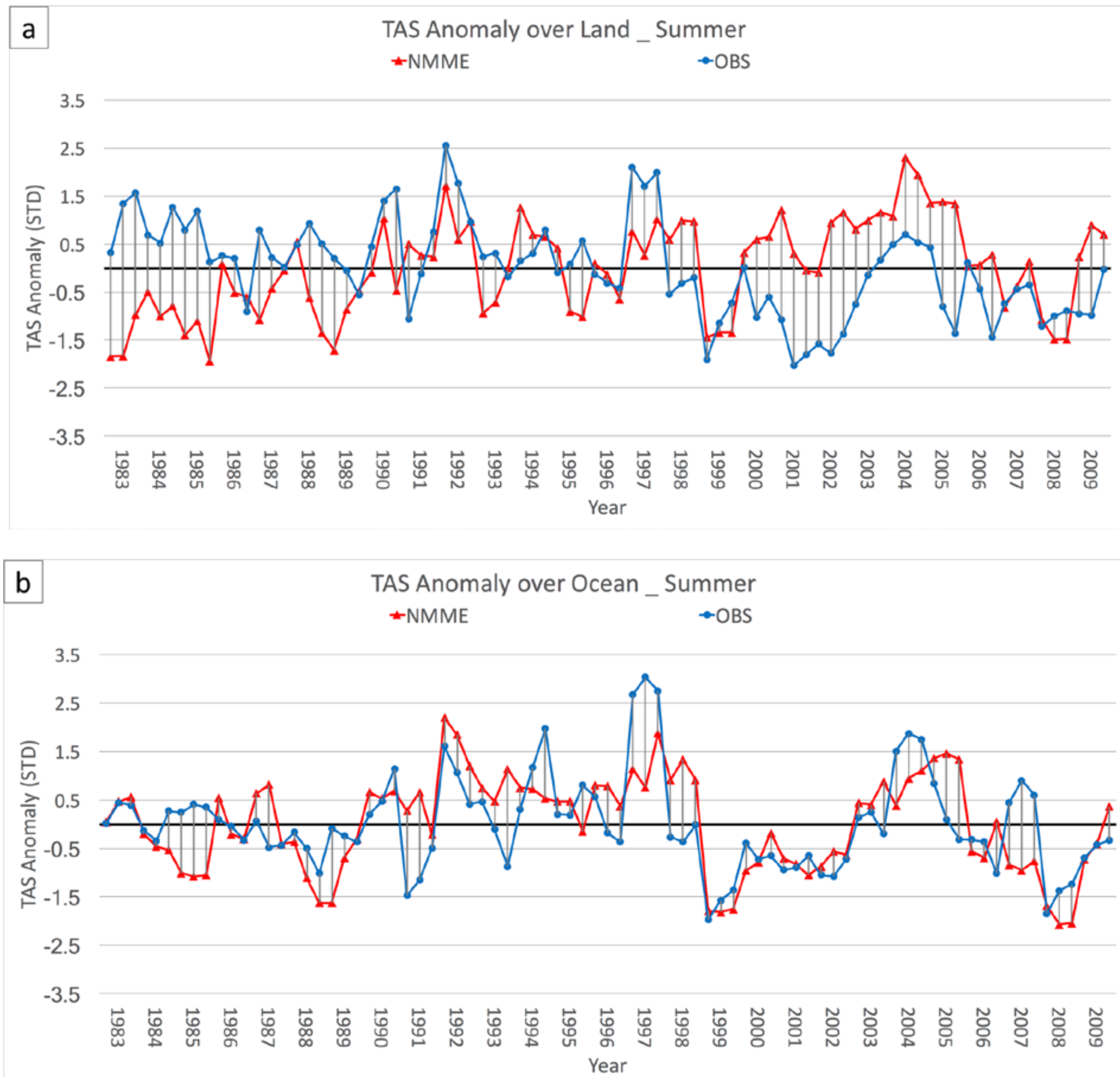

Figure 11. Time series of surface temperature anomaly (standard deviation) for (a) San Francisco and (b) adjacent location over the ocean, in 1983-2009 summer seasons (MJJ, JJA and JAS) for the NMME ensemble mean at lead time 1 month (red) and the observed temperature (blue).

