

Toward Identifying Subseasonal Forecasts of Opportunity Using North American Weather Regimes

ANDREW W. ROBERTSON, NICOLAS VIGAUD, AND JING YUAN

International Research Institute for Climate and Society, Columbia University, Palisades, New York

MICHAEL K. TIPPETT

Department of Applied Mathematics and Applied Physics, Columbia University, New York, New York

(Manuscript received 28 August 2019, in final form 27 February 2020)


ABSTRACT

Large-scale atmospheric circulation regime structures are used to diagnose subseasonal forecasts of wintertime geopotential height fields over the North American sector, from the NCEP CFSv2 model. Four large-scale daily circulation regimes derived from reanalysis 500-hPa geopotential height data using K -means clustering are used as a low-dimensional basis for diagnosing the model's forecasts up to 45 days ahead. On average, hindcast skill in regime space is found to be limited to 10–15 days ahead, in terms of anomaly correlation of 5-day averages of regime counts, over the 1999–2010 period. However, skill up to 30 days ahead is identified in individual winters, and intraseasonal episodes of high skill are identified using a forecast-evolution graphical tool. A striking vacillation between the West Coast and Pacific ridge patterns during December–January 2008/09 is shown to be predicted 20–25 days in advance, illustrating the possibility to identify “forecasts of opportunity” when subseasonal forecast skill is much higher than the average. The forecast-evolution tool also provides insight into the poor seasonal forecasts of California precipitation by operational centers during the 2015/16 El Niño winter. The Pacific trough regime is shown to be greatly overpredicted beyond 1–2 weeks in advance during the 2015/16 winter, with weather-scale features dominating the forecast evolution at shorter lead times. A similar though less extreme situation took place during the weaker El Niño of 2009/10, with the Pacific trough overforecast at S2S lead times.

1. Introduction

Subseasonal to Seasonal (S2S) forecasting at lead times of about 10–100 days has garnered much recent attention due to advances in dynamical model skill at predicting the Madden–Julian oscillation (MJO; Vitart 2014; Kim et al. 2018) and through the construction of new multimodel databases of subseasonal forecasts and reforecasts (Vitart et al. 2017; Pegion et al. 2019). However, the degree of useful forecast skill is still unclear in the extratropics where weather and atmospheric low-frequency variability (10–100 days) are dominated by vigorous synoptic-scale activity associated with baroclinic waves, and by persistent large-scale circulation patterns, respectively.

In the midlatitudes, certain large-scale flow patterns appear repeatedly at fixed geographical locations and persist beyond the lifetimes of individual weather disturbances. These flow patterns were termed “weather regimes” (WRs) by Reinhold and Pierrehumbert (1982) to convey the important dynamical role of the synoptic-scale transients in maintaining quasi-stationary dynamical balance with the planetary scales. Robertson and Ghil (1999) emphasized the connection between WRs and surface weather, by using probability distribution function (PDF) estimation and K -means clustering of daily 700-hPa geopotential height maps to link daily temperature and rainfall statistics at synoptic stations over the western United States with ENSO. While clustering of the circulation variability into a few preferred discrete “regimes” makes no formal dynamical-systems requirement of quasi-stationarity, performing the cluster analysis on low-pass-filtered data in the subspace of the leading principal components (PCs) emphasizes the large-scale subseasonal component of

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Dr. Andrew W. Robertson, awr@iri.columbia.edu

DOI: 10.1175/MWR-D-19-0285.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

TABLE 1. Contingency table of MERRA reanalysis (1999–2010) weather regime transition counts, with the probabilities (%) in parentheses. Asterisks (*) indicate statistical significance at the 0.1% level of a χ^2 test. The transitions should be read “from row to column.”

From/to	Regime 1	Regime 2	Regime 3	Regime 4	Total
Regime 1	558* (88)	31 (5)	22 (3)	25 (4)	636
Regime 2	30 (5)	534* (89)	23 (4)	12 (2)	599
Regime 3	25 (5)	31 (7)	399* (86)	8 (2)	463
Regime 4	21 (7)	6 (2)	19 (6)	258* (85)	304
Total	634 (32)	602 (30)	463 (23)	303 (15)	2002

the variability, as opposed to the faster synoptic scales (cf. Straus et al. 2017). The resulting regime self transition probabilities (shown in Tables 1 and 2 below) all exceed 0.8, indicating that the regimes are indeed highly persistent. However, the weather regime concept is used here in a qualitative sense, and does not address the issue of whether multiple equilibria exist on S2S time scales between two weeks and a season.

The regime approach incorporates the variance compression of PC analysis, enabling the identification of preferred polarities and linear combinations of the PCs, relaxing the orthogonality requirement and providing a set of more physically realistic patterns (Straus et al. 2017). Several regime-based methodologies have demonstrated robust identification of the leading patterns of atmospheric low-frequency behavior over the Pacific North American sector (Robertson and Ghil 1999; Smyth et al. 1999; Straus et al. 2007). The weather regime methodology has been used to link both MJO and ENSO to extreme floods over the Midwest (Robertson et al. 2015), and ENSO impacts over North America (Riddle et al. 2013). Becker et al. (2013) found the CFSv2 skill for extremes in monthly or seasonal forecasts to be higher than when all forecasts are considered, providing further evidence for the predictability of extremes.

The S2S scale lies between the realms of daily weather forecasts and seasonal forecasts, which deal with seasonal averages or daily weather statistics such as rain/day frequency (e.g., Moron et al. 2006). It is not clear how the intermediate subseasonal scale should be treated.

TABLE 2. Contingency table of weather regime transitions (from row to column), from the CFSv2 1–45 day hindcasts. Asterisks (*) indicate statistical significance at the 0.1% level of a χ^2 test. The 5-day running averages of the CFSv2 4-member ensemble-mean hindcasts were projected into the space of the four MERRA regimes.

From/to	Regime 1	Regime 2	Regime 3	Regime 4	total
Regime 1	22 668* (90)	819 (3)	602 (2)	1236 (5)	25 325
Regime 2	407 (4)	8285 (81)	1089 (10)	502 (5)	10 283
Regime 3	945 (4)	807 (4)	19 979* (89)	728 (3)	22 459
Regime 4	1341 (7)	421 (2)	732 (3)	17 517* (88)	20 011
Total	25 361 (32)	10 332 (13)	22 402 (29)	19 983 (26)	78 078

A subseasonal forecast might be judged accurate according to its ability to capture regime transitions within a tolerance of a few days. The weather regime concept has received renewed interest as a tool for diagnosing medium- and extended-range forecasts (Ferranti et al. 2015). Palmer (1988) found that the predictive skill at medium and subseasonal range was dependent on the phase of the Pacific/North American (PNA) pattern. Vigaud et al. (2018, hereafter VRT) diagnosed ECMWF reforecasts from the World Weather Research Programme/World Climate Research Programme Subseasonal to Seasonal Prediction Project database (Vitart et al. 2017), in terms of four extended-winter regimes identified over North America, using a *K*-means clustering applied daily 500-hPa geopotential height fields. They found very similar regime spatial structures, daily regime occurrences and seasonal regime counts in ECMWF week-1 reforecasts when compared with MERRA reanalysis data, indicating that the model captures the deterministic evolution of large-scale circulation well during the first 7 days. VRT then used the MERRA regime structures to define a four-dimensional subspace to track the ECMWF forecast evolution from week 1 to week 4; they found anomaly correlation skills for two of the four regimes of 0.5 up to 14 days ahead.

The goal of this paper is to apply the weather regime approach to diagnose the S2S reforecasts and forecasts made by the NCEP CFSv2 model over North America, as a means to identify skillful episodes in individual winters that could provide a framework for identifying S2S “forecasts of opportunity.” The analysis exploits the daily initializations of the NCEP CFSv2 forecasts, enabling a continuous analysis of forecast evolution. Section 2 provides methodological details. The results are presented in section 3, with concluding remarks in section 4.

