

Prediction Skill of the 2012 U.S. Great Plains Flash Drought in Subseasonal Experiment (SubX) Models

ANTHONY M. DEANGELIS,^a HAILAN WANG,^b RANDAL D. KOSTER,^c SIEGFRIED D. SCHUBERT,^{a,c}
YE HUI CHANG,^{c,d} AND JELENA MARSHAK^c

^a *Science Systems and Applications, Inc., Lanham, Maryland*

^b *Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland*

^c *Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, Maryland*

^d *Goddard Earth Sciences Technology and Research, Morgan State University, Baltimore, Maryland*

(Manuscript received 18 November 2019, in final form 26 March 2020)

ABSTRACT

Rapid-onset droughts, known as flash droughts, can have devastating impacts on agriculture, water resources, and ecosystems. The ability to predict flash droughts in advance would greatly enhance our preparation for them and potentially mitigate their impacts. Here, we investigate the prediction skill of the extreme 2012 flash drought over the U.S. Great Plains at subseasonal lead times (3 weeks or more in advance) in global forecast systems participating in the Subseasonal Experiment (SubX). An additional comprehensive set of subseasonal hindcasts with NASA's GEOS model, a SubX model with relatively high prediction skill, was performed to investigate the separate contributions of atmospheric and land initial conditions to flash drought prediction skill. The results show that the prediction skill of the SubX models is quite variable. While skillful predictions are restricted to within the first two forecast weeks in most models, skill is considerably better (3–4 weeks or more) for certain models and initialization dates. The enhanced prediction skill is found to originate from two robust sources: 1) accurate soil moisture initialization once dry soil conditions are established, and 2) the satisfactory representation of quasi-stationary cross-Pacific Rossby wave trains that lead to the rapid intensification of flash droughts. Evidence is provided that the importance of soil moisture initialization applies more generally to central U.S. summer flash droughts. Our results corroborate earlier findings that accurate soil moisture initialization is important for skillful subseasonal forecasts and highlight the need for additional research on the sources and predictability of drought-inducing quasi-stationary atmospheric circulation anomalies.

1. Introduction

Droughts that develop on time scales of several weeks to a few months, known as flash droughts, have been receiving increased attention by the scientific community (e.g., Svoboda et al. 2002; Otkin et al. 2018; Pendergrass et al. 2020). Characterized by a lack of precipitation, enhanced evapotranspiration, and a rapid decline in soil moisture, flash droughts can disrupt agricultural production, deplete water resources, and impair natural ecosystems and vegetation (Smith and Katz 2013; Otkin et al. 2016; Choat et al. 2018; He et al. 2018, 2019; Vogel et al. 2019). The central United States, extending from the southern Plains to the Midwest, has been identified as a region that is particularly prone to

flash drought (Christian et al. 2019; Koster et al. 2019; Chen et al. 2019). While precipitation deficits are a major factor contributing to flash drought onset in this region, abnormally high temperatures and evaporative demand are often associated with drought intensification as well (Otkin et al. 2013; Livneh and Hoerling 2016; Koster et al. 2019; Chen et al. 2019). This is partly because the central United States is a region where dry soils can strongly alter the land surface energy budget and enhance surface temperatures (Koster et al. 2009a).

One of the best recent examples of a flash drought affecting the central United States occurred in the summer of 2012. During this season, the region transformed from near-average to extremely dry conditions within a 3-month period from May through July. As seen in Fig. 1, the event was fueled, to first order, by persistent monthly precipitation deficits and extremely warm temperatures, both of which led to a rapid drying of the soil.

Corresponding author: Anthony M. DeAngelis, anthony.deangelis@nasa.gov

DOI: 10.1175/JCLI-D-19-0863.1

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

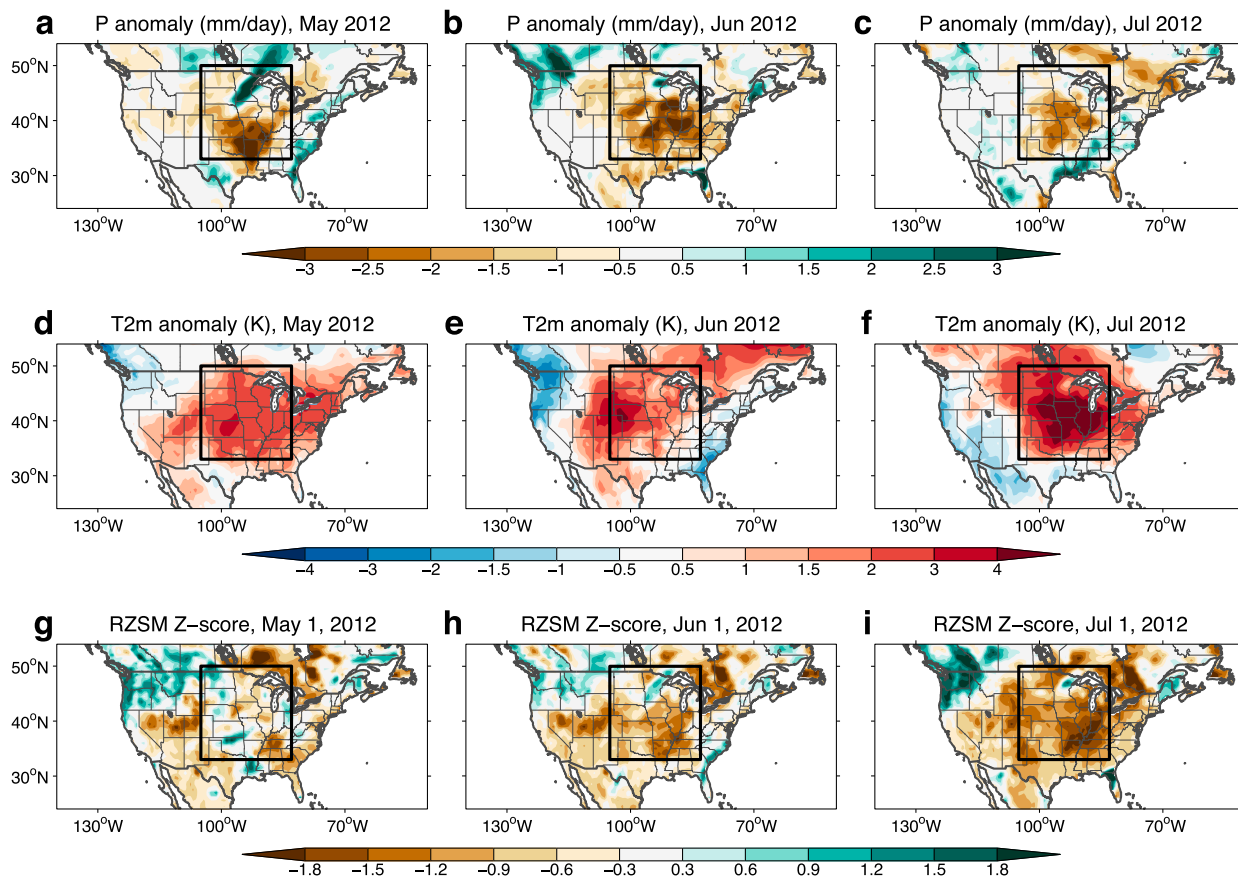


FIG. 1. Observed spatiotemporal evolution of anomaly fields during the 2012 Great Plains flash drought. (a)–(c) Anomalies of precipitation (P ; mm day^{-1}) for May, June, and July 2012, respectively. (d)–(f) Anomalies of 2-m air temperature ($T2m$; K) for May, June, and July 2012. (g)–(i) The Z scores of root zone soil moisture (RZSM; unitless) on 1 May, 1 Jun, and 1 Jul 2012, respectively. All anomalies are from MERRA-2 and are computed relative to the 1999–2015 climatology. The corrected P product from MERRA-2, as described in section 2c, is shown in (a)–(c). The Z scores are computed as anomalies normalized by the standard deviation over 1999–2015. The box indicates the core drought region (105° – 83° W, 33° – 50° N), which is used for computing regional averages throughout the paper.

