Understanding Skill of Seasonal Mean Precipitation Prediction over California during Boreal Winter and Role of Predictability Limits

ARUN KUMAR AND MINGYUE CHEN

Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland

(Manuscript received 11 April 2019, in final form 10 April 2020)

ABSTRACT

Using extensive hindcasts from seasonal prediction systems participating in the North American Multi-Model Ensemble (NMME), possible causes for low skill in predicting seasonal mean precipitation over California during December–February (DJF) are investigated. The analysis focuses on investigating two possibilities for low prediction skill: role model biases or inherent predictability limits. The motivation for the analysis was the seasonal prediction during DJF 2015/16 that called for enhanced probability for above normal precipitation over southern California (which was consistent with expected conditions during an extreme El Niño) while the observed precipitation was below normal. Based on various analysis approaches and using hindcast datasets from multiple seasonal prediction systems, we build up the evidence that low skill in predicting seasonal mean precipitation over California is likely to be due to inherent predictability associated with a low signal-to-noise (SNR) regime. For the same set of seasonal prediction systems, the precipitation variability over California is contrasted with that over the southeast United States where prediction skill, as well as the SNR, is higher. The discussion also notes that building a knowledge base that goes beyond the well-known response to ENSO (based on the linear regression or composite techniques) has proven to be difficult and a systematic approach to reaching resolution to some of the overarching questions is required, and toward that end, a pathway is suggested.

1. Introduction

Boreal winter of 2015/16 saw the emergence of an extreme El Niño event in the equatorial Pacific. The event was well anticipated from the prediction by statistical and dynamical coupled models (L’Heureux et al. 2017). The emergence of the extreme El Niño event coincided with the prolonged drought over California (Seager et al. 2015; Swain et al. 2014; Wang et al. 2014). Conforming to the current knowledge base for precipitation signals associated with strong El Niño conditions to be above normal over the region (Ropelewski and Halpert 1986; Hoerling and Kumar 1997; Kumar and Hoerling 1997, 1998; Trenberth et al. 1998; Chen and Kumar 2015), anticipation for an El Niño event generated widespread expectations that it would lead to above normal precipitation and relief from drought conditions, particularly over southern California. Indeed, the official seasonal prediction from the Climate Prediction Center (CPC) for seasonal mean of December–February

(DJF) 2015/16 indicated a probability for above normal precipitation over this region (Fig. 1). The CPC forecast was consistent with the mean El Niño response as inferred based on the regression or composite analysis using observational data and model simulations, and further, the dynamical tools [e.g., the Climate Forecast System version 2 (CFSv2) at the National Centers for Environmental Prediction (NCEP)] also predicted similar anomalies (Fig. 1).

In contrast to the expectations generated by the seasonal prediction, the observed seasonal mean precipitation anomalies over California turned out to be below normal and positive precipitation anomalies occurred only over the northern parts of the west coast of the United States (Fig. 1). Although given the probabilistic nature of seasonal mean prediction [e.g., CPC prediction for 60% probability that seasonal mean precipitation over southern California would be above normal (Fig. 1)], a forecast verification based on a single instance of opposite sign observed anomaly cannot be considered as a failure per se, nonetheless the general perception of the user community for the performance of seasonal prediction tilted toward that sentiment.

Corresponding author: Dr. Mingyue Chen, mingyue.chen@noaa.gov

FIG. 1. (top left) Official Climate Prediction Center seasonal precipitation outlook for DJF 2015/16 released November 2015. Green (brown) regions indicate probability for seasonal mean precipitation to be in above (below) normal category. Contours indicate the forecast probability for seasonal mean precipitation to be in the respective category. Regions in white are where forecast tools do not have a clear indication for seasonal mean precipitation to be in a preferred category. (bottom left) NMME forecast for DJF 2015/16 (mm day^{-1}). (top right) Regression of the observed DJF precipitation with the observed Niño-3.4 SST index over 1950–2016 (mm day^{-1}); regression is computed over the 1950–2016 period. (middle right) Composite precipitation response based on the seasonal prediction systems in the NMME database (mm day^{-1}). (bottom right) Observed DJF 2015/16 precipitation (mm day^{-1}).

Subsequently, a similar perception has also been shared by the scientific community, which has strived to explain why the observed seasonal mean anomalies may have differed from the expected El Niño response.

A common approach for understanding differences in the seasonal forecasts and the observed precipitation outcome is 1) to quantify how different the El Niño response for the DJF 2015/16 may have been from the commonly, and now well recognized and widely internalized, response to El Niño, and 2) to determine if the departure in the observed pattern from the forecasts also followed differences. In this reasoning, the differences in response may have been due to differences in the sea surface temperature (SST) pattern for the 2015/16 El Niño from a composite El Niño in that the SST anomalies for 2015/16 El Niño were stronger in the

central Pacific—a scenario commonly referred to as the central Pacific (CP) El Niño (Yu and Kim 2010; Paek et al. 2017). Differences in the atmospheric response could also have been due to the influence of some other anomalous boundary conditions. Such possibilities for DJF 2015/16 include warmer SST anomalies in the extratropical Pacific (referred to as the warm blob; Bond et al. 2015; Hu et al. 2017), a general warming of SSTs in the Indian Ocean and western Pacific, drier land conditions over California (as a result of prolonged drought; Yang et al. 2018), and changes in sea ice over the Arctic due to decline of sea ice in past decades (Cohen et al. 2017). Differences in atmospheric and terrestrial response to El Niño conditions in 2015/16 could also have been due to the nonlinearity in the atmospheric response to the strength of SST anomalies that cannot be

captured either by regression or composite-based analysis approaches.

The attempts of the scientific community to understand the unique aspects of atmospheric and terrestrial response during 2015/16 El Niño highlight the conundrum that even after an extensive history of research in understanding and analyzing the atmospheric response to ENSO following the seminal analysis of Horel and Wallace (1981), possible variations in ENSO response beyond the mean response are still not well understood (or agreed upon). Examples of such studies include atmospheric response to central versus eastern Pacific El Niño events (Yu and Kim 2010; Hu et al. 2012; Patricola et al. 2020), changes in the location of equatorial convection associated with El Niño (Palmer and Mansfield 1986; Larkin and Harrison 2005; Barsugli and Sardeshmukh 2002; Chiodi and Harrison 2015; Johnson and Kosaka 2016), and the influence of SST anomalies in other ocean basins (Robinson 2000; Kushnir et al. 2002; Quan et al. 2006; Schubert et al. 2009; Bond et al. 2015). A lack of consensus about the atmospheric response to ENSO and other SST anomalies is reflected in that forecasters still heavily rely on the ENSO composites. In other words, despite extensive efforts, answers to the question of how the atmospheric response may vary from one ENSO event to another have not been conclusive; further, such differences have not been internalized by the operational seasonal forecasting community that continues to rely on ENSO composites in the framing of their forecasts.

For the current generation of dynamical seasonal forecast systems for which an extensive set of hindcasts is generally available, the performance of the precipitation forecast during the 2015/16 event can be evaluated in a larger framework response during other El Niño events. Considering the specific example of CFSv2, seasonal hindcast is available from 1982. An analysis of skill of seasonal mean California precipitation for CFSv2 over the entire forecast history indicates that the average forecast skill is low; for example, the anomaly correlation over California is only about 0.3 (Fig. 2). Based on this measure alone, the perceived notion of failure of the seasonal forecast should not have been a surprise. This argument, however, can be criticized from at least two perspectives discussed below.

In the context of ENSO variability, the average skill assessment computed over the entire forecast history is the skill of the seasonal prediction system over all ENSO conditions including neutral, weak, moderate, and strong ENSO years. Such an assessment of forecast skill is often referred to as the unconditional skill (Kumar 2009). To the first order, a linear dependence of atmospheric signal on the strength in the amplitude of ENSO SST anomalies

FIG. 2. The skill based on anomaly correlation of DJF precipitation for the forecast ensemble means from CFSv2 hindcasts in 1982–2010. The two blue boxes are the areas defined as the U.S. west coast (wCoast, left side) and the U.S. southeast coast (seCoast, right side).

(Hoerling and Kumar 1997), and the theoretical relationship between the signal-to-noise ratio (SNR) and skill measures (Kumar 2009) will strongly suggest that skill during strong El Niño events (generally referred to as the conditional skill; Kumar 2009) should be larger than the estimate of unconditional skill. The concept of conditional versus unconditional skill, and differences among them, although intuitively appealing, cannot be quantified because of the shortness of the observational record that only contains few strong El Niño events; this is so because for small samples, skill estimates can have large uncertainty (Kumar 2009; Pegion and Kumar 2013).