2. Data

MERRA reanalyses version 1 (Rienecker et al. 2011) geopotential height data at 500 hPa (Z500) is used to

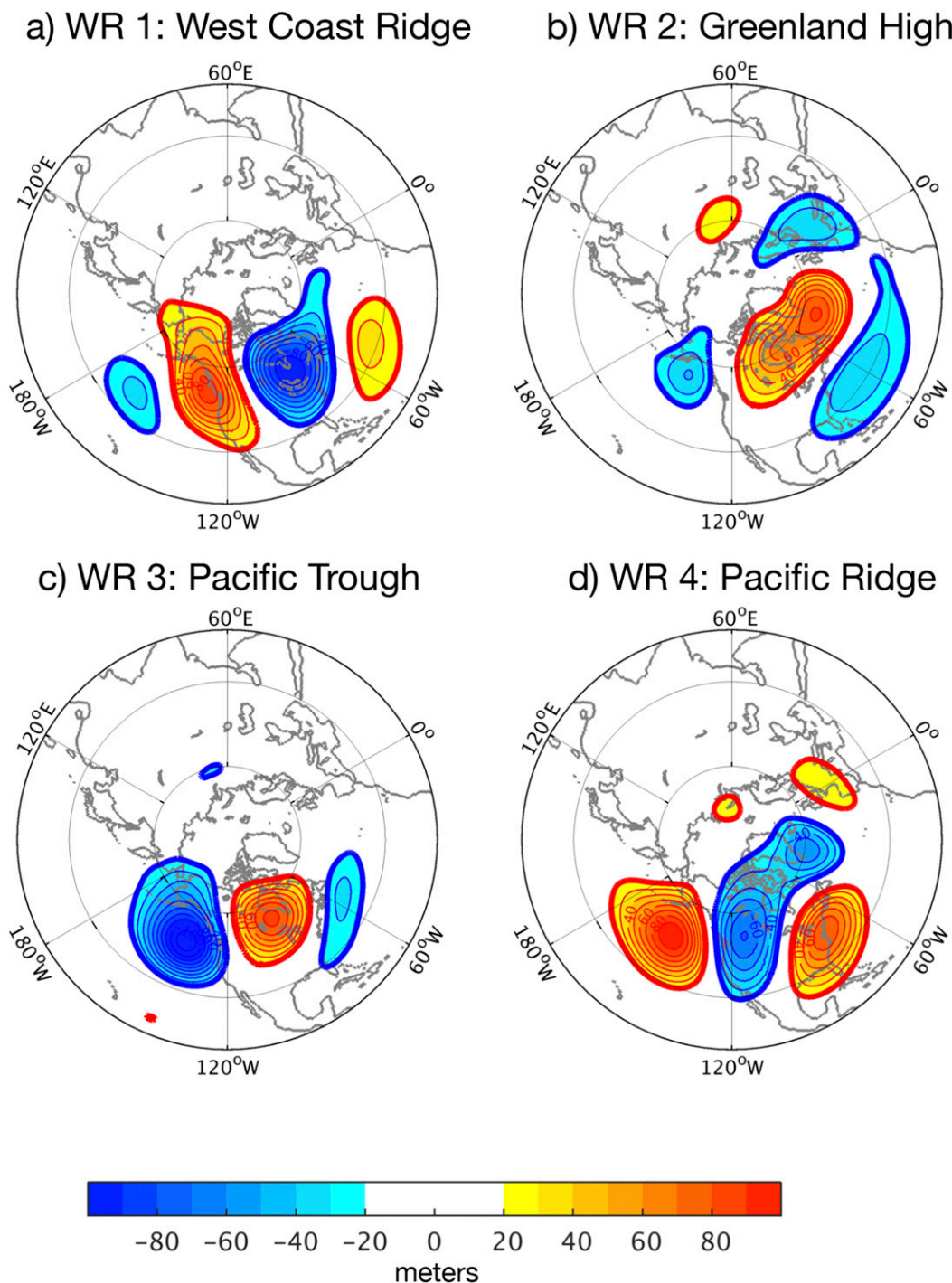


FIG. 1. Regime composites of 500-hPa geopotential height anomalies (in gpm), computed from MERRA reanalysis data. Shaded anomalies are significant at the 5% level of a Student's *t* test.

derive the “observed” regimes on a $1/2^\circ \times 2/3^\circ$ grid from 1982 to 2014, following VRT. The National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) Reanalysis Project (NNRP), version 1, 2.5° gridded data (Kalnay et al. 1996), is used

in section 4 for additional diagnostic analysis (including pentad geopotential height anomalies constructed with IRI Data Library Maproom <http://iridl.ldeo.columbia.edu/maproom/> in Fig. 4), along with the NCEP Climate Prediction Center (CPC) Unified Precipitation on a $1/4^\circ$ grid

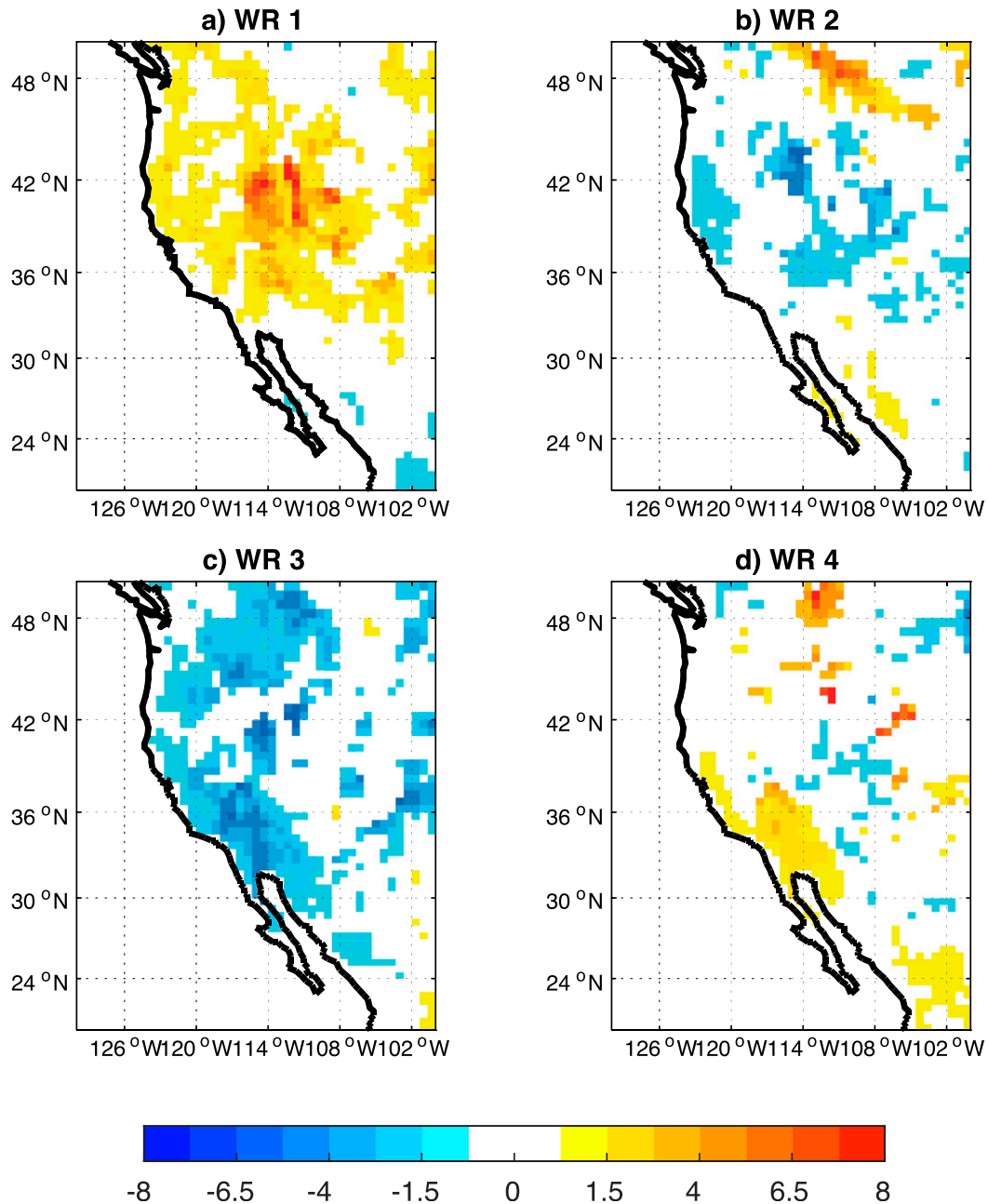


FIG. 2. Regime composites of fire weather index (dimensionless), expressed as deviations from the 1982–2014, October–March long-term average.

(Chen et al. 2008). These datasets were obtained via IRI Data Library <http://iridl.ldeo.columbia.edu>.

The NCEP CFSv2 model (Saha et al. 2014) has 64 vertical levels and T126 (ca. 1°) in the horizontal. Over the reforecast period 1999–2010, 4 ensemble members are run each day (1 member initialized every 6 h) up to 45 days ahead, with 16 members per day available for the 2015/16 forecast period analyzed. The model data are provided on a 1.5° grid in

the S2S database, which was accessed from IRI Data Library.

3. Methodology

a. Cluster analysis

Daily MERRA Z500 anomaly maps for the October–March season were first obtained by 1) taking 5-day running averages to filter out daily time-scale weather

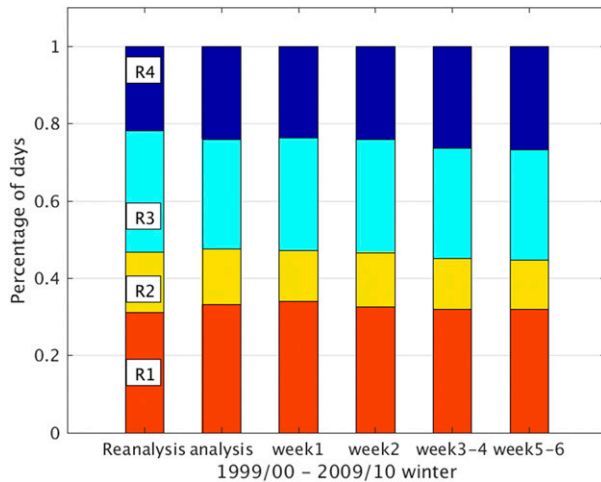


FIG. 3. Percentage of days assigned to each MERRA reanalysis regime (labeled R1 to R4) during the (left column) 1999–2010 reanalysis period, and for the 1999/00–2009/10 four-member ensemble mean CFSv2 hindcasts as a function of lead time.