The drought had significant and widespread impacts on agriculture, vegetation, and the U.S. economy (Hoerling et al. 2014; Rippey 2015; Otkin et al. 2016), with agricultural losses amounting to more than \$30 billion (NCDC 2019).

The causes of the 2012 flash drought have been extensively studied yet remain an active area of research. While there is substantial evidence for a connection between North American drought and sea surface temperatures (SSTs) on seasonal and longer time scales (e.g., Schubert et al. 2016), SSTs appear not to have played a role in causing the 2012 drought (Hoerling et al. 2013; Wang et al. 2014). Instead, the drought appears to have been induced by a series of atmospheric circulation anomalies that originated from internal atmospheric variability (Wang et al. 2014). Figure 2 shows the temporal evolution of the event, highlighting two key periods in which the atmospheric circulation induced warm

and dry conditions: the first in mid-May and the second and more prominent event in late June through early July (Figs. 2a,b). The second period, in particular, was associated with an upper-level atmospheric ridge that remained nearly stationary over the United States for 2–3 weeks (Fig. 2h). This quasi-stationary atmospheric ridge induced an extreme heatwave and precipitation deficits, which were associated with an unusually dry atmosphere and led to a rapid additional drying of the soil (Figs. 2a–d). The atmospheric ridge was part of a quasi-stationary Rossby wave train (Hoskins and Ambrizzi 1993; Ambrizzi et al. 1995), that is, a series of atmospheric troughs and ridges that remained nearly stationary in geographic location but with energy slowly propagating from west to east along the zonal jet stream. The wave train of 2012 developed between Eurasia and the western North Pacific in early to mid-June and propagated eastward to North America over subsequent weeks (Figs. 2g,h).

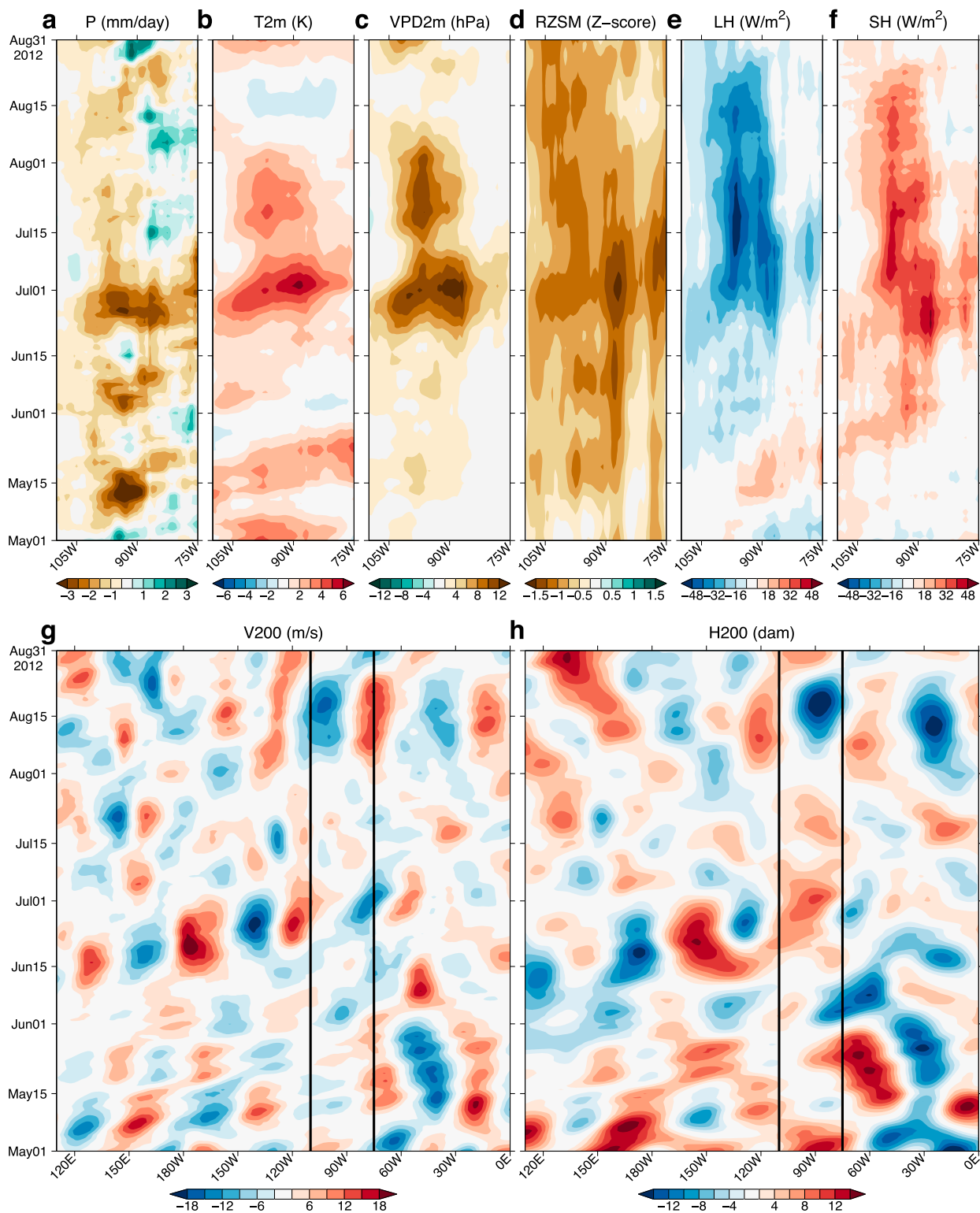


FIG. 2. Observed temporal evolution of the 2012 Great Plains flash drought. Hovmöller diagrams depict anomalies as a function of longitude and time from 1 May through 31 Aug for (a) P (mm day^{-1}), (b) T_{2m} (K), (c) vapor pressure deficit at 2 m (VPD_{2m} ; hPa) (d) RZSM (Z score), (e) surface latent heat flux (LH; W m^{-2}), (f) surface sensible heat flux (SH; W m^{-2}), (g) meridional wind at 200 hPa (V_{200} ; m s^{-1}), and (h) geopotential height at 200 hPa (H_{200} ; dam). All anomalies are averaged over 33° – 50°N (see box in Fig. 1) and are shown for the central United States in (a)–(f) and for a domain covering the western North Pacific to eastern North Atlantic in (g) and (h). Vertical lines in (g) and (h) indicate the central U.S. longitude domain that is shown in (a)–(f). All anomalies are from MERRA-2 and are computed as described in Fig. 1. A 10-day running mean has been applied for smoother plotting.

Another important element of the 2012 flash drought was land–atmosphere feedback (Roy et al. 2019; Herrera-Estrada et al. 2019). This aspect of the event is clearly demonstrated in Fig. 2. Despite extremely warm temperatures in March and below average precipitation in April of 2012 (not shown), soil moisture was only marginally drier than average at the start of the warm season in early May (Fig. 2d). Initially, the warm and dry atmospheric conditions in mid-May caused an increase in latent heat flux (hereafter LH), consistent with a sufficient supply of soil moisture and an energy-controlled evaporative regime (Figs. 2b,c,e). Upon further drying of the soil in late May, a transition to a soil moisture–controlled evaporative regime occurred with a reduction in LH and increase in sensible heat flux (hereafter SH) (Figs. 2e,f). This surface response (i.e., increased SH at the expense of evapotranspiration) intensified markedly with the arrival of the quasi-stationary ridge and associated extreme dry and warm atmospheric conditions in late June, and then remained nearly constant through mid-August. The persistent surface heat flux anomalies combined with very low atmospheric moisture from late June through mid-August likely helped to intensify and/or maintain the precipitation deficits and warm temperatures that led to the surface flux anomalies in the first place. In addition to the local land–atmosphere feedbacks just described, recent studies further show that dry land conditions, by reducing moisture transport to downwind areas, possibly aided in the spatial propagation of drought conditions in 2012 (Basara et al. 2019; Herrera-Estrada et al. 2019).