The second line of criticism for the low skill in the CFSv2 is that it is influenced by the model bias that may compromise the realization of (possibly higher) predictability inherent in nature. This is a valid argument in that issues like inference about skill estimates, responses to external forcings, and so on, if based on a single model, generally cannot be trusted. Reducing reliance on the analysis based on a single model has led to major internationally coordinated efforts involving multimodel approaches. Examples include estimating climate sensitivity to changes in external forcings as part of the Coupled Modeling Intercomparison Project (CMIP) (Eyring et al. 2016), seasonal predictions based on multimodels (Graham et al. 2011; Kirtman et al. 2014), subseasonal predictions based on a multimodel (Vitart et al. 2016), and improvement in weather predictions (Swinbank et al. 2016).

With the multimodel seasonal prediction datasets available at present, specifically the North American Multi-Model Ensemble (NMME) (Kirtman et al. 2014), we can address both the criticisms discussed above. The availability of a large ensemble may also allow us to

investigate the question of how sensitive the atmospheric response is to different flavors of ENSO.

Using the NMME dataset we explore these questions: What is the skill for predicting seasonal mean precipitation over California across various prediction systems? For the same set of prediction systems, how does it differ from other regions that are known to be influenced by ENSO, such as the southeastern United States (Ropelwski and Halpert 1986)? The dataset also gives us the means to address additional questions: Can we infer a difference in the response of CA rainfall to the flavors of ENSO SST anomalies? What is the nonlinearity in the atmospheric response to ENSO? We note that exploration of such factors has already been brought up in the context of 2015/16 California rainfall and in various investigations that have appeared in the literature (e.g., Hoell et al. 2016; Cohen et al. 2017; Paek et al. 2017; Chen and Kumar 2018; Jong et al. 2018; Singh et al. 2018; Zhang et al. 2018).

We approach the analysis with the possibility that even an extensive hindcast database like the NMME may not be sufficient for answering the above questions with confidence, and if so, we discuss what are the implications for the predictability of California seasonal mean precipitation. For example, is our lack of ability to provide answers using large databases a portent that the inherent SNR for California is, indeed, low? We also discuss what further could be done to alleviate the current situation we find ourselves in; that is, our knowledge base beyond the composite response to ENSO does not seem to exist with a degree of confidence that could be used to develop and communicate operational forecasts.

Using the same NMME database we also contrast the analysis over California with that over the southeastern United States where models have higher skill (Fig. 2). This regional contrast in levels of prediction skill using the same set of models raises a fundamental question about the general fidelity of models and our perceptions about their utility in addressing some basic questions associated with assessment of predictability.

2. Data and analysis approach

a. The data

We focus our analysis on the seasonal precipitation variability 1) averaged over all years, and 2) in all individual El Niño years for the target season of December–February (DJF).

The model data used are forecast data from the North American Multi-Model Ensemble (NMME) project (Kirtman et al. 2014; Becker et al. 2014). The forecast data include ensemble forecasts from seven participating

TABLE 1. The list of ensemble size and initial conditions for the seven models used in the analysis.

Model	Ensemble size	Initial conditions
CFSv2	24	0000, 0600, 1200, and 1800 UTC 18, 23, and 28 Oct and 2, 7, and 12 Nov
A (Environment Canada CMC1)	10	0000 UTC 1 Nov
B (Environment Canada CMC2)	10	0000 UTC 1 Nov
C (NCAR CCSM4)	10	0000 UTC 1 Nov
D (GFDL CM2.1)	10	0000 UTC 1 Nov
E (GFDL FLOR)	24	0000 UTC 1 Nov
F (NASA GEOS5)	11	Four members from 28 Oct and seven members on the last day of October

models with retrospective forecasts (also referred to as hindcasts) for 1982–2010 and real-time forecasts over 2011–16. The seven models are the CFSv2 and other six models, which for brevity are referred to as models A–F; mapping of models A–F to their corresponding seasonal forecast system is given in Table 1. The forecast members for the models A–E are initialized at 0000 UTC on the first day of each month for both the hindcasts and real-time forecast. For model F, four members are initialized every fifth day and seven members on the last day of each month. CFSv2 has four forecast members initialized at 0000, 0600, 1200, and 1800 UTC every fifth day starting at 1 January each year for hindcasts and four members every day for real-time forecasts. For the analysis, we subsample the CFSv2 real-time forecast members to match the frequency of its hindcast configuration and use the members from the initial conditions, which are close to the most other NMME models. Following this, the ensemble forecasts for the DJF season used in the analysis are the 24 members initialized on 18, 23, and 28 October, and 2, 7, and 12 November for CFSv2, and all members available in the NMME project for other six models (except for model F; see Table 1). The analysis period covers forecasts from 1982 to 2016. The ensemble size and initial conditions for each model are also listed in Table 1. More detailed information about the models can be found in Kirtman et al. (2014).

In addition to the forecast data from the NMME seasonal forecast systems, the Atmospheric Model Intercomparison Project (AMIP)-type simulation data from four models are also used in the analysis. The setup of AMIP simulation is the atmospheric general circulation model (AGCM) simulation forced by the observed evolution of SSTs. The four models include the atmospheric component of the CFSv2, the National Center for Atmospheric Research (NCAR) Community

Atmosphere Model 4 (CAM4), the Max Planck Institute for Meteorology ECHAM5, and the NOAA Earth System Research Laboratory (ESRL)–NCAR CAM5. The CFSv2 AMIP simulation is from the atmospheric component of CFSv2 and has 55 ensemble members. The simulations for the other three models are obtained from the ESRL (<https://www.esrl.noaa.gov/psd/repository/facts>). There are 20, 50, and 40 ensemble members for the CAM4, ECHAM5, and ESRL-CAM5 simulations, respectively. For consistency, we use the common period of 1982–2014 and interpolate all AMIP simulation data to 2.5° latitude/longitude spatial resolution.

Estimates of the observed conditions used in the analysis include the monthly precipitation from the Climate Prediction Center (CPC) monthly precipitation reconstruction over land (PREC/L; [Chen et al. 2002](#)) and the 200-mb (1 mb = 1 hPa) geopotential height (z200) from the NCEP–NCAR Reanalysis ([Kalnay et al. 1996](#)) for the time period from 1950–2017.

Following the same practice as in the NMME forecast system, the seasonal anomalies for the forecasts of each model and the AMIP simulation of each model are defined as the DJF seasonal mean departures from its own model's climatology computed over the hindcast period of 1982–2010. The seasonal anomalies for the observations are the departures from the observed climatology over the same period (i.e., 1982–2010).

b. The analysis approach

In the first set of analysis, we examine how well the basic features of the DJF forecast precipitation in terms of its mean, total interannual variability, and mean response to predicted ENSO SST for each model compare with corresponding observational estimates. The total interannual variability is defined as the average of seasonal mean precipitation standard deviation from individual members in the ensemble forecasts. The mean response to ENSO SST is quantified by the average of the linear regression of the seasonal mean precipitation forecast against the concurrent forecast value of seasonal mean Niño-3.4 SST index based on each individual member.

We also assess the anomaly correlation skill and the signal-to-noise ratio (SNR) for each model to assess whether the low prediction skill of DJF precipitation over the west coast in CFSv2 is an outlier compared to other models. The anomaly correlation skill is defined as the correlation of anomalies between the forecast ensemble means for each model and the observation. The signal component in the SNR is the standard deviation of ensemble mean while the noise component is the standard deviation of departure in the forecast individual members from the ensemble mean ([Kumar and Hoerling](#)

1995). The analysis of correlation skill is repeated with AMIP model simulations to contrast skill in seasonal prediction systems but without errors in the prediction of SSTs.

In the second part of the analysis we analyze the west coast precipitation response in each model to anomalous SSTs for individual El Niño events to assess to what extent the responses during individual events differ from the composite response. Based on such an extensive dataset of the ensemble forecasts available in the NMME, this analysis can be approached in two ways: 1) by analyzing the consistency of precipitation responses across El Niño events in a single model, and thereby examining the influence of ENSO SST flavor and possible nonlinearity in the response; and 2) by analyzing the consistency of precipitation responses across seven models for a specific El Niño event to examine if the consistency in response among models improves as the amplitude of El Niño events gets larger.

Following the classification of El Niño events as used in the Climate Prediction Center (CPC) (https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php), a total of 11 DJFs (except DJF 1987/88) are selected as El Niño events. They are the DJFs of 1982/83, 1986/87, 1991/92, 1994/95, 1997/98, 2002/03, 2004/05, 2006/07, 2009/10, 2014/15, and 2015/16. The reason why DJF 1987/88 is excluded from the analysis is that the forecast SSTs were in a La Niña pattern for CFSv2.

In the third component of the investigation, we analyze the DJF precipitation for the region of U.S. southeast coast with the same set of forecast models. Over the southeast coast the prediction skill is higher ([Fig. 2](#)) and therefore provides a contrasting case study to the analysis over the southwest coast of the United States.