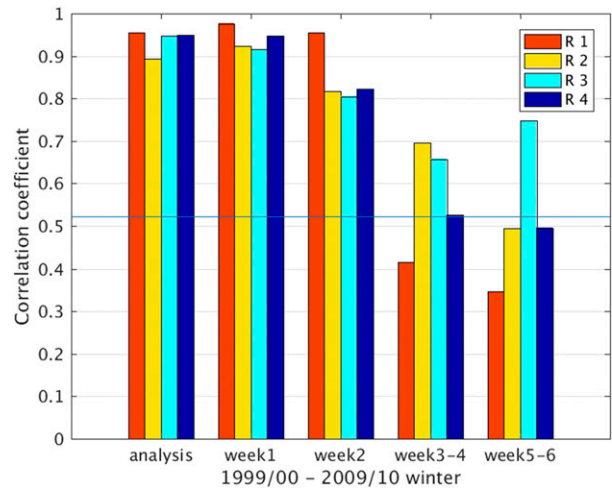


FIG. 4. Interannual correlations between the frequency of each regime in the CFSv2 hindcasts vs the MERRA reanalysis, as a function of lead time. The analysis refers to the hindcast initial condition (lead zero). The $r = 0.521$ 95% significance threshold is indicated.

variability, and 2) subtracting the mean 1982–2014 annual cycle on a daily basis and at each grid point. Low-pass filtering to remove periods less than about a week is commonly applied in weather regime studies to emphasize the low-frequency component (Straus et al. 2017). Four weather regimes were then identified over the (10° – 70° N, 150° – 40° W) sector by partitioning the 32 years \times 182 days (plus 8 leap days) of the dataset into four clusters using *K*-means clustering (Michelangeli et al. 1995). To reduce the dimensionality of the clustering problem and to ensure linear independence between input variables, principal component analysis was first performed on the data correlation matrix prior to clustering, retaining the first 12 PCs explaining 80.2% of the variance. This analysis is very similar to that of VRT, except that VRT used unfiltered daily Z500 data (and used 10-day running averages to compute the mean seasonal cycle). However, the resulting four spatial patterns obtained here using 5-day running averages prove to be very similar to those obtained by VRT based on unfiltered data.

b. Forecasts in weather-regime space

The intention of this paper is to use the low-dimensional space of MERRA reanalysis weather regimes to interpret CFSv2 forecast performance. To do this, Z500 maps were first constructed from the model hindcasts over the domain (10° – 70° N, 150° – 40° W), based on 1) daily forecast starts, 1 October–31 March, 1999/2000–2009/10; 2) 5-day running means over forecast lead days 1–45; and 3) the 4-member ensemble average. Next, anomaly maps were defined by subtracting the lead-dependent climatology for (1)–(3),

averaged over all 11 hindcast winters. No cross validation is applied. Finally, these forecast anomaly maps were projected onto the four reanalysis-regime *K*-means composite maps. The projection was carried out in physical space using a simple scalar product, and the Z500 forecast map then assigned to closest regime (i.e., smallest Euclidean distance). In addition, 3-day lagged ensemble mean hindcasts were similarly constructed, in which the ensemble size is augmented by averaging the 4 ensemble members from the forecast start day, together with those from the two previous days (i.e., over 12 members in all).

4. Results

a. Regime structures

The four regimes derived from MERRA reanalysis 5-day running average anomaly data are shown in Fig. 1 in terms of anomaly composites of 500-hPa height fields averaged over the days assigned to each *K*-means centroid. To aid interpretation these composite maps are plotted hemispherically, but the clusters were computed only over the sector (10° – 70° N, 150° – 40° W). The four anomaly patterns in Fig. 1 are visually almost indistinguishable from those derived by VRT using unfiltered daily data, with pattern correlations of 0.99, 0.96, 0.93, and 0.98, respectively. The minimal role played by the 5-day running averaging used here implies that the PC truncation acts to filter out the daily synoptic-scale variability in VRT.

Regimes 1, 3, and 4 consist of meridionally oriented ridge and trough anomalies resembling Rossby wave-trains, similar to those of the intermediate 10–30-day

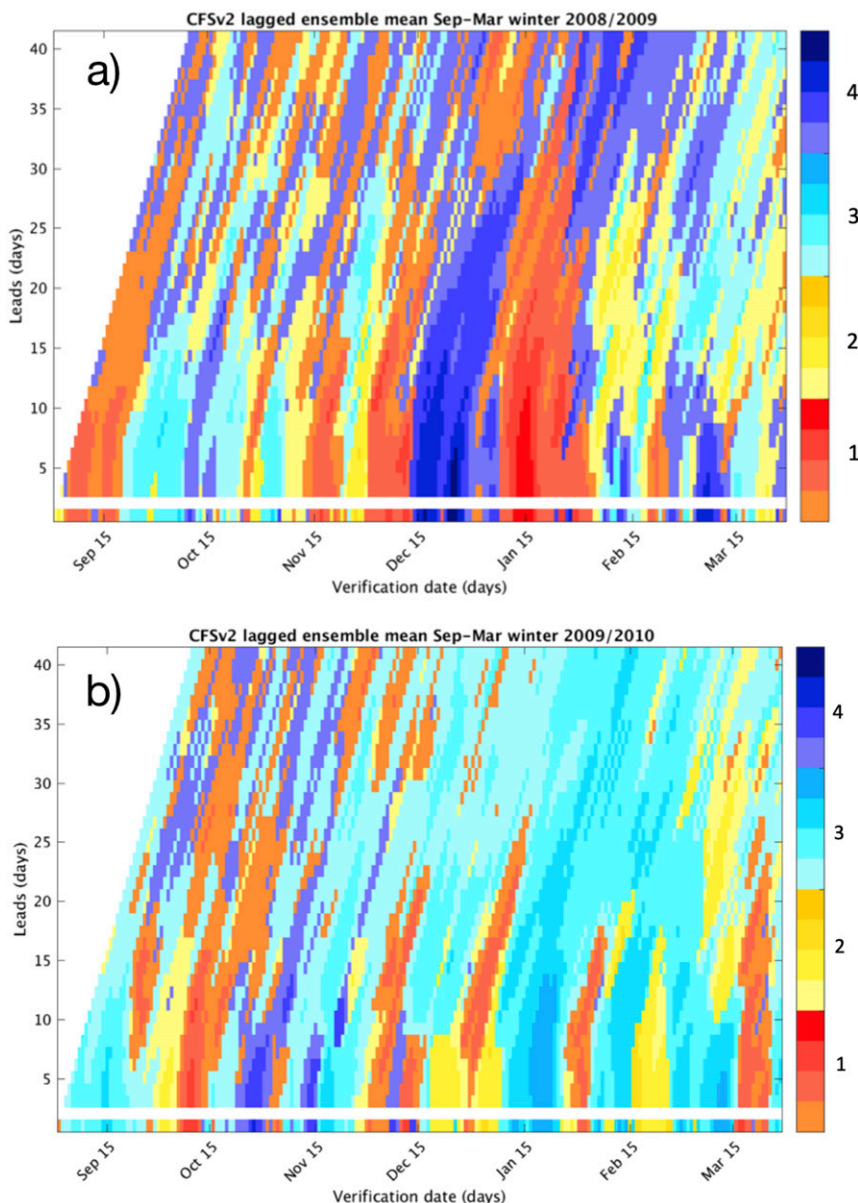


FIG. 5. Hindcast regime sequences vs daily lead time for the winters (a) 2008/09, and (b) 2009/10. Colors indicate the MERRA regime pattern that the hindcast 3-day lagged ensemble mean (12 members, smoothed with 5-day running averages) most closely resembles. Thus, the plotted sequence for lead = 3 days represents an average over days 1–5 of the forecast, and lead = 2 days is left white. The daily hindcast start date S is plotted on the abscissa, with lead time L on the ordinate. The daily evolution of the CFSv2 analysis (i.e., the lead-0 hindcast initial conditions) is shown along the bottom row of the plots ($L = 0$). For each regime color, the saturation provides an estimate of the similarity between the MERRA regime pattern and the hindcast ensemble mean.

time-scale waves identified by Blackmon et al. (1984). Regime 2, by contrast, is associated with a strong meridional pressure gradient between eastern North America and western regions of the North Atlantic coinciding with zonally elongated high and low height anomalies to the north and south of about 35°N , respectively; it is similar to

the negative phase of the North Atlantic Oscillation (NAO; e.g., Wallace and Gutzler 1981).