Dynamical seasonal forecasts issued in the spring of 2012 failed to predict intensifying drought conditions in the subsequent months (Hoerling et al. 2014), likely exacerbating the negative impacts of the severe flash drought that ensued. Here, we address the question of whether subseasonal forecasts, that is, forecasts for approximately three weeks to one month in the future, could have provided useful information if issued weekly in the summer of 2012. We do this by analyzing reforecasts (or hindcasts) from the Subseasonal Experiment (SubX), a multimodel global forecast ensemble that was recently developed to advance the research and operational capabilities of subseasonal prediction (Pegion et al. 2019).

Unlike weather and seasonal forecasting where skill is largely derived from a single source (atmospheric initial conditions and SSTs, respectively), the potential for skillful subseasonal predictions of summer flash drought in the central United States likely stems from multiple sources. One possible source of skill is the land, and in particular, soil moisture. In principle, knowledge of the initial soil moisture conditions at a given location can

lead to skillful subseasonal forecasts if the following are true: 1) the initial soil moisture is anomalously and sufficiently wet or dry, 2) the initial anomaly remains throughout the forecast (i.e., soil moisture memory is high), and 3) soil moisture conditions strongly influence surface heat fluxes and therefore temperature and precipitation over that location. The central United States is a region in which the above three criteria are often met during summer months (e.g., Koster and Suarez 2001; Koster et al. 2004a, 2009a). In fact, it has been demonstrated that when a global forecast model is initialized with accurate soil moisture anomalies over the central United States, in addition to the atmospheric and oceanic state, the subseasonal prediction of temperature and (to a lesser degree) precipitation generally improves (Koster and Suarez 2003; Koster et al. 2004b, 2010, 2011; Dirmeyer et al. 2018; Seo et al. 2019).

Another potential source of skill for the subseasonal prediction of flash droughts comes from quasi-stationary atmospheric circulation anomalies that can initiate or intensify drought conditions, such as that in late June of 2012 (Fig. 2). These circulation anomalies are often associated with quasi-stationary Rossby wave trains (e.g., Wang et al. 2014, 2017; Schubert et al. 2014). Boreal summer Rossby wave trains may initiate in a particular region (e.g., the Asian monsoon region) and travel eastward over the course of a few weeks (Jiang and Lau 2008; Schubert et al. 2011; Moon et al. 2013; Lopez et al. 2019). Thus, their development and future state may be inherently more predictable than random weather noise (Teng et al. 2013). Because these waves can induce extreme conditions in surface weather that last multiple weeks (e.g., Chen and Newman 1998), their adequate prediction in a forecast model can aid in the multiweek prediction of not only the intensity, but also the onset, of flash droughts. The prediction skill of Rossby wave trains in subseasonal forecast models is a topic that has previously received little attention among researchers [Chang et al. (2019) is a recent exception], and is one that demands further investigation.

In this study, we focus our analysis on the 2012 Great Plains flash drought due to its extreme impacts and its association with multiple flash drought mechanisms, but we also include analysis of central U.S. drought events in 2006 and 2011 to more robustly examine the impact of land initialization on forecasts. Our goal is to provide an overview of the subseasonal prediction skill of central U.S. summer flash droughts in SubX while highlighting the important roles of both land initialization and Rossby wave trains for the prediction skill. To further aid in the investigation of the 2012 drought, additional hindcasts from one of the SubX models, the Global Modeling and Assimilation Office (GMAO)–GEOS model, are generated.

TABLE 1. List of SubX models analyzed in this paper along with information about their hindcast ensemble size and initialization frequency. Models marked with an asterisk (*) provide soil moisture output. Note that the 46LCEM1 is not an official SubX participant (i.e., it does not provide real-time forecasts) but provides hindcast output on the IRI Data Library that archives SubX data. Additional details about the SubX models and references for individual models are provided in [Pegion et al. \(2019\)](#).

Model	Institution(s)	No. of members	Forecast length	Hindcast initializations
CESM-46LCEM1	National Center for Atmospheric Research and National Oceanic and Atmospheric Administration, Earth System Research Laboratory	10	45 days	Every Wednesday 1999–2015
ECCC-GEM	Environment and Climate Change Canada	4	32 days	Every 7 days 1995–2014
EMC-GEFS*	National Centers for Environmental Prediction, Environmental Modeling Center	11	35 days	Every Wednesday 1999–2016
ESRL-FIM*	National Oceanic and Atmospheric Administration, Earth System Research Laboratory	4	32 days	Every Wednesday 1999–2017
GMAO-GEOS*	National Aeronautics and Space Administration, Global Modeling and Assimilation Office	4	45 days	Every 5 days 1999–2016
NCEP-CFSv2	National Centers for Environmental Prediction	1	44 days	Every 6 h 1999–2017
NAVY-ESPC	Naval Research Laboratory	1	45 days	4 consecutive days each week 1999–2016
RSMAS-CCSM4*	National Center for Atmospheric Research, run at the University of Miami Rosenstiel School of Marine and Atmospheric Science	3	45 days	Every 7 days 1999–2016

Section 2 provides details about the SubX models, including the GMAO-GEOS and the additional hindcasts that are generated with this model. The datasets used for forecast verification and the analysis methodology used throughout the paper are also detailed in section 2. The results documenting the prediction skill of the entire SubX ensemble are presented in section 3, while results from the additional experiments with GMAO-GEOS are provided in section 4. Finally, a summary and discussion are given in section 5.

2. Data and methodology

a. SubX output

The SubX models analyzed in this paper are listed in Table 1. Throughout the remainder of this paper and in all figures, we refer to individual models by only their forecast system name for simplicity (these names are given in the first column of Table 1, following the hyphen). The models provide hindcasts over a recent 17-yr period (1999–2015) at an initialization frequency of at least once per week (see Table 1 for details). The number of ensemble members and method of initializing the members vary between models (see Table 1 and [Pegion et al. 2019](#)). Hindcasts range in length from 32 to

45 days depending on the model, allowing for an assessment of subseasonal (week 3–4) prediction skill. All models except Global Environmental Multiscale Model (GEM) and GEFS are coupled ocean–atmosphere systems and all modeling groups initialize the atmosphere and land for their hindcasts following their own procedures ([Pegion et al. 2019](#)). Additional details about the GEOS model are given in section 2b. For additional information about the other SubX models, including documentation for the models, see [Pegion et al. \(2019\)](#).

We use eight models that provide hindcast output to the International Research Institute (IRI) Data Library of Columbia University. Due to a technical error that affected land initialization during 2011 and 2012 in the GEFS model (Y. Zhu 2019, personal communication), we omit GEFS results from 2011 and 2012 from our analysis unless otherwise noted. GEFS output from other years is included in our analysis. For GEFS calculations that involve all years of the hindcast period (e.g., climatologies and overall prediction skill, described in section 2d), we remove the years 2011 and 2012 from the calculations unless noted otherwise.

Daily output of the following variables was obtained for all models: precipitation (hereafter P), 2-m air temperature (hereafter T2m), meridional wind at 200 hPa

(hereafter V200), and geopotential height at 200 hPa (hereafter H200). Four of the eight models also provide daily soil moisture output. Vertically integrated soil moisture (hereafter VISM) is provided for Flow-Following Icosahedral Model (FIM), GEOS, and CCSM4. For GEFS, volumetric soil moisture content is provided for four model layers. We analyze VISM directly from models that provide it, and for GEFS we compute VISM by scaling each layer's volumetric soil moisture content by the layer depth and summing over all layers. Daily averages over the period 0000–0000 UTC are provided for P , T2m, and VISM for all models except GEOS, for which the daily average is calculated over 2100 UTC of the previous day to 2100 UTC of the current day due to the initialization time of the model (see section 2b). For V200 and H200, daily fields are provided as the mean of instantaneous values at 0000, 0600, 1200, and 1800 UTC for all models. All SubX model output is available on a 1° latitude \times 1° longitude grid.

b. Additional hindcasts for GEOS

For a subset of initialization dates in 2012, additional hindcasts were performed with the GEOS model. The version of the model used for SubX and for the additional hindcasts performed here is the Subseasonal to Seasonal (S2S), version 2, prediction system (Molod et al. 2020). The model is a fully coupled land–atmosphere–ocean system and is run at an atmospheric resolution of 0.5° latitude \times 0.5° longitude with 72 hybrid sigma/pressure levels from the surface to 0.01 hPa. (Note that the output was regridded to a $1^\circ \times 1^\circ$ resolution for SubX; we analyze the regridded 1° output in section 3 for a fairer comparison with other SubX models, and we analyze the original 0.5° output in section 4 for a more direct comparison with the verification data.) The GEOS atmospheric model is similar to that used for the Modern-Era Retrospective Analysis for Research Applications, version 2 (MERRA-2) (Molod et al. 2015, 2020). The land surface model (LSM) in GEOS is the catchment-based LSM that is also used for MERRA-2 and is documented in Koster et al. (2000). The ocean component is the Modular Ocean Model, version 5 (MOM5), that was developed at the Geophysical Fluid Dynamics Laboratory (Griffies et al. 2005; Griffies 2012), and the sea ice component is the CICE 4.1 model developed by the Los Alamos National Laboratory (Hunke and Lipscomb 2010). For more details about the model components and parameterizations, see Molod et al. (2020).