In the final analysis the correlations of the DJF precipitation variability area averaged over the southwest coast and the southeast coast, respectively, to the 200-mb height field and SST are investigated to identify the origins and dynamical reasons for the DJF precipitation variability over these two regions and, further, why the precipitation skill over the regions may differ. In this observational analysis, the correlations are calculated over 1950–2016 and are complemented by AMIP simulations–based analysis over 1982–2014.

3. Results

As discussed in [section 1](#), [Fig. 2](#) shows the anomaly correlation skill of DJF precipitation for the forecast ensemble mean from CFSv2 computed over the entire hindcast period of 1982–2010 (a sample of 29 DJF seasons). It is found that the prediction skill along the west

FIG. 3. The DJF (top) precipitation climatology, (middle) standard deviation, and (bottom) linear regression (mm day^{-1}) to Niño-3.4 SST for seven models in the NMME (marked as CFSv2 and models A–F), and observations over the U.S. west coast are shown in the left side box of Fig. 2. Regression is for unit standard deviation of the Niño-3.4 SST index.

coast is low, and the largest values of correlation in the southwest do not exceed 0.3 (i.e., they explain approximately 10% of observed precipitation variability). In contrast, prediction skill is higher in the southeast with correlations reaching 0.6–0.7.

A low prediction skill based on the CFSv2 raises the question of whether it is an artifact of biases in this model or a consequence of low inherent predictability in nature. To investigate among two alternate possibilities, we analyze the ensemble forecasts from multiple models that are part of NMME.

a. Validation of precipitation variability in model forecasts

In this section, some basic features of DJF precipitation forecasts are first assessed and then compared with those in observations.

To keep the figure layout concise while displaying results from seven models and observations, the region of the U.S. west coast (wCoast) marked as the blue box in the left side of Fig. 2 is aligned to an $8^\circ \times 21^\circ$ longitude/latitude rectangle. Following this approach, each rectangle in Fig. 3 is the DJF precipitation climatology (top row), standard deviation (middle row), and linear regression with Niño-3.4 SSTs (bottom row) from each of seven models and for the observation.

For the DJF precipitation climatology, and its interannual variability, all seven models generally show a similar spatial pattern with smaller (larger) climatological precipitation in the south (north) and largest amplitude in interannual variability around 40°N . In general, the north–south variations in the model simulated precipitation climatology and standard deviation are like those in observations, although details differ.

It should be noted that because of the availability of an ensemble of forecasts, the estimates of climatology and standard deviation in models are based on a much larger sample, and hence estimates are more robust; the corresponding estimates in observations, however, do not have the same benefit. To obtain a somewhat more robust estimate of mean and standard deviation of precipitation in observations, estimates based on a longer period (1950–2016) were also computed, and the spatial structure as well as the amplitude over the two periods look very similar.

We next compare the ENSO-related response in models and observations. A simple approach to quantify the ENSO signal is to compute the regression between the interannual variability in seasonal mean precipitation with the Niño-3.4 SST index. For the estimate of the ENSO signal based on the linear regression (Fig. 3, bottom row), there is a general similarity in the spatial distribution with negative values in the north and positive in the south, a result consistent with the typical ENSO signal patterns documented in previous studies (e.g., Ropelewski and Halpert 1986; Hoerling and Kumar 1997; Kumar and Hoerling 1997, 1998; Trenberth et al. 1998; Chen and Kumar 2015). As we will show later, although the regression-based approach extracts the linear response in precipitation to ENSO variability, the spatial pattern looks remarkably like those based on the El Niño composites of precipitation anomaly. We also estimated the observed ENSO response in precipitation based on a dataset over 1950–2016 and basic features in the observed response remain the same.

It is noted that 1) there are differences in amplitude of regressions across models and a higher (lower) amplitude of regression implies a stronger (weaker) response to ENSO, and 2) just on its face value a higher or weaker response should not be taken as an indication that prediction skill for the respective models will also be higher or lower. It is because a stronger response, if erroneous compared to observations, can still lead to a lower skill, and 3) although regressions for the model forecasts could be influenced by model biases, bias is not an issue for regression based on observational data, which on the other hand can be influenced by limitations in the length of observational record.

Given that in all models, characteristics of the linear precipitation response differ from each other, and have differences from observations, it is not straightforward to anticipate which model is better in replicating the predictable component of observed interannual variability. The reasons for this are twofold. First, although the estimates of the statistical characteristics of precipitation variability in models is more robust (due to the availability of an ensemble of forecasts), the same

estimates in observations can be influenced by sampling, and therefore comparison between models and observations may not provide an indication of model's fidelity. The dependence on the analysis technique to create a gridded estimate for observed precipitation further exacerbates the observational uncertainty. Second, and more importantly, with the ensemble of forecasts while the predictable and unpredictable components of precipitation variability can be separated, the same, however, is not feasible for observations. Because of this, a comparison of signal and noise in the model and observations, beyond estimates based on an assessment of a linear signal (e.g., regression), cannot be made; also, as mentioned earlier, merely a stronger amplitude in the signal cannot be used as a criterion that skill for the model will also be higher.

Beyond comparing the precipitation response to ENSO as the linear signal, any attempts to validate differences in response to ENSO flavors (e.g., nonlinearity in response to ENSO amplitude or response to boundary forcings other than ENSO) are hindered by the limitations in the observational data. Faced with these difficulties and trying to understand if the low skill in predicting interannual precipitation variability in the CFSv2 (Fig. 2) is due to model biases or not, we next compare the prediction skill based on other models. The rationale is that prediction skill is determined by the separation of total interannual variability into predictable and unpredictable components (i.e., the SNR), and the model that has the most realistic (or unbiased) separation in comparison with observations will also have a higher skill. In other words, computation of prediction skill implicitly provides a check on the fidelity of the estimate of SNR in the model. Following this, it could be investigated if the relative strength of SNR across models is related to prediction skill, and whether the precipitation skill in the CFSv2 is systematically lower because of lower SNR.

Figure 4 shows the DJF precipitation anomaly correlation skill and its SNR for each model based on the 1982–2010 hindcast period. As mentioned in section 1, SNR is the ratio of ensemble mean variability (the predictable component) and the variability associated with the departures in the individual forecast members from the ensemble mean (the unpredictable component).

The basic feature to note in the comparison of skill is that the prediction skill across all the models is unanimously low and is near the skill for the CFSv2. Low prediction skill for CFSv2 (in Fig. 1), therefore, is not an outlier and is unlikely to be due to model biases influencing this model alone. In fact, a hurried opinion might lead one to draw the conclusion that CFSv2 may be

FIG. 4. The DJF precipitation correlation skill and SNR for the seven models over the U.S. west coast. Correlations (SNR) below 0.1 (0.3) are not shown. The area average AC for each model and NMME (going from left to right) is 0.24, 0.22, 0.00, 0.26, 0.14, 0.16, -0.02 , and 0.20 and the area average SNR is 0.46, 0.38, 0.46, 0.47, 0.48, 0.38, and 0.52. The seven-model average of AC is 0.14 and SNR is 0.45.

among the models with better prediction skill. However, for a short verification time series of 1982–2010, and for low skill regimes, the range of uncertainty in the estimate of skill score can be as big as the estimate of skill itself (Kumar 2009), and this can alone be responsible for some of the skill variations across the models. We also note that a conclusion of low prediction skill across all models is consistent with a long history of attempts to estimate seasonal predictability using observational and model simulations—a history that now stretches across a 40-yr period and has repeatedly led to low estimates of seasonal predictability in extratropical latitudes (Madden 1976; Jha et al. 2016).

Across the models, there is a much wider range of variability in the spatial pattern of skill than could have been anticipated from the corresponding differences in the models' linear response to ENSO (Fig. 3), and one cannot conclude that the amplitude of response relates to skill. For example, model A has a weak response in the north, but is also the one with highest skill; model F, on the other hand has the strongest signal in the north but has no skill; the response in model B has a reasonable depiction of the ENSO response in observation, but also has the smallest skill. In addition to sampling,

possible reasons for the lack of correspondence between the ENSO response (Fig. 3, bottom panels) and skill can include the following: 1) if the precipitation response to ENSO has appreciable nonlinearity then while skill will depend on fidelity in capturing the nonlinear response, an analysis of response based on a linear approach will not be able to capture that component; and 2) the estimate of the linear ENSO response in the model forecast may be biased and does not provide a prior assessment of what the skill for the forecast system may be.

To highlight the lack of correspondence between linear regression and skill, the SNR for each model is shown in Fig. 4 (bottom). A linear estimate of the observed SNR is also shown, and is based on the following approach: 1) by multiplying the linear ENSO response in precipitation (Fig. 3, rightmost panels) by the amplitude of the observed Niño-3.4 SST anomaly, a reconstruction of ENSO signal for each DJF in the analysis period is made; 2) the variability of the reconstructed ENSO signal is the linear estimate of precipitation associated with ENSO; 3) the noise component is estimated as the difference of total and signal variance; and 4) the ratio of signal and noise variability is the linear estimate of precipitation SNR.