We refer here to the four regimes as the West Coast ridge (WCR; regime 1), Greenland high (GH; regime 2), Pacific trough (PT; regime 3), and Pacific ridge (PR; regime 4). These four regime structures are similar to

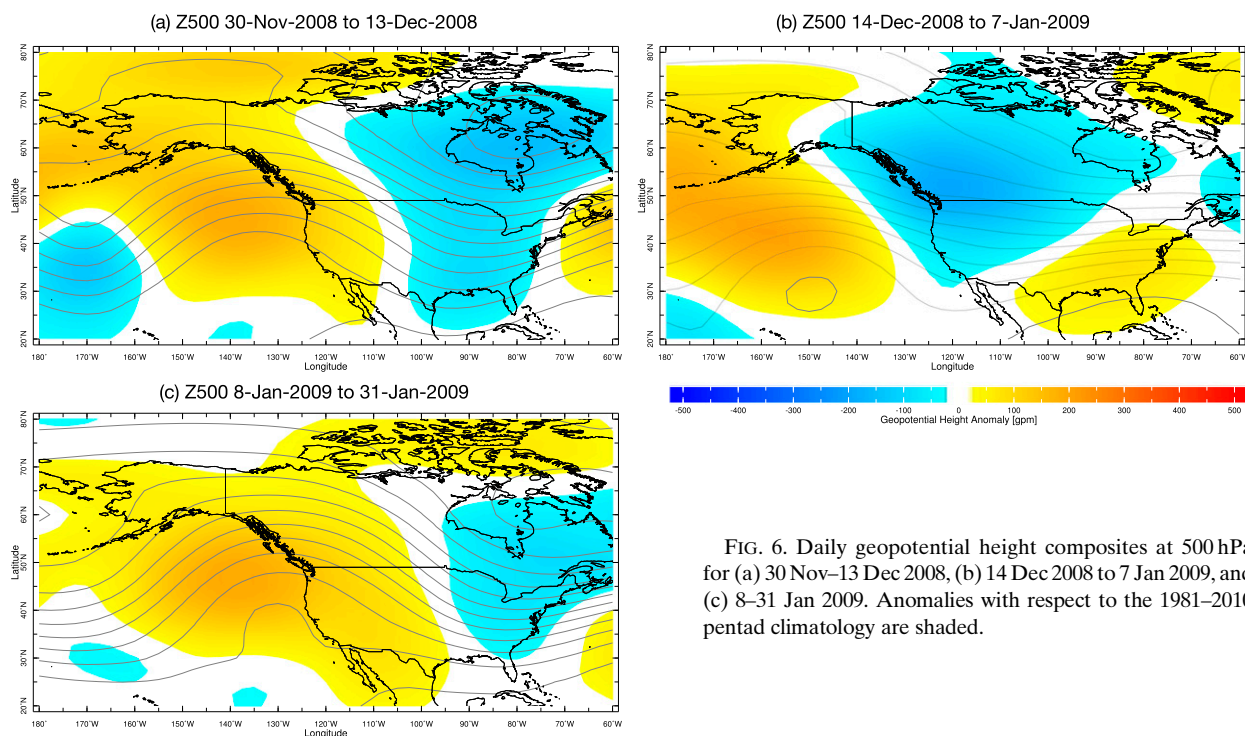


FIG. 6. Daily geopotential height composites at 500 hPa for (a) 30 Nov–13 Dec 2008, (b) 14 Dec 2008 to 7 Jan 2009, and (c) 8–31 Jan 2009. Anomalies with respect to the 1981–2010 pentad climatology are shaded.

those identified in the earlier Pacific-sector regime studies of Smyth et al. (1999) and Straus et al. (2007). The Pacific trough and Pacific ridge wavetrains resemble the Pacific–North American (PNA) teleconnection pattern and its reverse (i.e., negative) phase (reverse PNA or “RNA”) of Wallace and Gutzler (1981). Both regimes were identified as the sole Pacific sectorial regimes by Smyth et al. (1999) and both appear in the larger set of six sectorial regimes of Robertson and Ghil (1999). Together with the Greenland high (i.e., the negative phase of the NAO), regimes 2–4 correspond to the three hemispheric regimes identified by Cheng and Wallace (1993) using hierarchical clustering, and that were later reproduced by Smyth et al. (1999) using a Gaussian mixture model (see Fig. 9 of Smyth et al. 1999). We retain Straus et al. (2007)’s terminology for the Pacific trough, and name its reverse-phase counterpart the Pacific ridge for consistency. The West Coast ridge (regime 1) has less clear analogs in previous regime studies, although a very similar pattern has been implicated as a cause of recent droughts in California (S.-Y. S. Wang et al. 2017); VRT found that this regime has become more frequent over the last decade.

As documented in VRT, these 4 regimes strongly impact precipitation and near-surface temperature (their Figs. 5 and 6). As an example of how the regimes are related to impactful weather conditions, anomaly composites of an index of fire weather are plotted in

Fig. 2. We use the fire weather index (FWI) from the Global Fire Weather Database (GFWED), constructed by Field et al. (2015) and obtained via IRI Data Library. It is based on the Canadian Forest Fire Weather Index System, which uses local noon time temperature at 2 m, relative humidity at 2 m, and wind speed at 10 m, daily snow depth, and precipitation totaled over the previous 24 h. We use the version 2, which is based on MERRA2 reanalysis data and gridded CPC rainfall data. Figure 2 shows that the West Coast ridge is associated with substantially enhanced fire risk over large parts of the western United States.

b. Regime occurrence

Figure 3 shows the frequencies of each regime as a function of lead time (columns 2–6), averaged across the 11 hindcast winters, along with the average frequencies from the MERRA reanalysis for the 1999–2014 period (column 1). In MERRA, regimes 1 (WCR) and 3 (PT) are the most frequent (both 31% of days), with regimes 2 (GH) and 4 (PR) the least (16% and 22%). The second column shows the lead-0 forecast, labeled “analysis,” which exhibits very similar proportions to the MERRA. These percentages are remarkably well reproduced in the hindcasts at lead times up to 6 weeks, with surprisingly little model bias evident, even at 5–6 week lead. Larger biases were found in the ECMWF model by VRT, over the slightly longer 1995–2014 period.

The regime transitions are tabulated in [Table 1](#) for the MERRA data, and in [Table 2](#) for the 1–45-day 4-member ensemble mean hindcasts. Both tables are dominated by the diagonal self-transition probabilities, indicating a high degree of regime persistence, which can be expected given the 7-day low-pass filtering applied to the data. The persistence probabilities are quite similar for all four reanalysis regimes, with values slightly higher than found in [VRT](#) who used unfiltered daily data over the longer ECMWF 1995–2014 hindcast period. They are 10%–20% lower in the CFSv2 hindcasts, with regime 3 (PT) almost 20% less persistent. However, it should be recalled that the model regime occurrences are derived from 4-member ensemble mean Z500 anomaly fields, smoothed with a 5-day running average and then projected onto the MERRA regime centroid patterns; thus they do not solely represent the model's own intrinsic regime behavior, but also how well the hindcasts capture predictable observed regime evolution. None of the off-diagonal elements in either the reanalysis or model are statistically significant according to a χ^2 test.

[Figure 4](#) shows the variability of each regime's annual frequency (days per winter) in the hindcasts as a function of lead week, correlated with the observed number of days per winter spent in each regime. This measure reflects the CFSv2's ability to capture seasonal Z500 anomalies at subseasonal lead times, resulting either from slowly evolving surface boundary conditions (SST, sea ice, land surface), but also the seasonally rectified effect of subseasonal ones like MJO or sudden stratospheric warmings. At lead 0 ("analysis") and at week 1, the CFSv2 reproduces accurately the interannual variability in frequency of all four regimes, with an anomaly correlation exceeding about 0.9. There is a noticeable drop to about 0.8 at week 2, except for regime 1 (WCR), which maintains its week 1 skill level. Much larger drops are seen at weeks 3–4, especially for regime 1, where the correlation falls below the 95% statistical significance level, estimated as $r = 0.521$ from a Student's t test with 9° of freedom. A similar significance threshold was obtained using bootstrapping. Thus, the interannual variability in the West Coast ridge is captured very accurately in the forecasts up to week 2, with a large drop beyond 2 weeks, suggesting that its skill is largely associated with information in the atmospheric initial conditions, or that model errors accumulate more rapidly for this regime. In contrast, regime 3 (PT) sees the slowest falloff, which is consistent with its strong relationship with El Niño ([VRT](#)). It is notable that, except for regime 1, the interannual correlation reaches or exceeds the 95%

statistical significance level at weeks 3–4, and exceeds it at weeks 5–6 for regime 3 (PT).

c. Subseasonal forecast evolution

1) 2008/09 AND 2009/10 HINDCASTS

[Figure 5](#) shows the daily evolution of the CFSv2 hindcasts for the 2008/09 and 2009/10 hindcast winters, in terms of the color-coded MERRA regime that is closest to the forecast's ensemble mean 5-day running average Z500 anomaly pattern. Color saturation is used to give a qualitative estimate of how closely the hindcast ensemble mean matches the MERRA regime pattern. The ensemble average was augmented to form a 3-day lagged ensemble, consisting of 4 from the nominal start date, and 4 each from starts on the two previous days (12 members in all). The lagged ensemble yields slightly smoother plots than a 4-member ensemble. These plots can be read in two different ways. The lines sloping up to the right—seen immediately for the 1 September start—indicate the individual hindcast sequences $S + L$ for successive start days S . These sloping individual 45-day hindcast sequences are visible in the plots, especially beyond leads of about 10 days, and indicate regime persistence in the individual ensemble-mean hindcasts; however, there is quite a bit of variation in these sequences from day to day. Thus, although observed regime persistence is captured quite realistically in the hindcast sequences ([Table 2](#) versus [Table 1](#)), the hindcasts made on successive days (horizontal axis of plots) lack coherence at long leads, indicative of sensitivity to atmospheric initial conditions and loss of predictability. There is also a general tendency for the hindcast ensemble means to be less close to the regime centroids at longer lead, shown by less saturated colors.