The GEOS hindcasts prepared for SubX are initialized with MERRA-2 reanalysis data for the land and atmosphere. The ocean is initialized with output from a series of 5-day-long coupled ocean data assimilation runs (Molod et al. 2020). GEOS hindcasts are initialized

at 2100 UTC every 5 days, corresponding to the frequency of available oceanic initial conditions. Four ensemble members were produced for SubX, but here we extend the ensemble size to 35 members for a subset of the 2012 hindcasts in order to more robustly examine prediction skill. The additional ensemble members were produced by perturbing the atmospheric moisture, potential temperature, and horizontal wind component fields following the same approach used in generating the original SubX ensemble members (Molod et al. 2020). Hereafter, these 35-member GEOS hindcasts are referred to as control (CTL) hindcasts. (Note that in section 3, only the original four GEOS ensemble members prepared for SubX are used for a fairer comparison with other models).

To assess the impact of land surface initial conditions (including soil moisture) on hindcast prediction skill during 2012, we perform a parallel set of GEOS 35-member hindcasts in which we remove the influence of anomalies in land surface fields at initialization (hereafter referred to as noSFC hindcasts). This is accomplished by performing 35 simulations where the atmosphere and ocean are initialized from a particular calendar day during 2012 (e.g., 21 May 2012), but land surface fields over the entire globe are initialized from different years on that calendar day (e.g., 21 May 1981, 21 May 1982, etc.), thus effectively scrambling the initial land surface information.¹ The alternative years used to initialize the land surface are from 1981 to 2016, excluding 2012. The land surface fields that are modified encompass a variety of quantities that depict the land surface physical state, including various hydrologic and thermodynamic properties of the soil and surface skin temperature.

The 35-member CTL and noSFC hindcasts described above are performed for nine initialization dates, spaced 5 days apart, between 21 May and 30 June in 2012. These dates are chosen because the forecasts (falling in June and July 2012) coincide with the rapid intensification and peak of the 2012 Great Plains flash drought (Fig. 2). We also perform CTL hindcasts that are initialized every day between 5 and 15 June 2012 to assess the prediction skill of the quasi-stationary Rossby wave train in June 2012 (Fig. 2g).

c. Verification data

The prediction skill of the SubX hindcasts is evaluated using predominantly MERRA-2 reanalysis data (Bosilovich et al. 2015, 2017; Gelaro et al. 2017). Here we focus on daily

¹ The atmosphere is unperturbed in the 35 noSFC runs presented here. An additional set of noSFC runs with the atmosphere perturbed as in the CTL runs gave qualitatively similar results (not shown).

mean fields of P , T2m, V200, and H200 from MERRA-2. In the case of precipitation P , we use a “corrected” product, in which the precipitation generated by the atmospheric model underlying MERRA-2 was merged with gauge and satellite precipitation observations (Reichle and Liu 2014; Reichle et al. 2017). Additional variables related to the surface energy and moisture budgets are also analyzed from MERRA-2 to investigate the observed evolution of the 2012 flash drought and to evaluate the additional GEOS hindcasts performed for this study. These include LH, SH, vapor pressure deficit at 2 m (hereafter VPD2m, derived from T2m and 2-m dewpoint temperature), and root zone soil moisture (hereafter RZSM, the fraction of saturation in the top meter of soil). RZSM was generated using the corrected P product described above.

To assess the sensitivity of our results to verification dataset, we also analyze T2m data from the NOAA Climate Prediction Center (CPC) (<https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html>), which are derived directly from station data. Results using the CPC data are not explicitly shown but are discussed throughout the paper where appropriate.

To evaluate soil moisture initialization accuracy in the SubX ensemble, we utilize data from phase 2 of the North American Land Data Assimilation System (NLDAS-2) (Xia et al. 2012). NLDAS-2 is a collection of LSMs that were run offline and driven with common atmospheric forcing data to yield various surface fields over North America over the period from 1979 to present. The atmospheric forcing data are derived mostly from the North American Regional Reanalysis (NARR; Mesinger et al. 2006) as well as station and radar-based precipitation products (Xia et al. 2012). Because direct soil moisture observations are unavailable at the continental scale, NLDAS-2 is a reasonably independent estimate of soil moisture for the purposes of model evaluation. Four LSMs participated in NLDAS-2: Noah (Ek et al. 2003), Mosaic (Koster and Suarez 1996), the Variable Infiltration Capacity model (VIC) (Liang et al. 1994), and the Sacramento Soil Moisture Accounting model (SAC-SMA) (Burnash 1995). We analyze the mean VISM output from Noah, Mosaic, and VIC, which is available on a 1° latitude \times 1° longitude grid that matches that of the SubX output. We exclude SAC-SMA from our analysis due to differences in the way soil moisture is produced for this LSM (Xia et al. 2012) and due to differences in output format.

d. Processing of hindcast output and verification data

Throughout this paper, we analyze daily anomalies of the fields described in sections 2a and 2c that are relative to the daily climatology over the 17-yr period 1999–2015

(the SubX hindcast period).² For the MERRA-2 or NLDAS-2 verification data, anomalies are computed by simply subtracting the climatology for each calendar day from the corresponding daily values. For the SubX model hindcasts, one must account for climate drift when computing anomalies (e.g., Ma et al. 2014; Morcrette et al. 2018; Sun et al. 2018; Pegion et al. 2019). This is accomplished by removing model-dependent climatologies that are a function of both the day of year and length of time after initialization (or lead day) (see the appendix for details).

In this paper, we focus on the prediction skill of the ensemble mean of the hindcasts from each model. We also evaluate the overall performance of the SubX ensemble by computing a multimodel mean (hereafter MMM), which is derived from the ensemble-mean hindcasts of the individual models (see the appendix for details).

We also compute standardized anomalies (hereafter referred to as Z scores) of the soil moisture fields analyzed in this paper, in which anomalies are normalized by their corresponding standard deviation over the period 1999–2015. The Z scores for the SubX hindcast output are computed by normalizing the ensemble-mean anomalies by their corresponding standard deviations. The Z scores for the SubX MMM or NLDAS-2 mean are computed by averaging the Z scores for each model or LSM. In cases where the fields are smoothed (e.g., by applying a 7-day running mean), the standard deviations used to normalize anomalies are computed after first smoothing the data.

We assess the overall prediction skill of SubX models with temporal anomaly correlations, a metric commonly used for skill assessment (e.g., Koster et al. 2011; Becker et al. 2014; Kirtman et al. 2014; Pegion et al. 2019). Here, the anomaly correlations are computed by pooling 2-week-mean forecast anomalies that validate in June–August (JJA), for a given lead interval (e.g., forecast day 1–14), over the period 1999–2015 and correlating them with MERRA-2 anomalies for matching forecast validation dates (Note that SubX anomalies are first interpolated to the 0.5° latitude \times 0.625° longitude MERRA-2 grid for this computation, which is performed at every grid cell). We compute anomaly correlations corresponding to each 2-week lead interval (e.g., forecast days 1–14, 2–15, etc.) and use all 2-week forecasts for which the center of the 2-week period (defined as the seventh day)

² Anomalies computed for the GEM model are relative to the period 1999–2014 due to missing output for 2015, but this discrepancy is found to have a negligible effect on our conclusions (not shown).