A significance of the SNR is that for a perfect model, SNR also determines the expected value of skill (Kumar and Hoerling 2000; Kumar 2009). It is the biases in the model, together with small ensemble size and the influence of sampling over short verification time series, that can lead to departures from the relationship between SNR and skill. We note that while the model SNR (computed as the ratio of standard deviation of ensemble mean and the standard deviation of departures of individual forecasts from the ensemble mean) is not constrained by the assumption of linearity, the observed estimate, although unbiased, is constrained by the assumption of linearity.

The lack of correspondence that existed between linear response (Fig. 3, bottom) and skill (Fig. 4, top) also extends to between the nonlinear estimate of SNR and skill. The spatial pattern of SNR and linear response, in fact, has good spatial correspondence across all the models. The reason that SNR and skill do not bear a resemblance is the one discussed by Kumar et al. (2014), who indicated that because of model biases there does not have to be a relationship between the SNR (which has a theoretical relationship with skill under a *perfect* model assumption; Kumar 2009) and the *actual* skill. This is also the reason that a model with a large SNR (e.g., model F in northern latitudes) does not have to have higher correlation skill. An additional requirement for the SNR–skill relationship to hold is that the decomposition of total variability into predictable and unpredictable components (which determine the SNR) is like that in observations. One could conceive a forecast system in which precipitation response is very sensitive to ENSO, and thereby has a high SNR; however, if the same sensitivity does not exist in observation, the skill for that model will still be low. We also note that extending the analysis period to 1982–2016 (i.e., including the real-time forecasts also) only leads to minor changes in the skill and basic features of skill across all models are still replicated (not shown).

Apart from the possible influence of model biases (e.g., errors in tropical–extratropical teleconnection) that may lead to a low prediction skill, a low skill in SST prediction, particularly associated with ENSO, could also be an influencing factor in low skill for CFSv2 or in differences in skill across models. To demonstrate that SST prediction skill in all seasonal prediction systems over the ENSO region is unanimously high, skill in predicting DJF SSTs based on the same set of hindcasts is shown in Fig. 5. In the equatorial Pacific near and east of the date line, skill in predicting SST for all models exceeds 0.8 and has a similar spatial structure. We also attempted to relate errors in the prediction of SSTs in individual forecasts and errors in the prediction of area

DJF Tsfc Anomaly Correlation

FIG. 5. Skill in predicting DJF SSTs for various seasonal prediction systems. Prediction skill is computed as the forecast ensemble mean verifying against the observed seasonal mean over 1982–2010.

averaged precipitation over California; however, no systematic relationship was found (result not shown).

Another way to address the possible role of errors in SST prediction is to assess skill in simulating precipitation in AMIP simulations. In the AMIP simulations observed SSTs are prescribed, and therefore the possible influence of errors in prediction of SSTs can be discounted. We base our analysis on the AMIP simulations from four different models and simulation skill for each model is based on the ensemble mean for the respective model. Simulation skill for precipitation in the AMIP simulations (Fig. 6) has a structure like that for the

DJF Prec Anomaly Correlation

FIG. 6. Simulation skill of DJF mean precipitation from various AMIP integrations with models indicated in each panel. Simulation skill is computed for ensemble mean of AMIP integrations over 1982–2010. The two blue boxes are the areas defined as the U.S. west coast (wCoast) and U.S. southeast coast (seCoast). The area average AC over the wCoast is 0.28, 0.25, 0.34, and 0.25 for the CFSv2, CAM4, ECHAM5, and ESRL_CAM5, respectively, and over the seCoast is 0.49, 0.46, 0.55, 0.48 for the CFSv2, CAM4, ECHAM5, and ESRL_CAM5. The four-model average of AC is 0.28 and 0.50 over the wCoast and seCoast, respectively.

CFSv2, showing higher skill in the southern tier states, extending northward over the west and east coast of the United States. The magnitude of skill is generally similar to the other seasonal prediction systems over the west (Fig. 4, top panel) and the southeast coast (Fig. 11, top panel) with average skill for AMIP simulations being 0.28 (west coast) and 0.50 (southeast coast) and for seasonal predictions being 0.14 (west coast) and 0.47 (southeast coast). It is also interesting to note that for all seasonal prediction systems and AMIP simulations the skill over the southeast coast is systematically larger than over the west coast. These analyses of the skill of SST prediction (Fig. 5) and skill in simulating precipitation in the AMIP simulations indicate that errors in SST prediction are 1) not the cause for low prediction skill in the CFSv2 or 2) not linked to differences in precipitation skill across the models.

In summary, the analysis of prediction skill, and a lack of relationship with the SNR in the models, does not lead to the conclusion that low prediction skill in the CFSv2 (or in other models) is an artifact of some

systematic differences. Further, a similar feature of low prediction skill is shared across all the models in NMME and for AMIP simulations. A convergence of these results may start to hint toward the possibility that low skill in predicting west coast precipitation variability may just be a reflection of inherent predictability limits in nature.

b. Precipitation response in individual El Niño events

We next extend our analysis of forecast ensemble mean (i.e., the estimate of the response) for individual El Niño events, and thereby extend the analysis that primarily focused on an average response across all ENSO events. The analysis comparing individual El Niño events touches on questions like how sensitive the model response is to the different flavors and amplitude of ENSO SSTs, and whether some systematic conclusions about them can be obtained based on large forecast databases like NMME.

Figure 7 shows the forecast ensemble means of DJF precipitation over the west coast in each of 11 El Niño events (the columns) and for each of seven models (the

FIG. 7. The DJF precipitation ensemble means in each of 11 El Niño events arranged from the weak to strong event (from left to right) for each model, multimodel average (labeled as NMME), and observation (mm day^{-1}) over the U.S. west coast.

rows). Also included is the observed DJF precipitation (bottom row) and the forecast precipitation averaged across all the models (row labeled as NMME). The columns are arranged from the weak to strong El Niño events based on the strength of the observed Niño-3.4 SSTs during DJF. The maps shown in the rightmost column are the mean precipitation across all El Niño events, equivalent to the analysis based on the composite technique. It should be noted that during each El Niño the precipitation anomalies shown for models are the ensemble means and therefore constitute an estimate of mean precipitation response during that event.

This precipitation response is not constrained by any assumption of linearity and can fully incorporate the influence of ENSO flavors, ENSO amplitude, or other boundary conditions. The observed seasonal mean (equivalent to being a single realization of model forecast within an ensemble), on the other hand, is a combination of both the response and the contribution from the unpredictable noise, and therefore is expected to have much larger variations from one El Niño event to another (compared to the variations in the ensemble mean response). It is also noted that although an ensemble mean of 10 (which is the typical ensemble size for

the forecast systems in NMME) is generally sufficient to provide a good estimate for the mean value of the distribution of possible outcomes of seasonal means (Kumar and Hoerling 1995; Leutbecher 2018), on occasion it could still have a large contribution from internal variability.

The salient features in Fig. 7 are the following:

- A general comparison of ensemble mean response in the models across El Niño events with that in observations clearly indicates that the event-to-event variability in observations is much larger. If the observed seasonal mean anomalies have a large contribution from the internal variability, this would be the case (and would also imply a low SNR in observations and low prediction skill).
- Based on the general distribution of red and green colors an overall impression is that consistency in the response across models is better for stronger El Niño events (i.e., toward the rightmost columns), particularly over southern California (SCA) all models have above normal precipitation (except for model A in 1982). An increasing consistency in the model response with increasing strength of the El Niño would be consistent with the notion of the linearity in the amplitude of the response with the strength of the El Niño amplitude.
- For some models the response is very consistent across different El Niño events; this is true for the CFSv2 and model C, for which the north–south structure of negative–positive anomalies occurs for almost all El Niño events. In contrast, the response for models E and F gives a perception for a stronger nonlinearity with a predominately negative response over the entire west coast in the case of weak to moderate El Niño events changing to positive rainfall anomalies over SCA for strong El Niño events. Finally, for models A and D there is much more variability in response, with model A not having any consistent variation in response across all El Niño events and model D having large variations for weak-to-moderate El Niño events and a more consistent response for the strong amplitude El Niño events.
- Reasons for variations (or the lack of) in the precipitation response with the strength of the El Niño could be several, including just having strong amplitude linearity in response with a similarity in the spatial pattern of response across all El Niño events and a strong dependence in response to the flavors of El Niño leading to larger variations in the response. In addition, since the analysis is for initialized coupled predictions, differences (or errors) in SST predictions could also lead to inconsistent responses; a typical ensemble size of 10, at times, may be insufficient to dampen the contribution of internal variability in predictions. We did not explore further the possible influence of differences in SST predictions for specific El Niño event on the variations in response. As for the ensemble size, as mentioned earlier, an ensemble size of 10 is typically enough to provide a good estimate for the shift in the mean; however, for the models with weak response, contribution of noise may still dominate in the ensemble mean. For the CFSv2 (for which we have a 24-member ensemble), we repeated the analysis by randomly subsampling 12 ensemble members and found results to be the same (i.e., a similar consistency in the signal across different El Niño events).
- If the reasons for change in responses have a physical basis—nonlinearity in response to the El Niño amplitude or sensitivity to the flavors of El Niño—and if such variations in response are correct, one would expect that prediction skill for those models will be higher. This, however, is not the case; prediction skills for models E and F (with appreciable variations in response from weak to strong El Niño events) is not better than for CFSv2 and the model C.
- The precipitation response in NMME is largely consistent across all El Niño events with a positive response over California and a negative response over Oregon and Washington. The only exception is the El Niño event of 1994. We note that NMME response is based on an ensemble of 99 forecasts, and thus contribution of internal variability will be much smaller.
- Comparing precipitation responses across models for the same El Niño events presents a baffling scenario and does not lead to definitive conclusions. Weak El Niño events like 1994 seem to have a fair bit of consistency across models whereas the moderate event of 2009 has little consistency in precipitation response over California.