The date on the abscissa is also the verification date for the hindcasts initialized on earlier days, and that are stacked vertically above each verification date for the increasing lead times (ordinate); perfect hindcasts will align as vertical bars. A similar type of plot was used to visualize CFSv2 precipitation forecasts at a locations by [Tippett et al. \(2015\)](#). The hindcasts in 2008/09 and 2009/10 generally exhibit qualitatively high skill (vertically aligned bars) in predicting the observed (analysis) regime sequence (bottom row) up to about 10 days ($L = 0 - 10$) in advance. Beyond $L = 10$, the model skill in this regime space generally becomes much less uniform. However, there are some persistent episodes of high skill extending to much longer lead times seen clearly in these plots. In addition, the day-to-day coherence of the lagged-ensemble forecasts tends to reflect their ability to capture the intraseasonal regime episodes,

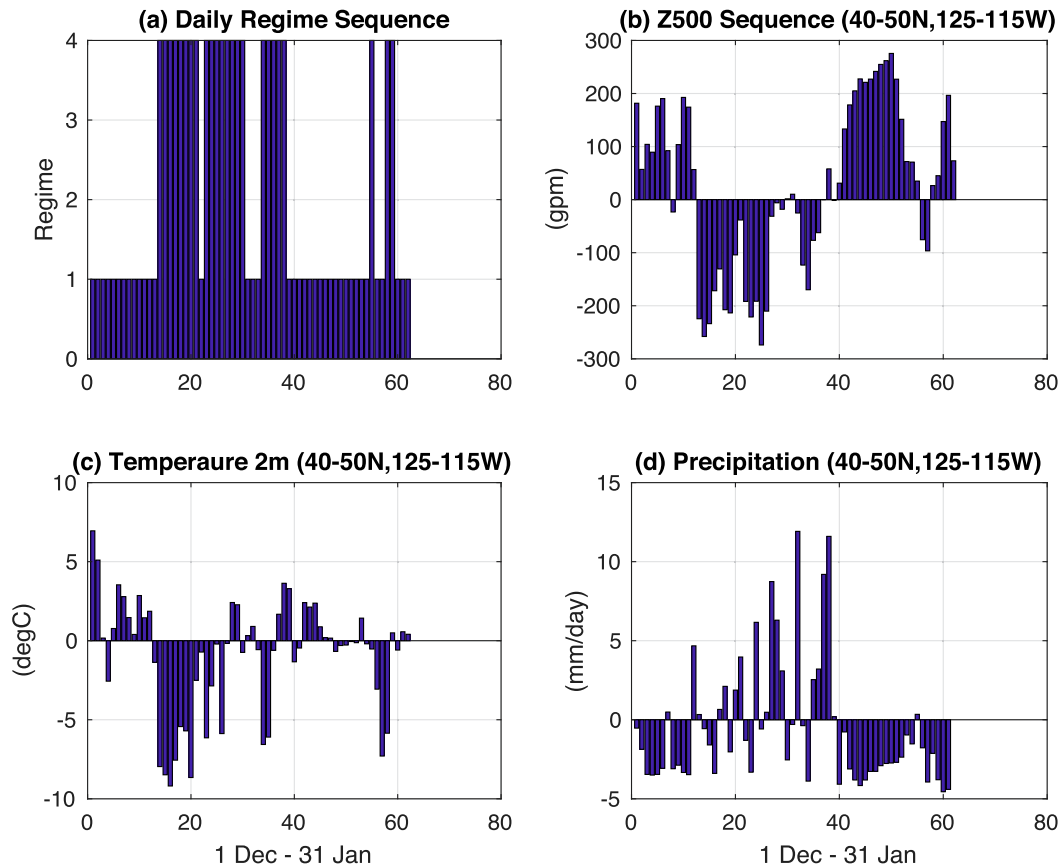


FIG. 7. Daily indices for the December 2008–January 2009 period, over the western United States (40° – 50° N, 125° – 115° W). (a) Weather regime sequence, (b) geopotential height at 500 hPa, (c) surface temperature (2 m), and (d) precipitation. Panels (b)–(d) show anomalies with respect to the 1981–2010 pentad climatology.

with a general loss of coherence after about 10 days and episodic coherence in the S2S range, such that forecast starts made several days apart still captured this event.

In 2008/09 (Fig. 5a) there was a major intraseasonal episode of the regime 4 (PR) from about 14 December to 8 January, with the remainder of December–February dominated by regime 1 (WCR). The contrast in 500-hPa geopotential height between the 14 December–8 January period and the periods either side is clearly seen in conventional composites plotted in Fig. 6; the ridge over the West Coast both before and after is replaced by a trough during this regime 4 episode, with the ridge displaced farther west over the Pacific. Figure 7 shows the regime sequence during December–January 2008/09, along with indices of observed anomalies of 500-hPa geopotential height, 2-m temperature, and precipitation, averaged over the box (40° – 50° N, 125° – 115° W) centered over the western United States. Impacts of the shift to regime 4 are expressed in low geopotential heights and temperatures, and higher precipitation, consistent with expectations (Robertson and Ghil 1999; VRT).

This intraseasonal episode and the timing of the transition from regimes 4 to 1 was remarkably well captured by the hindcasts up to 20–25 days in advance, seen as vertical stacks of regime color that indicate consistency of the hindcasts initialized on successively earlier days. Beyond 20 days, these stacks tend to tilt toward the right, indicating too much persistence and a failure of these longer lead forecasts to anticipate the early January transition from regime 4 to 1.

While the MJO was relatively weak during December 2008, there was a strong development in phases 6–7 during early January, approximately coinciding with the transition to regime 1 (Fig. 8a). The CFSv2 ensemble mean forecast initialized on 17 December captures this development quite well up to about 4 weeks ahead (after which the propagation is underestimated and the event decays too rapidly in the second half of January, Fig. 8b). MJO phases 6–7 corresponds to enhanced convection over the west Pacific, which is consistent with trough over North Pacific and development of ridge conditions over

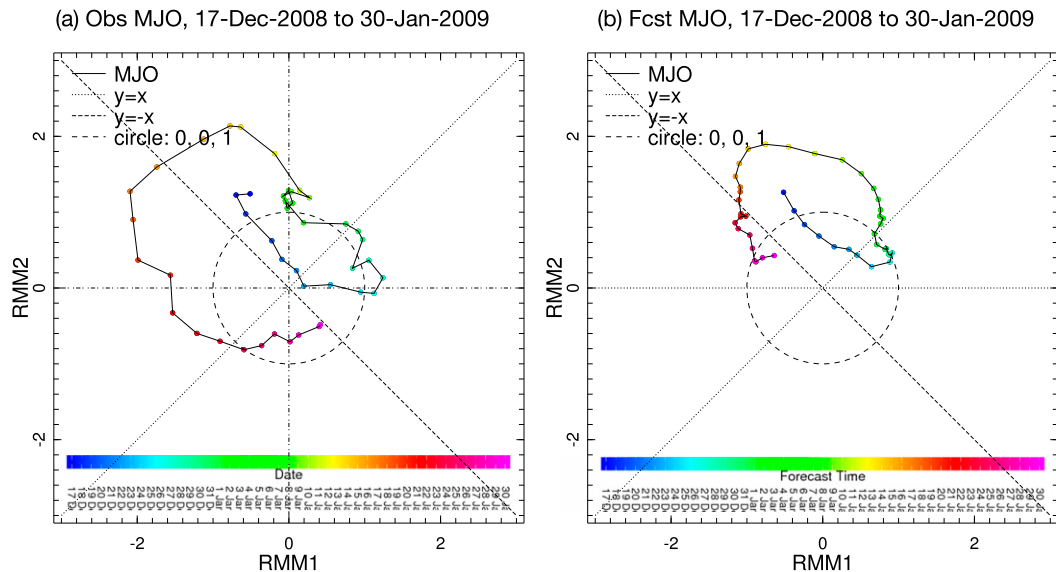


FIG. 8. Wheeler–Hendon MJO diagrams for (a) observed evolution 17 Dec 2008 to 30 Jan 2009, and (b) CFSv2 4-member ensemble mean forecast, initialized 17 Dec 2008. The color bar indicates the day corresponding to each point.

the western United States, as in regime 1 (Lin and Brunet 2018).