Overall SubX prediction skill for JJA

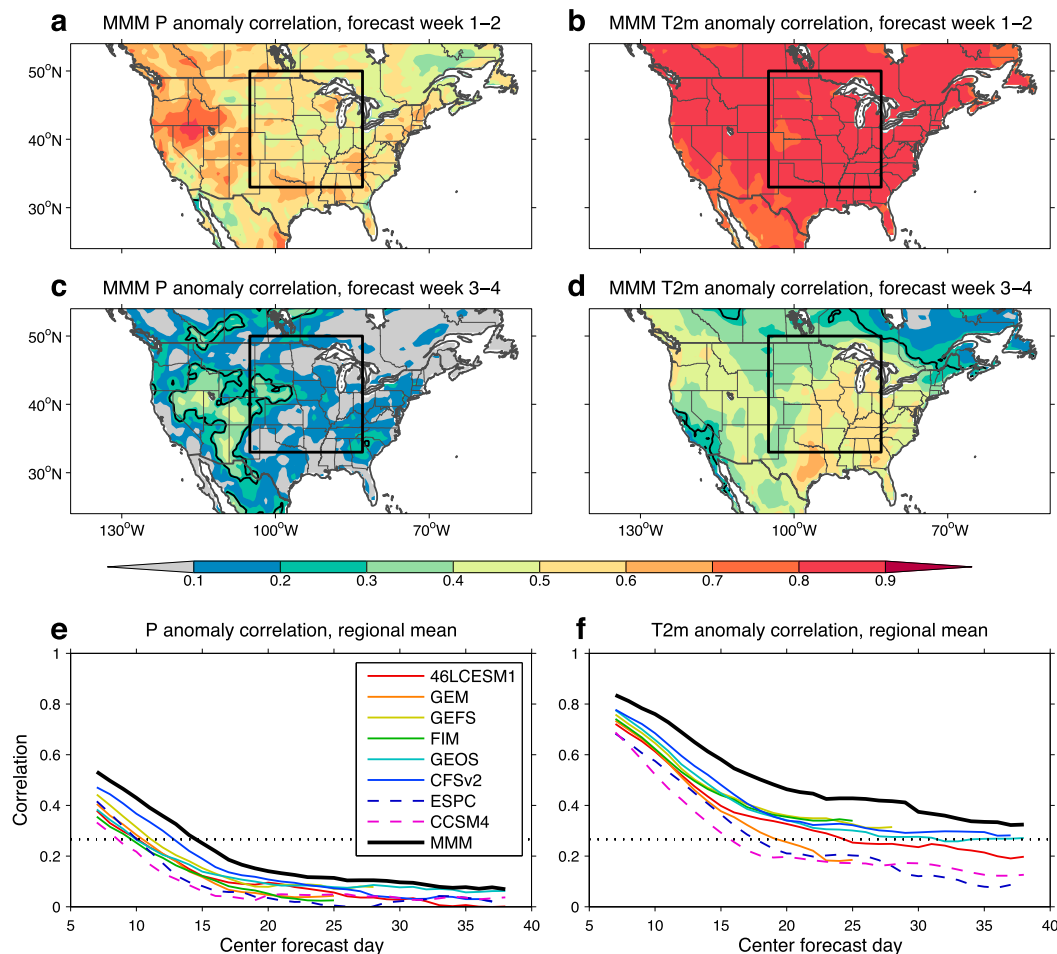


FIG. 3. Summer (JJA) prediction skill of SubX models over the period 1999–2015. Anomaly correlations between MERRA-2 and the SubX multimodel mean (MMM) are shown for JJA week 1–2 forecasts for (a) P and (b) T2m. (c),(d) Corresponding correlations for week 3–4 forecasts are shown. Correlations are averaged over the domain indicated with a box in (a)–(d) and shown for all 2-week lead intervals for all models for (e) P and (f) T2m. Regional-mean correlations in (e) and (f) are plotted at the center of the 2-week forecast period on the abscissa (e.g., the correlation for the day 1–14 forecast is plotted at day 7). Models that initialize soil moisture with an LSM-based product that utilizes a land model different from that in the full forecast system (which may result in reduced forecast skill; Koster et al. 2009b), are denoted with dashed lines in (e) and (f). A statistically significant correlation of 0.27 is indicated with a black contour on the maps and dotted black line on the graphs. Statistical significance is determined via a two-tailed t test with $\alpha = 0.05$ and degrees of freedom ($\nu = 53$) corresponding to approximately 13 forecasts per year and a decorrelation estimate of 4 weeks following Pegion et al. (2019).

falls in JJA. The anomaly correlations quantify the ability of a particular model's forecasts to collectively capture the observed subseasonal and interannual variability over the analysis period.

3. Prediction skill in the SubX ensemble

a. Overall prediction skill

We begin by presenting the overall prediction skill of SubX summer hindcasts for the full 17-yr period 1999–

2015 to provide context for the more focused analysis of the 2012 drought that follows. Anomaly correlations for the MMM are shown in the top panels of Fig. 3 for the week 1–2 and week 3–4 forecasts of P and T2m over the continental United States. Correlations for week 1–2 forecasts are quite high and are statistically significant over the entire domain (Figs. 3a,b), owing to the influence of atmospheric initial conditions on the forecasts (e.g., Lorenz 1982; Zhang et al. 2019). The week 3–4 correlations are considerably weaker (Figs. 3c,d), but

with substantially better skill for T2m, a common finding in subseasonal to seasonal prediction skill assessment (Koster et al. 2011; van den Hurk et al. 2012; Becker et al. 2014; Pegion et al. 2019). The skill of week 3–4 T2m forecasts is somewhat sensitive to verification dataset; while anomaly correlations exceed 0.5 over a substantial part of the central and southern United States when using MERRA-2 (Fig. 3d), they are lower (~ 0.3 – 0.5) when based on CPC (not shown). With both MERRA-2 and CPC, however, the correlations for T2m are statistically significant and can be physically justified. In particular, as mentioned in section 1, the central United States is a region in which soil moisture exerts a strong influence on evaporation and surface temperature and where subseasonal prediction benefits from accurate land surface initialization (e.g., Koster et al. 2011). Thus, it is likely that the significant correlations for T2m over this region reflect accurate land surface initialization in many SubX models [see section 3b(1) for a more in-depth analysis]. The reasons for the relatively high prediction skill for week 3–4 P forecasts in the western United States are less clear and require further investigation.

The regional-mean of the anomaly correlations over the central United States is shown for all 2-week lead intervals and all SubX models in the bottom panels of Fig. 3. The MMM outperforms any individual model for both P and T2m (Figs. 3e,f), consistent with the findings in Pegion et al. (2019) based on all seasons. The relatively good performance of CFSv2 compared to other models, especially for P (Fig. 3e), is likely due, in part, to the large effective ensemble size of this model (see appendix), which helps to improve the signal-to-noise ratio of the ensemble mean forecasts (Pegion et al. 2019). [Note that the CFSv2 prediction skill is degraded if only one initialization date per week, rather than the average of 7 days, is analyzed (not shown)]. For T2m, there is a substantial spread in anomaly correlations among models (Fig. 3f). (The spread is somewhat smaller when using CPC T2m data, which is mainly the result of weaker correlations among the more skillful models, not shown). Some of this model spread may be attributed to different approaches for initializing soil moisture [section 3b(1)].

b. Prediction skill of the 2012 flash drought

Ideally, the prediction skill assessment of flash droughts should focus on physical quantities that directly characterize drought conditions, including soil moisture and evaporative demand. Such an assessment, however, is hindered by the limited SubX output. Here, we instead focus on the key drivers of the 2012 Great Plains flash drought, that is, the extreme temperature and precipitation anomalies shown in Fig. 2. As seen in

Fig. 2, these P and T2m anomalies are coincident with responses in atmospheric moisture, soil moisture, and surface heat fluxes. An evaluation of P and T2m alone in SubX should thus provide a sufficient first-order indication of how well the 2012 flash drought is predicted.