Based on the forecast errors at San Francisco shown in Figure 11a, the ten seasons with the largest positive forecast errors (model forecast too warm) and ten seasons with the lowest forecast errors (anomalies in the model and observation very close) were selected for the period 1996-2009, when the cloudiness data are available. In the ten seasons with the largest positive errors, there are large positive cloudiness anomalies (0.5-0.6 standard deviations greater than the mean) around San Francisco in the composite cloudiness anomaly (Figure 12b). In comparison, for the ten seasons with the lowest errors at San Francisco, the coastal low cloudiness is very close to the climatological mean (Figure 12a). 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 cloudiness anomalies around this region (not shown). This indicates that compared to observations, the NMME models overestimate the surface temperature anomalies in May-September around San Francisco and San Diego during anomalously foggy/cloudy summers (and vice-versa).

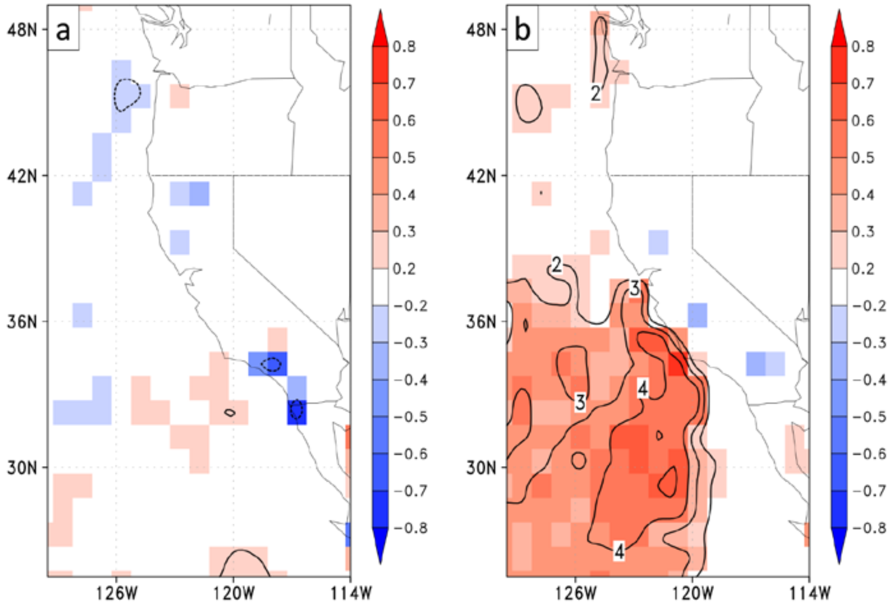

Figure 12. Composite coastal low cloudiness anomaly of the ten summer seasons having (a) the lowest forecast error and (b) the largest positive forecast error at lead time 1 month, at San Francisco during 1996-2009. Units are standard deviations indicated by colors, and contours in %.

4: Conclusions

Anticipating weather conditions with lead times of one to several months offers the prospect of improving natural gas system management through improved demand forecasts. This study investigates the skill of retrospective seasonal forecasts (sometimes termed "hindcasts") of near-surface air temperature over the California region using forecasts from six 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 in some regions for leads of 0 to 6 months, the results demonstrate a consistent pattern of low (near zero) forecast skill over California coastal land regions. To explore possible causes of this NMME forecast error, associations between NMME forecast errors and other atmospheric fields (500 hPa geopotential height and coastal low cloudiness) are examined. The major conclusions include:

1. At short time leads (forecast lead time 0 month) NMME ensemble mean forecast skill of seasonal average surface temperature is positive but close to or lower than the skill of persistence forecasts over the California region (Figure 3). At lead times 1 month out to 6 months, the NMME models produced positive skill that exceeds that of persistence.

2. Over California coastal land regions forecast skill in NMME is poor (ACC < 0.2) at all lead times, and markedly lower than the skill over the nearby offshore ocean and inland continent (Figures 2, 5, 6, and 7). This low skill area exists in the Southern California region

for NMME cool season (MAM and DJF) forecasts. For forecasts of the warm seasons (JJA and SON) the low skill over coastal land expands and extends northward along the coast. Yearly variations in NMME forecast skill correlate strongly between distant coastal locations (San Francisco and San Diego), but do not correlate well with NMME forecast skill over inland areas (Figure 8). Forecast skill at two separated California inland locations (Tahoe City and Parker Dam) also correlate highly. These results hold up for all seasons and lead time 0-6 months, confirming the distinct regimes of forecast performance that occurs in coastal vs. inland California land (Figure 5). Consistent with that, the contingency table analysis indicates that the NMME models produce skillful seasonal anomaly forecasts for some specific seasons and California inland locations that could inform practical applications, but there is little value in NMME forecasts over the California coastal region. Current NMME models have little improved forecasting ability over other methods for the California coastal region.