The analysis of precipitation response during El Niño years and across different models does not provide clear indications about how nonlinear the precipitation response to El Niño amplitude might be or what the influence of El Niño flavors is. Even with the availability of such an extensive hindcast dataset, our inability to draw definitive conclusions may again be consistent with a low signal-to-noise regime in that inferences about systematic variations in signal from one El Niño event to another would be hard to reach and the weak signal could easily be overwhelmed by the contribution of residual noise in the ensemble mean.

In the context of the analysis for DJF 2015/16, the analysis also does not provide conclusive evidence as to

FIG. 8. The consistency maps for the model ENSO responses across 11 El Niños over the U.S. west coast. For further explanation, see text.

whether the departures in the observed seasonal mean anomalies from the expected response were due to changes in the response to flavors of El Niño. As an alternative, the departure could just be explained by the contribution of internal variability in the observed seasonal mean. Even when comparing the three strongest events, where the precipitation response in the NMME ensemble mean has a good consistency over the SCA across three events, the negative observed anomalies over SCA for 2015 stand out. The same is true for the northern region of the analysis domain where the precipitation response had a different sign than the observed precipitation during 1982 and in 2015. In summary, an easy explanation for various features we have discussed so far is that over SCA, and over the west coast in general, the internal variability contributes a large fraction to the observed seasonal mean precipitation variability.

Next, we synthesize the consistency in the El Niño response during El Niño for a single model *or* across different models for an individual El Niño event. Shown in Fig. 8 are consistency maps across the 11 El Niño events for the model responses and for the observations. The consistency is calculated by counting the number of El Niño cases with the same sign anomaly (positive or negative) either in the model ensemble means or in the observations. The green-blue and yellow-red-brown colors represent the number of events with a similar-sign positive or negative anomaly, respectively. For example, all 11 El Niño events show positive anomalies over most of SCA and at least 8 of 11 El Niño events appear with negative anomalies over the northern parts of the west coast in the precipitation responses for CFSv2. It should be noted that the consistency in the models is for the precipitation response (with a smaller influence from the role of noise), whereas for the observed anomalies the contribution of noise could have a much larger influence during individual years (with the consequence that consistency will be lower).

In conformity with the results in Fig. 7, the consistency in response for model A is the weakest because of jumpiness in the sign of anomalies for different El Niño events. The level of consistency for model F is strongest in the northern latitudes. The response in the NMME is very good both over SCA and NCA. In general, across the models, consistency in the sign of anomalies is generally stronger in the northern part of the domain than in the southern latitudes. The consistency in the response, however, does not relate to the skill (Fig. 4) where model A (with least consistency) has a skill like others while model F (with highest consistency) has the lowest skill.

The uniformity in the sign of the *observed* anomalies across all El Niño events is weak, and this is consistent with the notion that the observed seasonal mean precipitation anomalies could have a large contribution from the internal variability. As to why the consistency in the models for responses across El Niño's events is not always high, there could have several reasons: how sensitive is the model's response to El Niño flavor; the quality of SST prediction; low SNR and the inability of the ensemble size of 10 to, at times, filter out the contribution of noise. As pointed out before, if differences in responses from event to event is correct, it should result in better skill; however, that is not the case. One could argue that skill itself could have been influenced by the sampling issues (because of the verification time series over 29 DJFs), and as the sampling issues are also much larger for smaller SNR ratios (Kumar 2009), this leads us back to the indication that the contribution of internal variability to seasonal mean precipitation anomalies is large.

Figure 9 shows consistency maps for the responses across seven models for individual El Niño events. In this figure, the green-blue and yellow-red-brown colors represent the number of models with similar sign positive or negative anomaly, respectively. The maps are arranged by the strength of El Niño events from weak to

FIG. 9. (top) The consistency maps in the model ENSO responses across seven models for each of 11 El Niño events and (bottom) the observed precipitation anomalies for the 11 El Niño events over the U.S. west coast (mm day^{-1}).

strong. The lower panels in Fig. 9 are the observed precipitation anomalies for the events and for ease of comparison; it is the repeat of the bottom panels in Fig. 7.

Except for 2014, the consistency in models is generally similar or higher in the northern latitudes than over California. This might once again indicate that the role of internal variability (the influence of ENSO) in determining why the observed outcome is larger (weaker) in the southern latitudes. One could also argue that low consistency in the model response over SCA is because some models capture changes in response to the flavors of ENSO while others do not, and because of this the consistency in the response across models is not maintained. However, the models that do seem to have a higher “sensitivity to the amplitude of ENSO” (e.g., models E and F) do not necessarily seem to have higher skill.

In summary, the analysis using a large set of hindcasts from multiple models does not provide convincing evidence that low skill in the CFSv2 in Fig. 2 may not have been due to model biases as it was similar to that based on other forecast systems. Among different possible explanations—for example, that all models share similar biases (including errors in tropical–extratropical teleconnections) or for all model prediction skill of ENSO SSTs is low—a much easier explanation for various results is that the observed anomalies could have large contributions from the unpredictable internal variability and, hence, have a small SNR. This explanation can

easily elucidate as to why the prediction skill for all models is low, why inferences about understanding differences in responses to the flavors of ENSO have been elusive, and why consistency in observed anomalies across El Niño events is low. This is not to say that model biases do not lead to a lower realization of predictability in observations, because they do, but the notion of inherent low practicability would be consistent with the results. To provide further support for this hypothesis, we next contrast the same analysis over the west coast with that over the southeast where for all prediction systems skill is consistently higher.

c. Comparison of precipitation variability over the U.S. southwest and southeast coasts

The basic features for simulating seasonal precipitation variability in the same ensemble of forecasts from seven models for the southeast coast area are shown in Fig. 10. Using the same layout as Fig. 3, each rectangle in Fig. 10 represents a $10^\circ \times 10^\circ$ longitude/latitude area for the southeast coast region (seCoast) marked as the blue box in the right side of Fig. 2.

Comparing the mean and standard deviation for models to that for observations, the magnitudes are generally similar, except for CFSv2 and model A, which have relatively larger mean values. The observed interannual variability in seasonal mean precipitation is fairly uniform throughout the domain and is a feature that is generally replicated in forecasts from all models.

FIG. 10. As in Fig. 3, but for the U.S. southeast coast area.

The linear regressions (ENSO signal) for all models have similar spatial patterns with northward decreasing gradient, and spatial distribution in models is consistent with that for observation. The amplitude of the precipitation response in the models is also comparable to observations, except for models A and B showing relatively larger values.

Comparing the results over the seCoast (Fig. 10) to those over the wCoast (Fig. 3) region, the magnitudes of mean precipitation and its standard deviation are lower, while the amplitude of linear regression for models and observation is similar. A lower variability and similar amplitude of response indicates larger SNRs, and would imply higher skill than over the wCoast.

Similar to Fig. 4, Fig. 11 shows the anomaly correlation and SNR for each model over the seCoast. As discussed, overall SNRs for all models are larger compared to that over the wCoast. This is also true for the linear estimate of SNR in observations. Furthermore, the overall prediction skill for all models is also higher and is consistent with larger SNRs.

The model precipitation responses during individual El Niño events and for each model for the seCoast area

are shown in Fig. 12. It is clearly seen that compared to the analysis for the wCoast, the responses are more consistent across events and across models for El Niños with moderate and strong strength. The sign of observed anomalies, which is a combination of ENSO response and the internal variability, also tends to be similar, except for the weak events shown in the left panels of Fig. 12. The composite anomalies for all models and observations also have a good resemblance. All of the results—higher skill and better consistency in response across different El Niño events—are inferences that would be implied if the SNR was higher.