To present a visual of how well the SubX models collectively forecast the anomalies in P and T2m, regional-mean and 2-week-mean forecast anomalies are shown for each model and every available initialization date in Fig. 4, and these are compared with corresponding anomalies from MERRA-2. Most 1–2 week forecasts from the SubX models are able to capture the observed periods of anomalously warm and dry conditions that occurred in the summer of 2012 (Figs. 4a,b, note how the dark colors, indicating stronger anomalies, align vertically on each of these panels). This high prediction skill is not surprising considering the first 1–2 weeks falls within the realm of numerical weather prediction, for which prediction skill is largely derived from initial atmospheric conditions.

Prediction skill for the week 3–4 forecasts (Figs. 4c,d), however, is much less consistent across models and is generally less satisfactory. The prediction skill for P is also notably inferior to that of T2m. Some models, however, exhibit satisfactory prediction skill. For example, most GEOS hindcasts initialized between mid-May and early July predict warm and dry anomalies 3–4 weeks later. These forecasts broadly agree with what occurred in nature despite a slight mismatch in the dates of the most intense anomalies. The GEOS prediction skill is investigated in greater depth in section 4. The CFSv2 also predicts positive T2m anomalies from early June through mid-July (Fig. 4d), in broad agreement with MERRA-2, but the amplitude of the anomalies is small due to the averaging over a large number of forecasts (see the appendix).

1) IMPORTANCE OF SOIL MOISTURE INITIALIZATION

It is important to understand the factors contributing to enhanced prediction skill in some SubX models during the 2012 flash drought and potentially other similar events. To the extent that land–atmosphere feedbacks contribute to the extreme temperatures and precipitation deficits during flash droughts, one should expect higher prediction skill from models that more accurately initialize the land and simulate land–atmosphere coupling. In this subsection, we examine the influence of soil moisture initialization, in particular, on the prediction skill. To do this, we compare soil moisture (specifically VISM) from the SubX models that provide output with NLDAS-2 verification data. For this evaluation, it is important to compare Z scores of VISM, as soil moisture

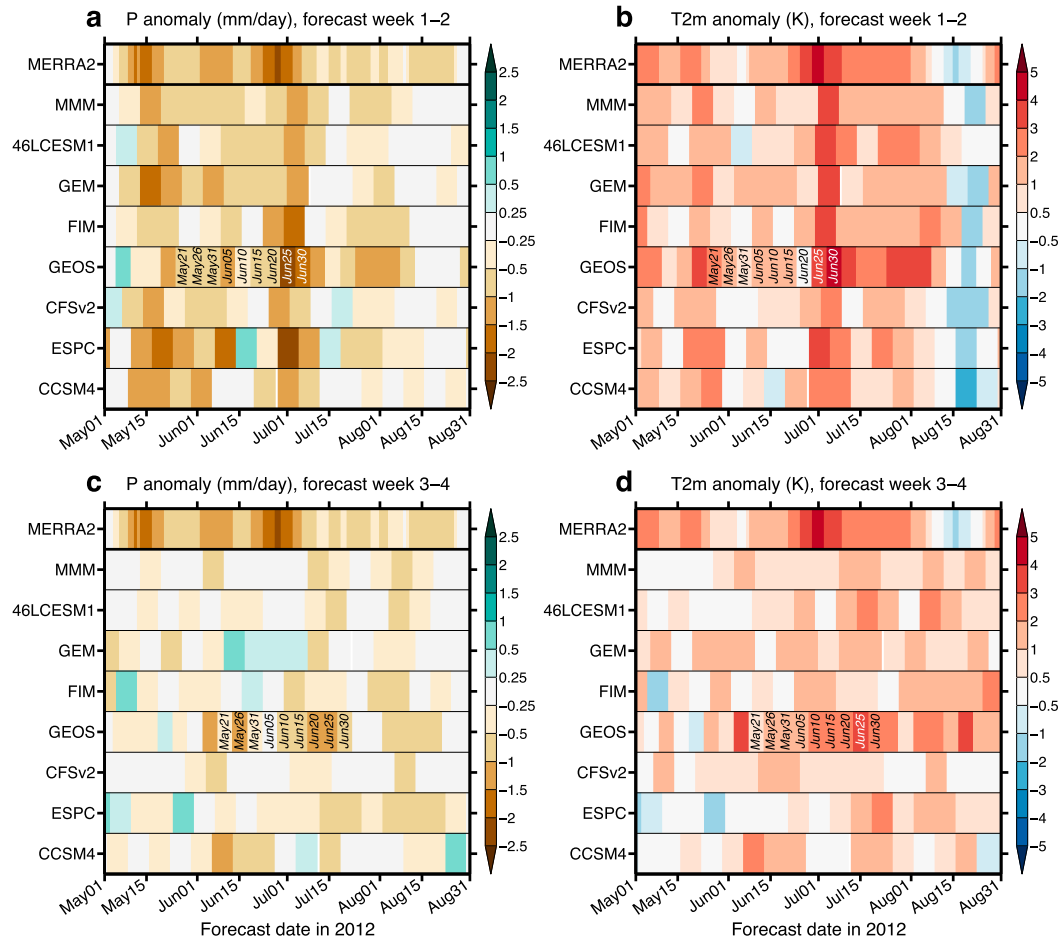


FIG. 4. Prediction skill of the 2012 Great Plains flash drought in SubX models. Regional-mean anomalies of (a),(c) P (mm day^{-1}) and (b),(d) $T2m$ (K) are shown for MERRA-2 and SubX hindcasts for the period 1 May through 31 Aug 2012. Anomalies are averaged over $105^{\circ}\text{--}83^{\circ}\text{W}$, $33^{\circ}\text{--}50^{\circ}\text{N}$ (see box in Fig. 1). For MERRA-2, the 14-day running mean anomaly is shown. For SubX models, the 14-day-average anomaly corresponding to the week 1–2 forecast in (a) and (b) or week 3–4 forecast in (c) and (d) is plotted as a color-filled box for each available initialization, with its position on the abscissa corresponding to the center date of the 14-day forecast. For reference, the initialization dates for certain GEOS forecasts are printed with italicized text on each panel. These initializations (21 May–30 Jun) are the focus of the 35-member hindcasts performed with the GEOS model that are presented in section 4.

fields produced by different LSMs (i.e., those from the SubX models or NLDAS-2 product) are highly model-dependent and difficult to compare without prior normalization (Koster et al. 2009b).

Figure 5 provides an example of the dependence of week 3–4 forecast skill on the accuracy of soil moisture initialization for hindcasts initialized around the middle of June 2012. The three models analyzed (CCSM4, FIM, and GEOS) initialize their hindcasts with roughly the correct spatial pattern of VISM anomalies, but all models underestimate the magnitude of negative VISM anomalies over the central United States to varying degrees, with the underestimation in GEOS being smallest (Fig. 5, first column). A comparison of the

VISM anomaly at initialization (Fig. 5, first column) with subsequent P and $T2m$ forecasts approximately 3–4 weeks later (Fig. 5, third and fourth columns), suggests that a larger underestimation in the initial VISM deficit leads to a less skillful forecast. To demonstrate this quantitatively, the regional-mean difference between the hindcast and verification data over the central United States (the core region of soil moisture deficit) is computed for each model and quantity in Fig. 5 and displayed on each panel. A correspondence between the bias in VISM anomaly at initialization and $P/T2m$ anomalies approximately 3–4 weeks later is clear: the smallest bias in VISM anomaly corresponds to the smallest biases in P and $T2m$ anomalies (GEOS;