A major sudden stratospheric warming (SSW) event took place in late January/early February 2009 (Harada et al. 2010), approximately corresponding with regime 1 (WCR). A pronounced stratospheric wave-2 pattern with deep troughs over eastern Canada and Siberia, and strong ridging over Alaska and Europe are consistent with the regime 1 pattern over North America. According to the analysis of Harada et al. (2010), the upper-troposphere ridge over Alaska played an important role in the development of the SSW through upward propagation of wave packets from the Alaska region. Our analysis indicates that this SSW development was associated with regime 1, suggesting that it may have been predictable up to a month in advance based on Fig. 5a. However, the WCR is not a common precursor of SSWs (Domeisen et al. 2020), and Karpechko (2018) found that this SSW was only predictable up to 10 days in advance in the ECMWF model. There was also no obvious subsequent impact of the SSW on our North American weather regimes in February–March 2009.

The 2009/10 winter (Fig. 5b) was strikingly different from 2008/09, with the hindcasts dominated by regime 3 (PT), reflecting the concurrent El Niño condition, especially from November–February. Episodes of regime 3 in late November 2009 were well forecast out to about 30 days, and in January 2010 to 45 days ahead. However, it is clear from Fig. 5b that the longer lead forecasts exaggerated the number of days in regime 3, particularly in the second half of the winter. Regime frequencies for

the 2009/10 winter as a whole are summarized in Fig. 9 over different ranges of lead time. Compared to the analysis (i.e., lead-0 forecast), only the week 1 hindcasts approximately represent the proportions of the 4 regimes across the winter. Regime 3 is considerably overrepresented at week 2 lead, and the problem worsens at longer lead times. The different character of the regime forecasts in 2008/09 and 2009/10 can be partly traced to ENSO, with the tropical Pacific dominated by La Niña and El Niño conditions, respectively. A smaller SSW event occurred in late January/early February 2010 (Dörnbrack et al. 2012), though its connections with the regime sequence is unclear.

2) 2015/16 FORECASTS

Figure 10 shows the forecast evolution of the 2015/16 winter, which was characterized by one of the strongest El Niño events on record, and was marked by poor performance of the CFSv2 forecasts at seasonal scale over the U.S. Southwest (Chen and Kumar 2018). In this case, 16 ensemble member forecasts are available each day, so the 3-day lagged ensemble contains 48 members (compared to 12 for the hindcasts). These forecasts were bias corrected by subtracting the lead-dependent weekly hindcast (1999–2010) climatology. As in 2009/10, the CFSv2 forecasts were dominated by regime 3 (PT), as might be expected. However to an even greater extent than in 2009/10, the El Niño–typical Pacific trough regime was exaggerated in the forecasts, as summarized in Fig. 11. Despite the strength of the 2015/16 El Niño SST anomalies, the expected El Niño precipitation signal

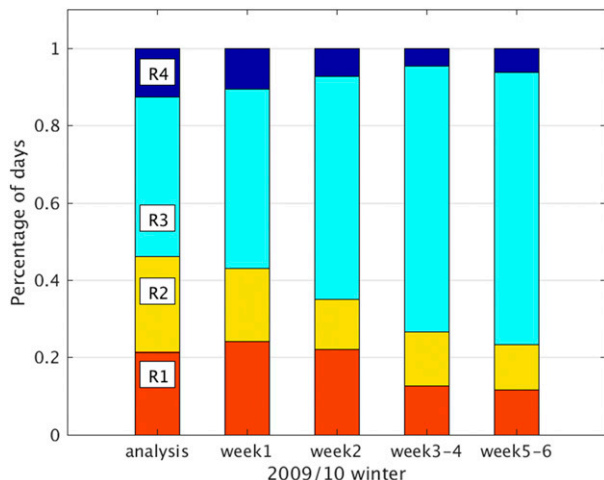


FIG. 9. Regime frequencies as a function of lead time for the 2009/10 winter.

over California of wet anomalies to the south and dry ones to the north failed to occur, and dry anomalies were observed over southern California during DJF 2015/16. This was not captured by the CFSv2 precipitation forecasts, which largely predicted the canonical wet-south/dry-north El Niño pattern (Chen and Kumar 2018).

According to Fig. 10, the CFSv2 forecasts during DJF 2015/16 consistently indicated regime 3 (PT), as would be expected associated with El Niño, at lead times of 45 to 7–14 days ahead. However, at shorter lead times the forecasts started to be much more heterogeneous,

particularly with the emergence of regime 4 (PR) in the second half of December 2015, and regime 1 (WCR) in February; particularly the latter is associated with dry conditions over California. This is consistent with seasonally unpredictable atmospheric dynamics playing an important role in the seasonal outcome, as put forward by (Chen and Kumar 2018), consistent with the paradigm of “seasonal noise vs subseasonal signal” discussed by S. Wang et al. (2017). The prevalence of regime 4 (PR) during October–December 2015 may be associated with a pronounced MJO event in phases 2–3 during late September/early October (Lin and Brunet 2018), although the relationship with MJO is only partial.

Figures 5b and 9 show that a similar (though less extreme) situation arose in 2009/10 and that 2015/16 was not an unprecedented case. However, January 2010 was wet over the U.S. Southwest, as was the December–February seasonal precipitation anomaly, which in turn was correctly anticipated by seasonal forecast from IRI and elsewhere.

d. Skill evaluation

Figure 12 shows the anomaly correlation skill for each regime, based on the 1999–2010 hindcast period. These curves are based on the anomaly correlation coefficient of 5-day moving averages of regime counts, between MERRA and the forecast ensemble-mean; thus the 3-day lead value corresponds to the lead 1–5-day sum of days that the forecast ensemble mean is assigned to regime *k*, versus the count in the MERRA sequence on

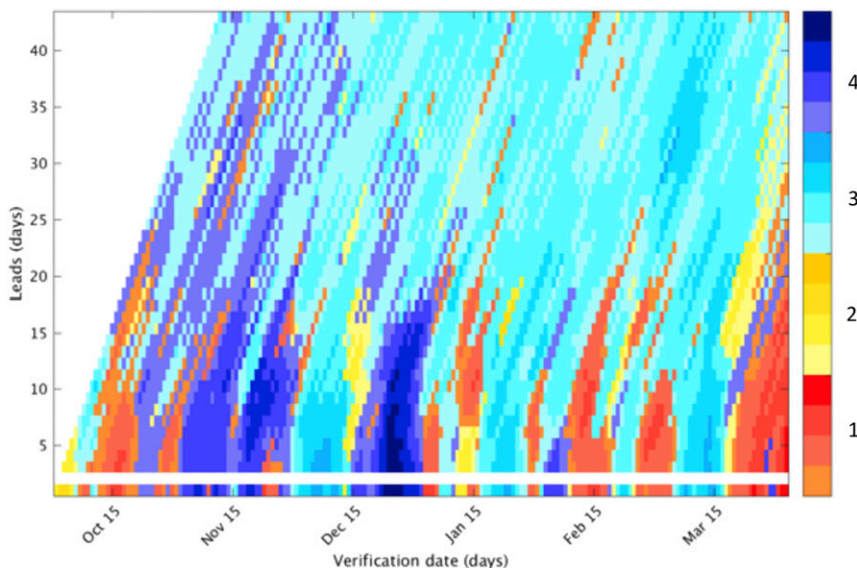


FIG. 10. Sequence of forecast regimes vs lead time for the 2015/16 winter. Details as Fig. 5, except that 16 forecast members are available each day, so the 3-day lagged ensemble contains 48 members.