3. Forecast errors at two coastal locations are associated with large scale anomalies of geopotential height (500 hPa) over the tropics and subtropics from the southwest to the southeast of the California region, with modest positive correlations 0.2-0.5 (Figures 10 a and c). In contrast, forecast errors at two inland locations exhibit strong negative correlations (<- 0.8) with 500hPa anomalies over the western U.S. (Figure 9 b and d).
4. For summer seasons (MJJ, JJA, and JAS), where NMME seasonal forecasts over the California coastal land region are notably poor, the NMME forecast errors correlate significantly with variations in stratus clouds along the California coast. Forecast error correlates positively (+0.2 to +0.6) with local low cloudiness over the coastal and offshore areas during summer seasons (Figure 10). NMME forecasts tend to be warmer than observed when coastal low cloudiness has a strong positive anomaly over the California coastal region (Figure 12).
5. All the models available in the NMME ensemble exhibit the dip in forecast skill along the coast. All the models' cloudiness results indicate that the locally poor forecast skill is related to errors in simulating low cloudiness. Therefore, we find no evidence that the current set of models can be subsetted to eliminate poor quality models while retaining high-performing models in this region.

Although poor forecast performance of seasonal mean surface temperature along the narrow California coastal region may be ignored in the studies for the global or North America domain, there are important regional consequences, particularly since the population density in California is highest near the coast. Having skillful forecasts of temperature within the subsequent month out to several months in advance would be very useful for energy, agriculture, and public health considerations over the highly-populated California coastal region. For example, a reliable forecast for an anomalously warm summer at lead time one to a few months will have practical application for natural gas as well as electric utilities because temperature anomalies drive energy demand. Skillful temperature forecasts will also elucidate conditions related to drought in California since anomalous temperature plays an important role in anomalous evaporation and snowpack loss. NMME temperature forecasts are notably poor in coastal region, but there are some suggestions that these forecasts can be improved. One source for optimism is that NMME forecasts produce substantially higher skill at the nearby and offshore oceanic region and the adjacent interior region of California, along with better persistence forecasts along the coastal land area.

Future studies on temperature forecast skill errors could pursue multiple approaches. These include understanding the development of low (or high) forecast skill over the California coastal region on shorter time scales (5-day, 15-day, monthly), investigating the influence of coastal low cloudiness on the forecast skill during the earlier- and later- than the core summer season, and exploring other potential factors impacting sub-seasonal to seasonal temperature forecast skill in the California region.

Our results suggest that poor forecast of coastal low cloudiness is a key local factor contributing to poor seasonal temperature forecast skill along the California coast, and is as important as other more widely recognized factors in controlling seasonal forecast skill, such as sea surface temperature, ENSO, and the Madden-Julian Oscillation. More effort might be made to resolve processes associated with stratus clouds in order to improve the sub-seasonal to seasonal temperature forecast over the California coastal region. Improved understanding of these processes would contribute to improved forecasts of energy demand on times scales of one to several months in advance, as well as how those demands will be expected to change in light of climate change.

5: References

Becker, E., den Dool, H.V. and Zhang, Q., 2014: Predictability and forecast skill in NMME. *Journal of Climate*, *27*(15), pp.5891-5906.

Becker, E. and Van Den Dool, H., 2016: Probabilistic Seasonal forecasts in the North American multimodel ensemble: a baseline skill assessment. *Journal of Climate*, *29*(8), pp.3015-3026.

Clemesha, R.E., Gershunov, A., Iacobellis, S.F., Williams, A.P. and Cayan, D.R., 2016: The northward march of summer low cloudiness along the California coast. *Geophysical Research Letters*, *43*(3), pp.1287-1295.

Chen, L.C., Van den Dool, H., Becker, E. and Zhang, Q., 2017: ENSO precipitation and temperature forecasts in the North American Multimodel Ensemble: Composite analysis and validation. *Journal of Climate*, *30*(3), pp.1103-1125.

Harris, I.P.D.J., Jones, P.D., Osborn, T.J. and Lister, D.H., 2014: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International Journal of Climatology*, *34*(3), pp.623-642.