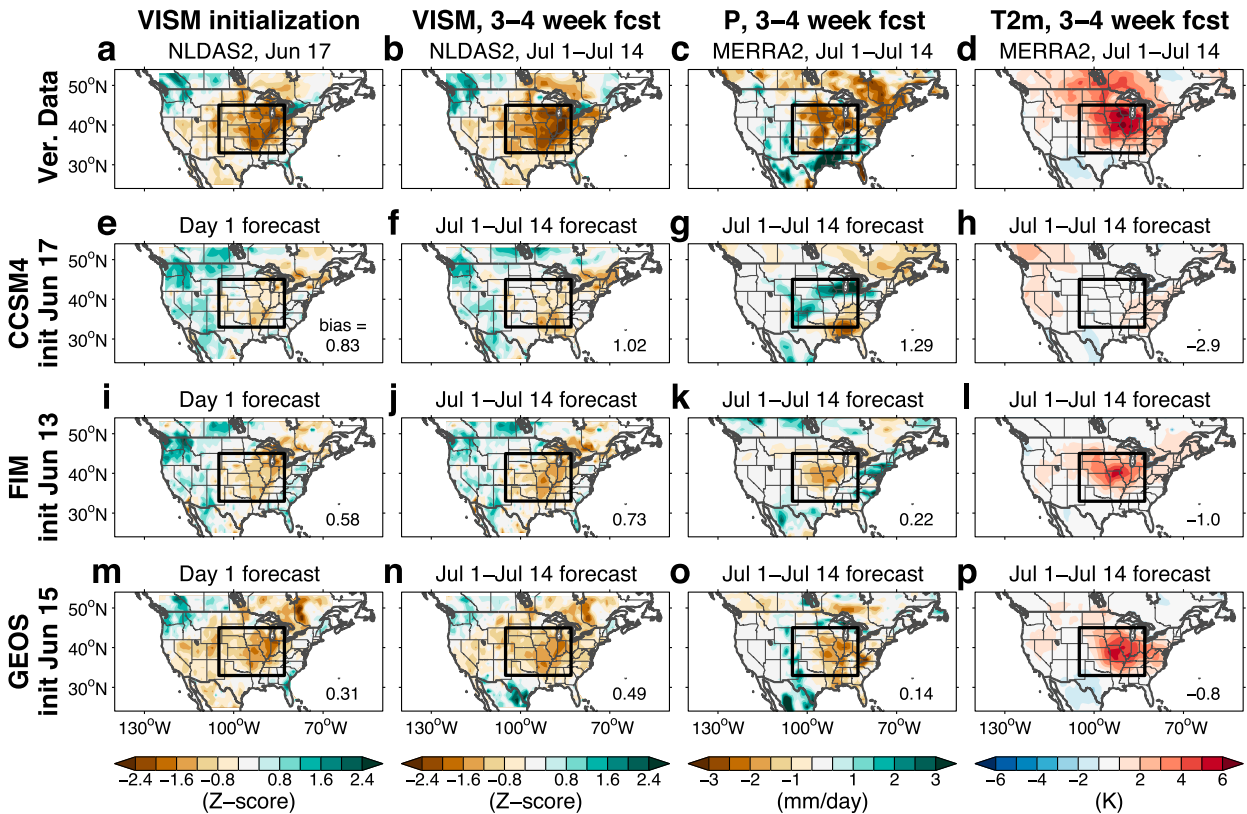


FIG. 5. Dependence of SubX prediction skill on soil moisture initialization for a mid-June initialization in 2012. Shown in the first row are (a) Z score of vertically integrated soil moisture (VISM) on 17 Jun 2012 from the NLDAS-2 mean, (b) NLDAS-2-mean VISM Z score for 1–14 Jul, (c) MERRA-2 *P* anomaly for 1–14 Jul, and (d) MERRA-2 T2m anomaly for 1–14 Jul. In the three subsequent rows, ensemble-mean forecasts for a single initialization on or just before 17 Jun are shown for SubX models that provide soil moisture output: (e)–(h) CCSM4 initialized 17 Jun, (i)–(l) FIM initialized 13 Jun, and (m)–(p) GEOS initialized 15 Jun. The fields shown in each column for the SubX models correspond to those shown for the verification data in the first row: from left to right, day 1 VISM forecast, 1–14 Jul VISM forecast (approximately 3–4 weeks after initialization), 1–14 Jul *P* forecast, and 1–14 Jul T2m forecast. On each panel showing model output, the bias (model minus verification data) is shown for the regional average over 105°–83°W and 33°–45°N (region indicated with a box on all panels) of the corresponding field. In the first column, the bias is computed as the difference between the NLDAS-2 Z score on 17 Jun and the day 1 forecast Z score from the initialization displayed for that model.

Figs. 5m–p), and vice versa (CCSM4; Figs. 5e–h). While factors such as the atmospheric circulation may contribute to model differences in the *P* and T2m forecast skill, Fig. 5 suggests differences in the initial VISM anomaly also play an important role. This is supported by the persistence of soil moisture anomalies throughout the forecasts (cf. the first and second columns of Fig. 5).

To examine the robustness of the results in Fig. 5, we compute regional-mean biases for other initializations during the summer of 2012 and during two other summers in which flash droughts occurred. These two additional summers are 2006 and 2011, when anomalously dry soils were coincident with a lack of precipitation and unusually warm temperatures over the upper Great Plains or southern Great Plains, respectively (Figs. 6c–f). The calculations are summarized in scatterplots in the top panels of Fig. 6, comparing the biases in initialization

day VISM anomalies with corresponding biases in week 3–4 forecast anomalies of *P* and T2m. For this larger sample of cases, there is again a clear relationship between soil moisture initialization accuracy and week 3–4 forecast skill: for summers in which a particular model accurately initializes soil moisture, biases in week 3–4 forecasts of *P* and T2m are, on average, smaller (Figs. 6a,b). The rather large variability in model biases for a particular summer and model (the whiskers about each symbol in Fig. 6), particularly for *P*, likely reflects differences in the predicted atmospheric circulation resulting from different initial conditions and internal atmospheric variability.

Our results clearly demonstrate the importance of accurate land initialization for the subseasonal prediction of central U.S. flash droughts. They are also consistent with earlier findings that accurate soil moisture initialization is beneficial for subseasonal forecasts in