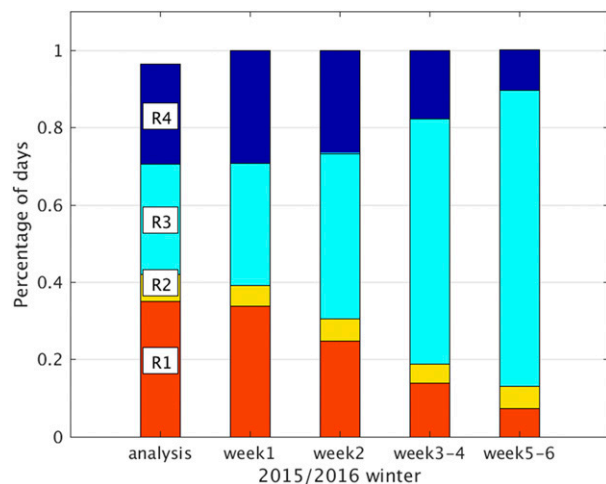


FIG. 11. Regime frequencies as a function of lead time for the 2015/16 winter.

those days. Averaged over the 11 hindcast winters, the skill is similar for the four regimes, with an anomaly correlation threshold of 0.5 exceeded for lead times less than about 12–14 days. This is consistent with the weekly averaged frequencies in Fig. 4, and is similar to that found using the ECMWF model by VRT. Figures 13a–c show the skill averaged over the individual winters plotted in Figs. 5 and 10. Large differences in the rate of dropoff of skill with lead time are visible for the individual regimes between these three winters. For example, the $r = 0.5$ threshold exceeds about 20 days for the regime 1 (WCR) in 2008/09, consistent with the highly skillful episode of regime 1 in late December 2008 seen (Fig. 5a). During 2015/16, the $r = 0.5$ threshold exceeds 15 days for regime 3 (PT) and reaches about 30 days for regime 4 (PR), consistent with the forecast evolution (Fig. 10), with regime 4 well predicted in October and November, and skillful intermittent episodes of regime 3. Figure 13d shows the interannual variability of skill averaged of week 3–4 lead time, which is seen to be substantial. We recall that anomaly correlations presented in Figs. 4 and 13 are not cross validated, and thus may be overestimated, noting the potential role of the mean bias correction for the small 11 year hindcast sample. The interannual variations in skill were not found to be significantly correlated with interannual variations in regime frequency or persistence.

5. Summary and conclusions

Large-scale weather regimes are used to diagnose subseasonal forecasts of wintertime (September–March) geopotential height fields over the North American sector, from the NCEP CFSv2 model hindcasts (1999–2010)

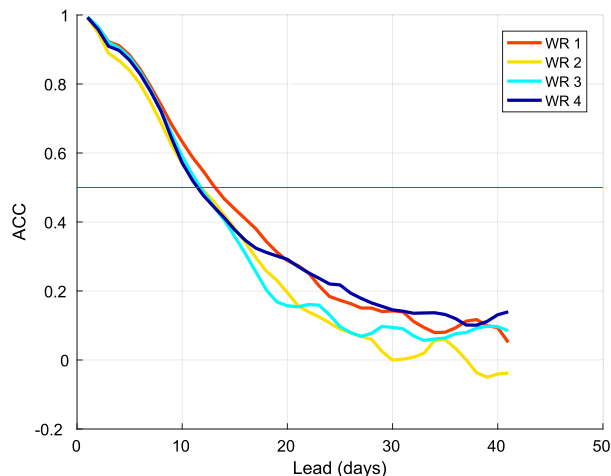


FIG. 12. Anomaly correlation skill for each regime, based on the 1999–2010 hindcast period. The 0.5 correlation value is indicated.

and forecasts in 2015/16. Four regimes derived from low-pass-filtered 500-hPa geopotential height daily MERRA reanalysis data (1982–2014) using *K*-means clustering are used as a low-dimensional basis for diagnosing the model's forecasts up to 45 days ahead (Fig. 1). The regime patterns obtained are almost identical to those obtained previously using unfiltered daily data (VRT). The West Coast ridge (regime 1) is shown to be associated with anomalously high fire-weather occurrence during winter over the 1982–2014 period (Fig. 2).

The CFSv2 hindcasts are shown to be remarkably unbiased, even up to 5–6 weeks ahead, in terms of their climatological weather regime frequencies (Fig. 3); that is to say, they do not drift erroneously toward particular circulation patterns at longer lead times.

Overall hindcast skill in regime space is found to be limited to 10–15 days ahead, in terms of anomaly correlation of 5-day averages of regime counts (Fig. 12), as well as interannual correlations in weekly regime frequency (Fig. 4). However, forecast skill is shown to be much higher for individual winters, reaching 30 days for regime 4 during 2015/16 (Fig. 13). Individual persistent episodes of high skill are identified graphically for individual winters using plots of forecast evolution, termed Chiclet Charts by Carbin et al. (2016). The CFSv2 lagged ensemble approach with initializations every day are plotted to reveal these episodes, using the winters of 2008/09, 2009/10 (Fig. 5), and 2015/16 (Fig. 10) as examples. A striking intraseasonal vacillation between the West Coast Ridge and Pacific Ridge patterns (regimes 1 and 4) during December–January 2008/09 (Figs. 5a, 6, 7) is shown to be predicted 20–25 days in advance (Fig. 5a). The transition back to regime 1 in early January 2009 coincides with an MJO event in phases 6–7

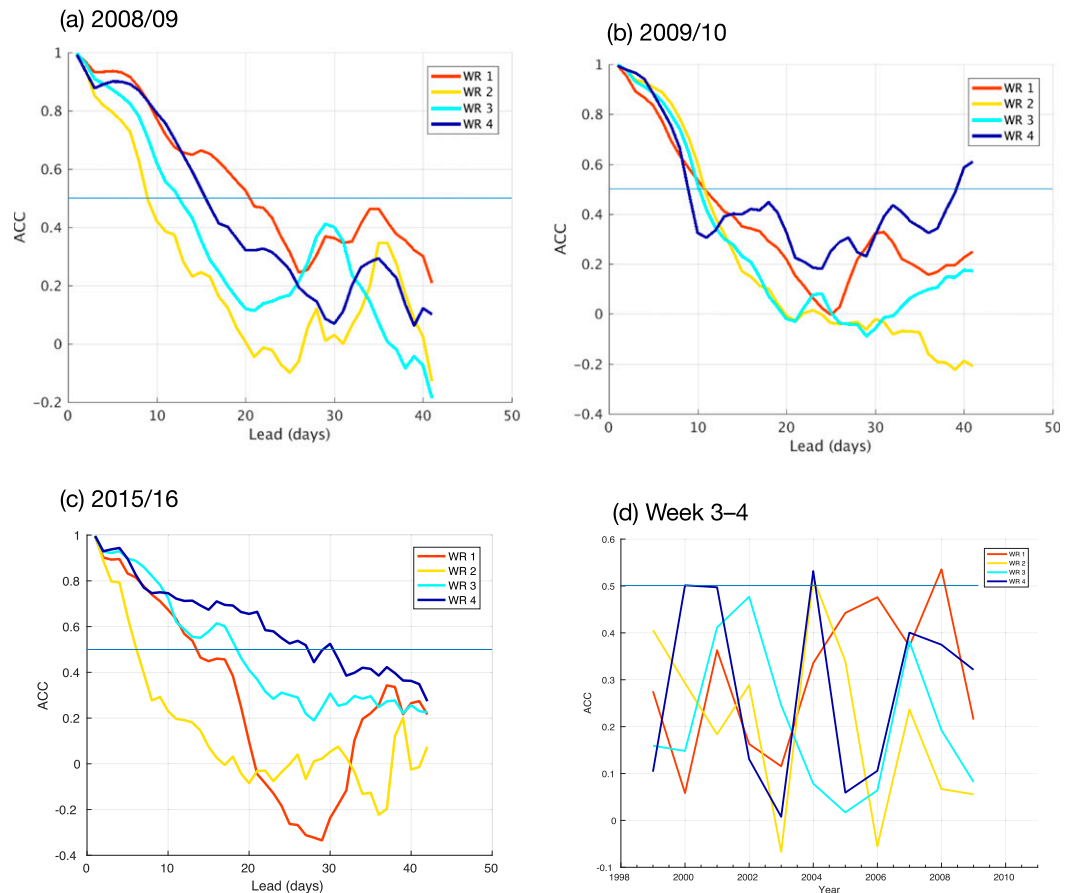


FIG. 13. Anomaly correlation skill for each regime, based on forecasts for the September–March period (a) 2008/09, (b) 2009/10, and (c) 2015/16. (d) Week 3–4 skill as a function of September–March year, where the year number given on the abscissa corresponding to September. The 0.5 correlation value is indicated.

that is relatively well captured by the CFSv2 forecast from mid-December (Fig. 8); this suggests that the MJO plays an active role in this regime transition. However, no such attribution is possible for the regime transition in early December 2008 when the MJO was weak.

The weather regime graphical forecast-evolution tool was then applied to diagnose the 2015/16 El Niño winter, which had been poorly predicted in terms of California precipitation at seasonal lead times. The Pacific trough regime 3 is shown to dominate the forecasts throughout much of the winter (Fig. 10), consistent with previous ENSO teleconnection studies. However, this regime is also shown to be highly over predicted beyond a week in advance, and especially beyond two weeks (Fig. 11). Weather-scale features start to dominate the forecast evolution diagram at lead times shorter than two weeks, consistent with the concept of “seasonal noise vs sub-seasonal signal” (S. Wang et al. 2017). Additional studies are needed to answer the question of why these subseasonal features infect the 2015/16 winter so strongly.

However, it is notable that a similar though less extreme situation took place during the weaker El Niño of 2009/10, with the Pacific trough overforecast at S2S lead times (Fig. 5b, 8).