From the consistency maps shown in Fig. 13, it can be seen that the consistency across different El Niño for model responses is high for models, especially in south, whereas it is much lower in the observation probably because of the contribution from the internal variability; the weak events at times had negative precipitation anomalies compared to the strong El Niño events (Fig. 12, bottom panel).

For individual El Niño events, the consistency across different models is high and improves as the strength of the event gets stronger, particularly for events that are stronger than 2002 El Niño (Fig. 14). Further, over this

FIG. 11. As in Fig. 4, but for the U.S. southeast coast. The area average AC for each model and NMME (from left to right) is 0.60, 0.44, 0.51, 0.56, 0.45, 0.41, 0.35, and 0.54 and the average SNR is 0.57, 0.82, 1.00, 0.52, 0.65, 0.47, and 0.61. The seven-model average of AC in 0.47 and SNR is 0.66.

region, higher consistency across models also better relates with the observed anomalies having the same sign. Only for the weak El Niño cases, the observed precipitation anomalies have a higher instance of having the opposite sign. This might be expected as weak El Niño events will have a weaker response and are more likely to be dominated by the contribution of internal variability.

In summary, for the same set of models the results for the analysis over the seCoast region present a contrasting situation compared to that for the wCoast region. The interannual variability over the seCoast has a higher SNR than that over the wCoast and consequences to this carry over to other measures that are influenced by the SNR: higher prediction skill (a feature also replicated in the AMIP simulation), higher consistency across model response, and a higher consistency in the sign of the observed seasonal mean precipitation anomaly and also with the models. Further, the same models that had difficulties in replicating interannual precipitation variability along the wCoast have a better performance in the seCoast.

d. Precipitation variability and its connection with circulation

In this section, we provide some dynamical basis for differences in SNR for precipitation variability along the wCoast and seCoast and attempt to illustrate that the differences in SNR over two regions may indeed be real. The analysis is based on two approaches. The first

component is an analysis based on the observational data alone, and therefore, is not influenced by biases. Observational analysis, however, can be influenced by sampling variability, and to reduce its impact, is repeated with the ensembles of AMIP simulations. The analysis investigates association between area averaged precipitation with the variability in 200-mb height (z200) and SSTs.

Figure 15 shows the correlations of the area-averaged southwest coast (swCoast) and seCoast precipitation and the Niño-3.4 SST index with z200, respectively, for the time period of 1950–2016. The swCoast region is the area south of 35° latitude of the wCoast (approximately the box outlined in Fig. 2). To enhance the statistical significance of the correlations to the extent we can, the analysis is for the 67 DJFs in 1950–2016. The correlation is repeated three times: over all 67 DJFs (left column), only over the 20 ENSO DJFs (when the Niño-3.4 SST index was at least one standard deviation; center column), and over the remaining 47 non-ENSO DJFs (right column). A comparison of the last two correlations breaks down the correlations between ENSO and the neutral years. The rationale for using total area-averaged precipitation variability is to analyze what fraction of variability is related to ENSO SSTs. The bottom row in Fig. 15 is the correlation between the Niño-3.4 and z200 and indicates the control of ENSO SST variability on seasonal mean heights. In all the

FIG. 12. As in Fig. 7, but for the U.S. southeast coast.

panels, correlation of the respective variable with the SST is also overlaid as contours.

The correlations between Niño-3.4 and z200 (Fig. 15, bottom panel) are largest over tropical latitudes and extend into extratropical Northern Hemisphere along the well-documented tropical–extratropical pathway associated with ENSO variability (Horel and Wallace 1981). As expected, the correlation is higher when done for ENSO years alone (bottom row, center) and weaker for neutral years (bottom row, right).

For the area-averaged swCoast precipitation, the amplitude of the correlation in tropical latitudes with z200 during the ENSO years is much weaker than what is indicated in the correlation of z200 with Niño-3.4. This is particularly true for tropical latitudes where variability in z200 is highly constrained by SSTs (e.g., correlation between Niño-3.4 and z200 near the date line is higher than 0.9). Lower correlation with tropical Pacific ENSO variability (SSTs and z200) implies that circulation associated with anomalous seasonal mean precipitation

FIG. 13. As in Fig. 8, but for the U.S. southeast coast.

FIG. 14. As in Fig. 9, but for the U.S. southeast coast.

does not have a dominant contribution from ENSO. We note that correlation of area-averaged precipitation south of 35° latitude is able to pick a dipole pattern of above (below) normal precipitation anomalies in the south (north) (not shown), and looks similar to the ENSO regression pattern in Fig. 3.

For the area-averaged precipitation over the seCoast, on the other hand, the correlation with z200 during ENSO years is stronger. Further, during ENSO years alone, the amplitude of correlation in tropical latitudes is closer to the ones indicated by ENSO SST variability. The same indication exists for correlation between seCoast precipitation and equatorial SSTs which are larger than the corresponding correlation with swCoast precipitation.

These results indicate that precipitation variations over California (seCoast) is less (more) constrained by ENSO SSTs and is influenced more (less) by internal variability, resulting in lower (higher) SNR.

As the analysis in Fig. 15 can be influenced by sampling in the observational data, the same analysis is repeated using AMIP simulations using multiple models. Instead of showing correlations with all four models individually, correlations are averaged and are shown in Fig. 16. The results based on observational data (Fig. 15) are essentially replicated in the model-based analysis, and once again, indicate that precipitation over the seCoast has a stronger connection with the interannual variability associated with ENSO.

4. Summary, discussion, and path forward

Faced with low skill in predicting seasonal mean precipitation over California, and the failure in predicting dry

conditions over southern California observed during the strong El Niño event of 2015/16 (which was opposite to the expected response during an El Niño; Fig. 7), we set out to investigate if the skill for the CFSv2 was low (Fig. 2) due to the possible influence of model biases or if it was a common feature also shared by other seasonal prediction systems, and therefore may point to inherent predictability limits in nature.

Based on the analysis of an extensive hindcast dataset in the NMME from seven seasonal forecast systems, no evidence was found that the skill of CFSv2 was an outlier; all models had a similarly low value in predicting seasonal mean precipitation variability over California. Various other analysis approaches, such as assessing the consistency in precipitation response across different models and determining the correlation of area-averaged precipitation with z200, also indicated that the signal-to-noise ratio in precipitation variability may inherently be low. In contrast, for the same hindcast dataset from the same seasonal prediction systems, analysis over the seCoast presented a different paradigm of being consistent with a higher SNR regime.

Another facet of the analysis was that for the precipitation variability over the wCoast the investigation did not provide answers to the questions like nonlinearity (or flavors) in the precipitation response to ENSO, even though the analysis was based on a large multimodel dataset. Such difficulties also point to a low SNR regime because under such scenarios a higher level of effort is required for extracting the signal (and variations therein) above the noise and drawing robust inferences becomes harder (e.g., requiring larger and larger ensemble sizes). In contrast, for the tropical latitudes where SNR in the context of ENSO variability is high,

FIG. 15. The spatial pattern of the correlation between the observed z200 and (top) the southwest coast area averaged precipitation, (middle) the southeast coast area averaged precipitation, and (bottom) the Niño-3.4 SST index, for (left) all 67 DJFs in 1950–2016, (center) 20 ENSO DJFs, and (right) 47 non-ENSO DJFs. In all the panels, the black contours are the correlations of respective variable with the SST.

even small size ensembles are generally enough to determine the signal (Kumar and Hoerling 1995).

The self-consistent relationship between SNR, ensemble size (required for quantifying the signal), and prediction skill is summarized in Fig. 17 and fits the contrasting situation of predictability of seasonal mean precipitation over the wCoast and seCoast. A low (high) SNR requires large (small) ensemble sizes to extract the predictable signal and is also associated with low (high) prediction skill. Conversely, a low prediction skill across a wide range of prediction systems, or difficulties in obtaining agreements in responses, may harbingers the possibility that the SNR is low.

In the practice of making seasonal predictions, going beyond the utility of ENSO composites has proven to be a challenging task and any information about flavors of response and so on has not been internalized to become part of routine forecast practices. We note that the classical ENSO response pattern in the atmospheric variability was first documented by Horel and Wallace

(1981); since then, further advances in reaching agreed-upon inferences about higher-order variation in responses have not been forthcoming. It may indeed be that the impediments in advances may not be the analysis methods or biases in the models or the length of the observational record; rather it may be that SNR is low and has been the fundamental constraint.

This analysis also leads to the question as what could be done to reach closure on some of questions that continue to bedevil the extended-range forecasting community. For example, what degree of flavor does the ENSO response have, and can this information be used in framing seasonal forecasts on a routine basis? And even if we may not be able to reach closure, can a community-wide understanding be developed that acknowledges that given the current generation of analysis tools we have, we are not in a position to answer these questions, and therefore, for now, the composite ENSO response is the best heuristics to draw upon?