Hervieux, G., Alexander, M.A., Stock, C.A., Jacox, M.G., Pegion, K., Becker, E., Castruccio, F. and Tommasi, D., 2017: More reliable coastal SST forecasts from the North American multimodel ensemble. *Climate Dynamics*, pp.1-16.

Iacobellis, S.F. and Cayan, D.R., 2013: The variability of California summertime marine stratus: Impacts on surface air temperatures. *Journal of Geophysical Research: Atmospheres*, *118*(16), pp.9105-9122.

Infanti, J.M. and Kirtman, B.P., 2014: Southeastern US rainfall prediction in the North American multi-model ensemble. *Journal of Hydrometeorology*, *15*(2), pp.529-550.

Jacox, M.G., Alexander, M.A., Stock, C.A. and Hervieux, G., 2017: On the skill of seasonal sea surface temperature forecasts in the California Current System and its connection to ENSO variability. *Climate Dynamics*, pp.1-15.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter III, J.L., Paolino, D.A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M.P., Becker, E. and Peng, P., 2014: The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), pp.585-601.

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Fletcher, C.G., Lawrence, P.J., Levis, S., Swenson, S.C. and Bonan, G.B., 2012: The CCSM4 land simulation, 1850–2005: Assessment of surface climate and new capabilities. *Journal of Climate*, *25*(7), pp.2240-2260.

Merryfield, W.J., Lee, W.S., Boer, G.J., Kharin, V.V., Scinocca, J.F., Flato, G.M., Ajayamohan, R.S., Fyfe, J.C., Tang, Y. and Polavarapu, S., 2013: The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Monthly weather review*, *141*(8), pp.2910-2945.

Merryfield, W. J., F. J. Doblas-Reyes, L. Ferranti, J.-H. Jeong, Y. J. Orsolini, R. I. Saurral, A. A. Scaife, M. A. Tolstykh, and M. Rixen 2017: Advancing climate forecasting, *Eos, 98,* https://doi.org/10.1029/2017EO086891. Published on 27 November 2017.

Mo, K.C. and Lyon, B., 2015: Global meteorological drought prediction using the North American multi-model ensemble. *journal of Hydrometeorology*, *16*(3), pp.1409-1424.

Newman, M., Sardeshmukh, P.D., Winkler, C.R. and Whitaker, J.S., 2003: A study of subseasonal predictability. *Monthly weather review*, *131*(8).

Pegion, K. and Sardeshmukh, P.D., 2011: Prospects for improving subseasonal predictions. *Monthly Weather Review*, *139*(11), pp.3648-3666.

Randall, D., Khairoutdinov, M., Arakawa, A. and Grabowski, W., 2003: Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), pp.1547-1564.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C. and Kaplan, A., 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(D14).

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.T., Chuang, H.Y., Iredell, M. and Ek, M., 2014: The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), pp.2185-2208.

Saha, S., Moorthi, S., Pan, H.L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D. and Liu, H., 2010: The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), pp.1015-1058.

Schneider, T., Teixeira, J., Bretherton, C.S., Brient, F., Pressel, K.G., Schär, C. and Siebesma, A.P., 2017: Climate goals and computing the future of clouds. *Nature Climate Change*, *7*(1), p.3.

Shukla, S., Safeeq, M., AghaKouchak, A., Guan, K. and Funk, C., 2015: Temperature impacts on the water year 2014 drought in California. *Geophysical Research Letters*, *42*(11), pp.4384-4393.

Slater, L.J., Villarini, G. and Bradley, A.A., 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 Dynamics*, pp.1-16.

Vecchi, G.A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A.T., Zeng, F., Anderson, W., Balaji, V., Dixon, K. and Jia, L., 2014: On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, *27*(21), pp.7994-8016.

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, 61.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M. and Hendon, H., 2017: The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), pp.163-173.

Wang, H., 2014: Evaluation of monthly precipitation forecasting skill of the National Multi-Model Ensemble in the summer season. *Hydrological processes*, *28*(15), pp.4472-4486.

White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J., Lazo, J.K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A.J., Murray, V. and Bharwani, S., 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*.

Winkler, C.R., Newman, M. and Sardeshmukh, P.D., 2001: A linear model of wintertime low-frequency variability. Part I: Formulation and forecast skill. *Journal of climate*, *14*(24), pp.4474-4494.