The examples presented here illustrate the potential for “forecasts of opportunity” when subseasonal forecast skill is much higher than the average up to 3–4 weeks in advance and the utility of the weather regime concept as a means for identifying these windows in terms of the large-scale flow in midlatitudes and S2S sources of predictability. The analysis presented here highlights the 2008/09 case and further research is needed to demonstrate the general utility of the approach, to quantify the relationships with MJO phase and other S2S drivers, and to identify such windows in real time. The forecast-evolution plots or Chiclet Charts (Carbin et al. 2016) provide a new tool for visualization of the forecasts in real time, providing forecasters with an easy-to-read summary of the forecasts in terms of large-scale circulation patterns that

have occurred to date during the winter, and how the forecasts are evolving in lead time and for each successive initialization day. It can be argued that forecast-to-forecast consistency is a prerequisite for skill, and this format provides a way for forecasters to note it if it happens. The Chiclet Charts have been implemented at IRI during the 2018/19 and 2019/20 winters to monitor the CFSv2 forecasts in real time, with the forecasts updated every day.

This paper has used the weather regime concept to create a compact large-scale circulation diagnostic for S2S forecasts. A cluster analysis identified four recurrent patterns of observed geopotential height S2S variability, providing a “low-order” view of the forecasts on a 2D chart. The weather regimes provide a guide to the position of troughs and ridges over the Pacific–North American sector, how they evolve in the ensemble-mean forecasts, and give a concise picture of S2S forecast performance over a winter, either retrospectively or in real time. However, it is still an open question as to whether quasi-stationary circulation regimes exist on the S2S time scale from the dynamical systems perspective, which would imply additional predictability based on atmospheric midlatitude dynamics. It also remains to be shown how well these regime-based forecasts could be translated into surface weather probabilistic forecast skill.

Acknowledgments. We are grateful to the thoughtful comments from three anonymous reviewers and the editor, which substantially improved the manuscript. This work was supported by Grant NA16NWS4680014, as part of the NOAA Next Generation Global Prediction System (NGGPS) program. The CFSv2 reforecasts and forecasts were obtained from the S2S database, via IRI Data Library (<http://iridl.ldeo.columbia.edu>). S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF as an extension of the TIGGE database.

REFERENCES

- Becker, E. J., H. van den Dool, and M. Pena, 2013: Short-term climate extremes: Prediction skill and predictability. *J. Climate*, **26**, 512–531, <https://doi.org/10.1175/JCLI-D-12-00177.1>.
- Blackmon, M. L., Y.-H. Lee, and J. M. Wallace, 1984: Horizontal structure of 500 mb height fluctuations with long, intermediate and short time scales. *J. Atmos. Sci.*, **41**, 961–980, [https://doi.org/10.1175/1520-0469\(1984\)041<0961:HSOMHF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041<0961:HSOMHF>2.0.CO;2).
- Carbin, G. W., M. K. Tippett, S. P. Lillo, and H. E. Brooks, 2016: Visualizing long-range severe thunderstorm environment guidance from CFSv2. *Bull. Amer. Meteor. Soc.*, **97**, 1021–1031, <https://doi.org/10.1175/BAMS-D-14-00136.1>.
- Chen, M., and A. Kumar, 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>.
- , W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, <https://doi.org/10.1029/2007JD009132>.
- Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.*, **50**, 2674–2696, [https://doi.org/10.1175/1520-0469\(1993\)050<2674:CAOTNH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<2674:CAOTNH>2.0.CO;2).
- Domeisen, D. I. V., and Coauthors, 2020: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res. Atmos.*, **125**, e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Dörnbrack, A., M. C. Pitts, L. R. Poole, Y. J. Orsolini, K. Nishii, and H. Nakamura, 2012: The 2009–2010 Arctic stratospheric winter—General evolution, mountain waves and predictability of an operational weather forecast model. *Atmos. Chem. Phys.*, **12**, 3659–3675, <https://doi.org/10.5194/acp-12-3659-2012>.
- Ferranti, L., S. Corti, and M. Janousek, 2015: Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quart. J. Roy. Meteor. Soc.*, **141**, 916–924, <https://doi.org/10.1002/qj.2411>.
- Field, R. D., and Coauthors, 2015: Development of a global fire weather database. *Nat. Hazards Earth Syst. Sci.*, **15**, 1407–1423, <https://doi.org/10.5194/nhess-15-1407-2015>.
- Harada, Y., A. Goto, H. Hasegawa, N. Fujikawa, H. Naoe, and T. Hirooka, 2010: A major stratospheric sudden warming event in January 2009. *J. Atmos. Sci.*, **67**, 2052–2069, <https://doi.org/10.1175/2009JAS3320.1>.
- 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).
- Karpechko, A., 2018: Predictability of sudden stratospheric warmings in the ECMWF extended-range forecast system. *Mon. Wea. Rev.*, **146**, 1063–1075, <https://doi.org/10.1175/MWR-D-17-0317.1>.
- Kim, H., F. Vitart, and D. E. Waliser, 2018: Prediction of the Madden–Julian oscillation: A review. *J. Climate*, **31**, 9425–9443, <https://doi.org/10.1175/JCLI-D-18-0210.1>.
- Lin, H., and G. Brunet, 2018: Extratropical response to the MJO: Nonlinearity and sensitivity to the initial state. *J. Atmos. Sci.*, **75**, 219–234, <https://doi.org/10.1175/JAS-D-17-0189.1>.
- Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.*, **52**, 1237–1256, [https://doi.org/10.1175/1520-0469\(1995\)052<1237:WRRASQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1237:WRRASQ>2.0.CO;2).
- Moron, V., A. W. Robertson, and M. N. Ward, 2006: Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Mon. Wea. Rev.*, **134**, 3248–3262, <https://doi.org/10.1175/MWR3252.1>.
- Palmer, T. N., 1988: Medium and extended range predictability and stability of the Pacific/North American mode. *Quart. J. Roy. Meteor. Soc.*, **114**, 691–713, <https://doi.org/10.1002/qj.49711448108>.
- Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, <https://doi.org/10.1175/BAMS-D-18-0270.1>.
- Reinhold, B., and R. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.*, **110**, 1105–1145, [https://doi.org/10.1175/1520-0493\(1982\)110<1105:DOWRQS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2).

- Riddle, E. E., M. Stoner, N. Johnson, and M. L'Heureux, 2013: The impact of the MJO on clusters of wintertime circulation anomalies over the North American region. *Climate Dyn.*, **40**, 1749–1766, <https://doi.org/10.1007/s00382-012-1493-y>.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
- Robertson, A. W., and M. Ghil, 1999: Large-scale weather regimes and local climate over the western United States. *J. Climate*, **12**, 1796–1813, [https://doi.org/10.1175/1520-0442\(1999\)012<1796:LSWRAL>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1796:LSWRAL>2.0.CO;2).
- , Y. Kushnir, U. Lall, and J. Nakamura, 2015: Weather and climatic drivers of extreme flooding events over the Midwest of the United States. *Extreme Events: Observations, Modeling and Economics*, M. Chavez, M. Ghil, and J. Urrutia-Fucugauchi, Eds., John Wiley and Sons, 115–124.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Smyth, P., M. Ghil, and K. Ide, 1999: Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, **56**, 3704–3723, [https://doi.org/10.1175/1520-0469\(1999\)056<3704:MRINHH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<3704:MRINHH>2.0.CO;2).
- Straus, D. M., S. Corti, and F. Molteni, 2007: Circulation regimes: Chaotic variability versus SST-forced predictability. *J. Climate*, **20**, 2251–2272, <https://doi.org/10.1175/JCLI4070.1>.
- , F. Molteni, and S. Corti, 2017: Atmospheric regimes: The link between weather and the large-scale circulation. *Nonlinear and Stochastic Climate Dynamics*, C. L. E. Franzke and T. J. O'Kane, Eds., Cambridge University Press, 105–135.
- Tippett, M. K., M. Almazroui, and I.-S. Kang, 2015: Extended-range forecasts of areal-averaged rainfall over Saudi Arabia. *Wea. Forecasting*, **30**, 1090–1105, <https://doi.org/10.1175/WAF-D-15-0011.1>.
- Vigaud, N., A. W. Robertson, and M. K. Tippett, 2018: Predictability of recurrent weather regimes over North America during winter from submonthly reforecasts. *Mon. Wea. Rev.*, **146**, 2559–2577, <https://doi.org/10.1175/MWR-D-18-0058.1>.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, <https://doi.org/10.1002/qj.2256>.
- , and Coauthors, 2017: 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>.
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2).
- Wang, S., A. Anichowski, M. K. Tippett, and A. H. Sobel, 2017: Seasonal noise versus subseasonal signal: Forecasts of California precipitation during the unusual winters of 2015–2016 and 2016–2017. *Geophys. Res. Lett.*, **44**, 9513–9520, <https://doi.org/10.1002/2017GL075052>.
- Wang, S.-Y. S., J.-H. Yoon, E. Becker, and R. Gillies, 2017: California from drought to deluge. *Nat. Climate Change*, **7**, 465–468, <https://doi.org/10.1038/nclimate3330>.