AMIP DJF Prec ~ z200 Correlation

FIG. 16. As in Fig. 15, but for AMIP simulations. For each model, correlations are computed for each simulation and are then averaged across all ensemble members over 1982–2014 period. For sake of brevity, ensemble averaged correlations are averaged further over all four models.

An approach toward reaching consensus may be a systematic and coordinated set of experiments done within the purview of seasonal prediction systems run at the operational centers, and analysis of those experiments can be used to develop a synthesis about the current state of knowledge. Similar to the CMIP exercise (Eyring et al. 2016), a limited set of experiments can be repeated periodically as seasonal prediction systems improve. This approach contrasts with the current paradigm of isolated case studies that generally do not contribute to advancement in forecast practices. For example, more than 15 papers that appeared in the peer-reviewed journals on analyzing seasonal mean precipitation over California during 2015/16 that attribute differences in seasonal forecasts and observed precipitation to factors such as internal variability (in line with the conclusions presented here; Jong et al. 2018; Lim et al. 2018; Zhang et al. 2018; Cash and Burls 2019;

Swenson et al. 2019), decline in the Arctic sea ice (Cohen et al. 2017), sensitivity in precipitation response to flavors of ENSO (Paek et al. 2017; Siler et al. 2017; Patricola et al. 2020), and dry land surface conditions over southern California (Yang et al. 2018) have not led to usable advancements in our understanding and, if anything, might have led to more confusion. A coordinated assessment of responses in seasonal prediction systems would also help shed light on the questions like 1) what specific biases in models matter the most for the realization of seasonal predictability and 2) given that model biases are often the blame for low prediction skill (or when seasonal predictions fail), what metrics can quantify the “fidelity” of the models and their suitability for seasonal predictions (Kumar et al. 1996). Answering questions like these is important because although models will continue to improve and biases will reduce with time, it is unlikely that they will go to zero in the near future. Even with

FIG. 17. The schematic illustrating the self-consistent relationship between signal-to-noise ratio, expected skill of seasonal prediction, and the required ensemble size (either to infer the signal or to realize the expected skill). The low (high) signal-to-noise scenario has a low (high) expected skill and requires large (small) ensemble sizes to realize the skill.

improved models, if the model-based estimates of predictability stay low, the question of the influence of model biases will continue to cast doubt on the validity of updates in the estimates of predictability.

We also need to recognize that if inferences about limits of predictability cannot be obtained based on first principles, then the alternative is to build them based upon the convergence of circumstantial evidence using different analysis tools. It is noted that because the evidence for the estimates of predictability are circumstantial, there are always several alternatives that can be used to explain results in the present analysis. These include the following: all models share common biases (including errors in tropical–extratropical teleconnection), and hence all seasonal prediction system have low skill in predicting west coast precipitation; all seasonal prediction systems have some errors in predicting SSTs associated with ENSO (although skill for short-lead prediction is ~ 0.9), and this could be responsible for low prediction skill in west coast precipitation; and inherent SNR for precipitation is low and is responsible for low precipitation skill. Out of various possible explanations, the alternative of low inherent SNR is the simplest. This alternative is also consistent with a 40-yr history of attempts to estimate seasonal predictability in extratropical latitudes, and all results based either on observational data or based on generation of models (which have improved over time) have led to the same conclusion of low predictability (Madden 1976; Jha et al. 2016). The choice of this alternative, of course, is open to further analysis and debate, and as better seasonal prediction systems are put in place, can be easily tested.

One should also recognize that the same models that fail to predict extratropical features are much better at predicting seasonal mean anomalies in the tropical latitudes compared to those in the extratropics for some variables, such as 200-mb heights (Kumar and Hoerling 1995; Phelps et al. 2004; Kumar and Chen 2015). The difference is because of much higher SNR in the tropics compared to that in the extratropics. It is also noted that the question of model biases gets accentuated when high-profile predictions (based on a collection of prediction systems) fail and the null hypothesis that forecast failure could be due to contribution of internal variability is ignored.

Acknowledgments. We thank Dr. Emily Becker for providing the forecast data from the NMME project and Dr. Bhaskar Jha for maintaining the CFSv2 AMIP runs. We are also very thankful for ESRL providing online access to their AMIP simulations. We also thank the editor and four anonymous reviewers for their comments that led to improvements in the final version of the manuscript.

REFERENCES

- Barsugli, J. J., and P. D. Sardeshmukh, 2002: Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific basin. *J. Climate*, **15**, 3427–3442, [https://doi.org/10.1175/1520-0442\(2002\)015<3427:GASTTS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3427:GASTTS>2.0.CO;2).
- Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, <https://doi.org/10.1175/JCLI-D-13-00597.1>.
- Bond, N. A., M. F. Cronin, H. Freeland, and N. Mantua, 2015: Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophys. Res. Lett.*, **42**, 3414–3420, <https://doi.org/10.1002/2015GL063306>.
- Cash, B. A., and N. J. Burls, 2019: Predictable and unpredictable aspects of U.S. west coast rainfall and El Niño: Understanding the 2015/16 event. *J. Climate*, **32**, 2843–2868, <https://doi.org/10.1175/JCLI-D-18-0181.1>.
- Chen, M., and A. Kumar, 2015: Influence of ENSO SSTs on the spread of the probability density function for precipitation and land surface temperature. *Climate Dyn.*, **45**, 965–974, <https://doi.org/10.1007/s00382-014-2336-9>.
- , and —, 2018: Winter 2015/16 atmospheric and precipitation anomalies over North America: El Niño response and the role of noise. *Mon. Wea. Rev.*, **146**, 909–927, <https://doi.org/10.1175/MWR-D-17-0116.1>.
- , P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global land precipitation: A 50-yr monthly analysis based on gauge observations. *J. Hydrometeorol.*, **3**, 249–266, [https://doi.org/10.1175/1525-7541\(2002\)003<0249:GLPAYM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2).
- Chiodi, A. M., and D. E. Harrison, 2015: Global seasonal precipitation anomalies robustly associated with El Niño and La Niña events—An OLR perspective. *J. Climate*, **28**, 6133–6159, <https://doi.org/10.1175/JCLI-D-14-00387.1>.
- Cohen, J., K. Pfeiffer, and J. Francis, 2017: Winter 2015/16: A turning point in ENSO-based seasonal forecasts. *Oceanography*, **30**, 82–89, <https://doi.org/10.5670/oceanog.2017.115>.

- Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>.
- Graham, R. J., and Coauthors, 2011: Long-range forecasting and the global framework for climate services. *Climate Res.*, **47**, 47–55, <https://doi.org/10.3354/cr00963>.
- Hoell, A., M. Hoerling, J. Eischeid, K. Wolter, R. Dole, J. Perlwitz, T. Xu, and L. Cheng, 2016: Does El Niño intensity matter for California precipitation? *Geophys. Res. Lett.*, **43**, 819–825, <https://doi.org/10.1002/2015GL067102>.
- Hoerling, M. P., and A. Kumar, 1997: Origins of extreme climate states during the 1982–83 ENSO winter. *J. Climate*, **10**, 2859–2870, [https://doi.org/10.1175/1520-0442\(1997\)010<2859:OOECSD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<2859:OOECSD>2.0.CO;2).
- Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813–829, [https://doi.org/10.1175/1520-0493\(1981\)109<0813:PSAPAW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2).
- Hu, Z.-Z., A. Kumar, B. Jha, W. Wang, B. Huang, and B. Huang, 2012: An analysis of warm pool and cold tongue El Niños: Air-sea coupling processes, global influences, and recent trends. *Climate Dyn.*, **38**, 2017–2035, <https://doi.org/10.1007/s00382-011-1224-9>.
- , —, —, J. Zhu, and B. Huang, 2017: Persistence and predictions of the remarkable warm anomaly in the north-eastern Pacific Ocean during 2014–16. *J. Climate*, **30**, 689–702, <https://doi.org/10.1175/JCLI-D-16-0348.1>.
- Jha, B., A. Kumar, and Z.-Z. Hu, 2016: An update on the estimate of predictability of seasonal mean atmospheric variability using North American Multi-Model Ensemble. *Climate Dyn.*, **53**, 7397–7409, <https://doi.org/10.1007/S00382-016-3217-1>.
- Johnson, N. C., and Y. Kosaka, 2016: The impact of eastern equatorial Pacific convection on the diversity of boreal winter El Niño teleconnection patterns. *Climate Dyn.*, **47**, 3737–3765, <https://doi.org/10.1007/s00382-016-3039-1>.
- Jong, B.-T., M. Ting, R. Seager, N. Henderson, and D.-E. Lee, 2018: Role of equatorial Pacific SST forecast error in the late winter California precipitation forecast for the 2015/16 El Niño. *J. Climate*, **31**, 839–852, <https://doi.org/10.1175/JCLI-D-17-0145.1>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- 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>.
- Kumar, A., 2009: Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Mon. Wea. Rev.*, **137**, 2622–2631, <https://doi.org/10.1175/2009MWR2814.1>.
- , and M. Hoerling, 1995: Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345, [https://doi.org/10.1175/1520-0477\(1995\)076<0335:PALOSA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0335:PALOSA>2.0.CO;2).
- , and —, 1997: Interpretation and implications of observed inter-El Niño variability. *J. Climate*, **10**, 83–91, [https://doi.org/10.1175/1520-0442\(1997\)010<0083:IAOTO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0083:IAOTO>2.0.CO;2).
- , and —, 1998: Annual cycle of Pacific–North American predictability associated with different phases of ENSO. *J. Climate*, **11**, 3295–3308, [https://doi.org/10.1175/1520-0442\(1998\)011<3295:ACOPNA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3295:ACOPNA>2.0.CO;2).
- , and —, 2000: Analysis of conceptual model of seasonal climate variability and implication for seasonal predictions. *Bull. Amer. Meteor. Soc.*, **81**, 255–264, [https://doi.org/10.1175/1520-0477\(2000\)081<0255:AOACMO>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0255:AOACMO>2.3.CO;2).
- , and M. Chen, 2015: Inherent predictability, requirements on the ensemble size, and complementarity. *Mon. Wea. Rev.*, **143**, 3192–3203, <https://doi.org/10.1175/MWR-D-15-0022.1>.
- , M. Hoerling, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate*, **9**, 115–129, [https://doi.org/10.1175/1520-0442\(1996\)009<0115:AAGSFM>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0115:AAGSFM>2.0.CO;2).
- , P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, <https://doi.org/10.1175/MWR-D-13-00287.1>.
- Kushnir, Y., W. A. Robinson, I. Bladé, N. M. Hall, S. Peng, and R. Sutton, 2002: Atmospheric GCM response to extra-tropical SST anomalies: Synthesis and evaluation. *J. Climate*, **15**, 2233–2256, [https://doi.org/10.1175/1520-0442\(2002\)015<2233:AGRTES>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2233:AGRTES>2.0.CO;2).
- Larkin, N. K., and D. E. Harrison, 2005: On the definition of El Niño and associated seasonal average U.S. weather anomalies. *Geophys. Res. Lett.*, **32**, L13705, <https://doi.org/10.1029/2005GL022738>.
- Leutbecher, M., 2018: Ensemble size: How suboptimal is less than infinity? *Quart. J. Roy. Meteor. Soc.*, **145**, 107–128, <https://doi.org/10.1002/qj.3387>.
- L'Heureux, M., and Coauthors, 2017: Observing and predicting the 2015/16 El Niño. *Bull. Amer. Meteor. Soc.*, **98**, 1363–1382, <https://doi.org/10.1175/BAMS-D-16-0009.1>.
- Lim, Y.-K., S. D. Schubert, Y. Chang, A. M. Molod, and S. Pawson, 2018: The impact of SST-forced and unforced teleconnections on 2015/16 El Niño winter precipitation over the western United States. *J. Climate*, **31**, 5825–5844, <https://doi.org/10.1175/JCLI-D-17-0218.1>.
- Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.*, **104**, 942–952, [https://doi.org/10.1175/1520-0493\(1976\)104<0942:EOTNVO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<0942:EOTNVO>2.0.CO;2).
- Paek, H., J.-Y. Yu, and C. Qian, 2017: Why were the 2015/2016 and 1997/1998 extreme El Niño different? *Geophys. Res. Lett.*, **44**, 1848–1856, <https://doi.org/10.1002/2016GL071515>.
- Palmer, T. N., and D. A. Mansfield, 1986: A study of wintertime circulation anomalies during past El Niño events using a high resolution general circulation model. II: Variability of the seasonal mean response. *Quart. J. Roy. Meteor. Soc.*, **112**, 639–660, <https://doi.org/10.1002/qj.49711247305>.
- Patricola, C. M., and Coauthors, 2020: Maximizing ENSO as a source of western US hydroclimate predictability. *Climate Dyn.*, **54**, 351–372, <https://doi.org/10.1007/s00382-019-05004-8>.
- Pegion, K., and A. Kumar, 2013: Does an ENSO-conditional skill mask improve seasonal predictions. *Mon. Wea. Rev.*, **141**, 4515–4533, <https://doi.org/10.1175/MWR-D-12-00317.1>.
- Phelps, M. W., A. Kumar, and J. J. O'Brien, 2004: Potential predictability in the NCEP CPC dynamical seasonal forecast system. *J. Climate*, **17**, 3775–3785, [https://doi.org/10.1175/1520-0442\(2004\)017<3775:PPITNC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3775:PPITNC>2.0.CO;2).
- Quan, X., M. Hoerling, J. Whitaker, G. Bates, and T. Xu, 2006: Diagnosing sources of U.S. seasonal forecast skill. *J. Climate*, **19**, 3279–3293, <https://doi.org/10.1175/JCLI3789.1>.
- Robinson, W. A., 2000: Review of WETS—The workshop on extra-tropical SST anomalies. *Bull. Amer. Meteor. Soc.*, **81**,

- 567–578, [https://doi.org/10.1175/1520-0477\(2000\)081<0567:ROWTWO>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0567:ROWTWO>2.3.CO;2).
- Ropelewski, C. F., and M. S. Halpert, 1986: North America precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, [https://doi.org/10.1175/1520-0493\(1986\)114<2352:NAPATP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2).
- Schubert, S. D., and Coauthors, 2009: A U.S. CLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: Overview and results. *J. Climate*, **22**, 5251–5272, <https://doi.org/10.1175/2009JCLI3060.1>.
- Seager, R., M. Hoerling, S. Schuber, H. Wang, G. Lyon, A. Kumar, J. Nakamura, and N. Henderson, 2015: Causes of the 2011–14 California drought. *J. Climate*, **28**, 6997–7024, <https://doi.org/10.1175/JCLI-D-14-00860.1>.
- Siler, N., Y. Kosaka, S.-P. Xie, and X. Li, 2017: Tropical ocean contributions to California’s surprisingly dry El Niño of 2015/16. *J. Climate*, **30**, 10 067–10 079, <https://doi.org/10.1175/JCLI-D-17-0177.1>.
- Singh, D., M. Ting, A. A. Scaife, and N. Martin, 2018: California winter precipitation predictability: Insights from the anomalous 2015–2016 and 2016–2017 seasons. *Geophys. Res. Lett.*, **45**, 9972–9980, <https://doi.org/10.1029/2018GL078844>.
- Swain, D. L., M. Tsiang, M. Haugen, D. Singh, A. Charland, S. Rajaratnam, and N. S. Diffenbaugh, 2014: The extraordinary California drought of 2013/14: Character, context, and the role of climate change [in “Explaining Extremes of 2013 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **95** (9), S3–S7, https://www.ametsoc.org/ams/assets/File/publications/BAMS_EEE_2013_Full_Report_high-res.pdf.
- Swenson, E. T., D. M. Straus, C. E. Snide, and A. Fahad, 2019: The role of tropical heating and internal variability in the California response to the 2015/16 ENSO event. *J. Climate*, **76**, 3115–3128, <https://doi.org/10.1175/JAS-D-19-0064.1>.
- Swinbank, R., and Coauthors, 2016: The TIGGE project and its achievements. *Bull. Amer. Meteor. Soc.*, **97**, 49–67, <https://doi.org/10.1175/BAMS-D-13-00191.1>.
- Trenberth, E. K., Q. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. F. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperature. *J. Geophys. Res.*, **103**, 14 291–14 324, <https://doi.org/10.1029/97JC01444>.
- Vitart, F., and Coauthors, 2016: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Wang, S.-Y., L. Hipps, R. R. Gillies, and J.-H. Yoon, 2014: Probable causes of the abnormal ridge accompanying the 2013–14 California drought ENSO precursor and anthropogenic warming footprint. *Geophys. Res. Lett.*, **41**, 3220–3226, <https://doi.org/10.1002/2014GL059748>.
- Yang, X., L. Jia, S. B. Kapnick, T. L. Delworth, G. A. Vecchi, R. Gudgel, S. Underwood, and F. Zeng, 2018: On the seasonal prediction of the western United States El Niño precipitation pattern during the 2015/16 winter. *Climate Dyn.*, **51**, 3765–3783, <https://doi.org/10.1007/s00382-018-4109-3>.
- Yu, J.-Y., and S. T. Kim, 2010: Identification of Central-Pacific and Eastern-Pacific types of ENSO in CMIP3 models. *Geophys. Res. Lett.*, **37**, L15705, <https://doi.org/10.1029/2010GL044082>.
- Zhang, T., and Coauthors, 2018: Predictability and prediction of southern California rains during strong El Niño events: A focus on the failed 2016 winter rains. *J. Climate*, **31**, 555–574, <https://doi.org/10.1175/JCLI-D-17-0396.1>.