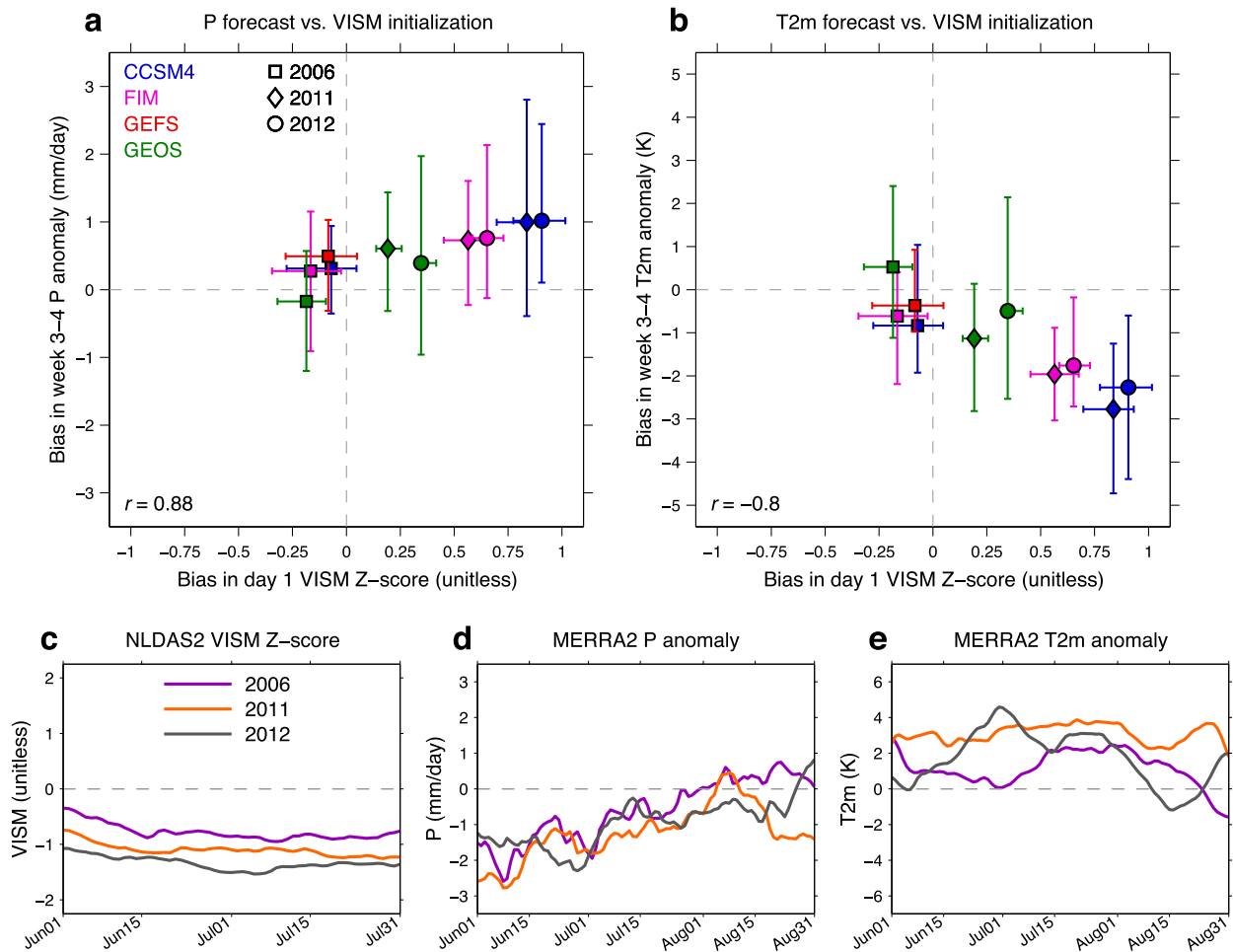


FIG. 6. (top) Relationship between soil moisture initialization accuracy and subseasonal forecast skill in SubX models for three Great Plains drought events. Shown on the abscissa is the regional-mean bias in the day 1 forecast Z score of VISM. On the ordinate is the regional-mean bias in the week 3–4 forecast anomaly of (a) P or (b) $T2m$. Biases are relative to the NLDAS-2 mean for VISM or MERRA-2 for P and $T2m$. Each symbol represents the mean of forecast biases for several summer initializations within a particular year and for a particular model (see legend), while whiskers depict the corresponding range of forecast biases for that summer and model. Considered initializations are from 1 Jun to 15 Jul for both 2006 and 2012, and from 1 Jun to 31 Jul for 2011. Biases are averaged over 100° – 85° W and 33° – 49° N for 2006, 104° – 93° W and 30° – 40° N for 2011, and 105° – 83° W and 33° – 45° N for 2012. The region used for 2012 is the same as that shown in Fig. 5. Also shown on each panel is the correlation (r) between mean biases of VISM and $T2m/P$ forecasts. (bottom) For reference, the regional-mean anomalies of (c) VISM, (d) P , and (e) $T2m$ are shown for the verification data for the three summers analyzed.

general, particularly when initial soil moisture conditions are more extreme (e.g., Koster et al. 2010). The implication is that the precipitation and temperature anomalies in June–July 2012 (Fig. 2) were partly induced by the dry land state that developed in the late spring of 2012. To predict the component of these anomalies originating from the prior soil state and land–atmosphere coupling, it is critical that forecasts are initialized with accurate soil moisture anomalies.

What factors contribute to more accurate soil moisture initialization and subsequent prediction skill in some models? To address this question, we investigate

the soil moisture initialization approaches of all SubX models. For initialization, all modeling centers utilize soil moisture data that are ultimately derived from an LSM driven with observed meteorology (Saha et al. 2010, 2014; Bleck et al. 2015; Sun et al. 2018; Y. Zhu 2019, personal communication; H. Lin 2019, personal communication; Y. Richter 2019, personal communication; B. Kirtman 2019, personal communication; N. Barton 2019, personal communication). Ideally, the LSM used to derive the initialization data should be the same as the LSM used in the fully coupled model that generates the forecasts (Koster et al. 2009b). Two models that do not meet this criterion are

CCSM4 and ESPC, both utilizing inconsistent LSMs in the initialization data and forecast system (B. Kirtman and N. Barton 2019, personal communication). CCSM4 attempts to correct for the inconsistency by scaling initial soil moisture anomalies to have the same Z score in the model climate and observed climate, but this scaling is subject to the possibly inaccurate assumption of a Gaussian soil moisture distribution (Koster et al. 2009b). Both CCSM4 and ESPC produce noticeably less skillful forecasts during the 2012 drought (Figs. 4–6) and over the entire 1999–2015 period (Figs. 3e,f).

In addition to soil moisture initialization, other aspects of land–atmosphere dynamics may influence the prediction skill of flash droughts and contribute to some of the scatter in Fig. 6. These include soil moisture memory (e.g., Lorenz et al. 2010), the strength of the land–atmosphere coupling in models, and the degree of climate drift over the course of a forecast. These factors are not assessed in this study, partly due to a lack of necessary SubX output for the investigation. These factors should be investigated in future work to better understand the sources of flash drought prediction skill and identify potential shortcomings in current-generation forecast systems, possibly through controlled GCM experiments.

2) PREDICTION OF ROSSBY WAVE TRAIN

In this subsection, we briefly evaluate the prediction skill of the quasi-stationary Rossby wave train during the 2012 flash drought in the SubX ensemble. Due to different initialization dates and both limited and variable ensemble sizes among SubX models, a robust and fair intermodel comparison of wave train prediction skill is challenging. Nevertheless, an assessment of the model output shows that some models exhibit skill in predicting features of the wave train. For example, GEFS,³ FIM, and GEOS are able to sufficiently predict the development of the wave train and its propagation to the central North Pacific when hindcasts are initialized between 10 and 13 June (Figs. 7a–d). The downstream propagation of the wave train to North America, however, is slower than observed in both GEFS and FIM, with a ridge developing over western North America rather than central North America in late June in those models (Figs. 7f,g). The eastward propagation of the wave and North America ridge development are more accurate in GEOS (Fig. 7h). In section 4b, we examine the prediction skill of the 2012 Rossby wave train in GEOS more extensively with additional ensemble members.

³Based on our assessment, the V200 and H200 fields do not appear to be adversely impacted by the land initialization error mentioned in section 2a. To compute the climatologies for V200 and H200, the years 2011 and 2012 are included in the calculation.

4. Factors contributing to skillful predictions in GEOS

The additional ensemble members and experiments generated with the GEOS model (described in section 2b) serve several purposes. The first is to assess the dependence of the GEOS prediction skill on ensemble size. Specifically, we assess whether the GEOS SubX forecasts during the 2012 drought (e.g., Figs. 4–7), which are based on four ensemble members, are similar in a larger ensemble. The larger ensemble size better samples the noise originating from internal and inherently unpredictable atmospheric variability. That is, an ensemble-mean forecast from a large ensemble is less likely to reflect coincidental (and chaotic) large anomalies from a few realizations and more likely to reflect robust sources of prediction skill. The fundamental question addressed here is, does the skill seen in the four-member GEOS ensemble (Figs. 4–7) occur by chance, or does it originate from robust sources? Assuming GEOS's forecasts are skillful in a larger ensemble, we then further examine the effect of soil moisture initialization with a controlled experiment in which the influence of the observed 2012 land initial conditions is removed. By comparing the hindcasts that are initialized with and without land information, one can also infer the sources of prediction skill originating solely from atmospheric initial conditions.

Figure 8 presents an overview of the results from the 35-member GEOS hindcasts initialized between 21 May and 30 June 2012. All CTL hindcasts (which are initialized with both atmospheric and land conditions) predict positive temperature anomalies of at least 1–2 K over the central United States throughout their 45-day forecast period. Additionally, most hindcasts predict negative precipitation anomalies with magnitude 0.5–1 mm day⁻¹ or greater over the same period and region. These predictions are in broad agreement with the anomalies in MERRA-2, but exhibit smaller temporal fluctuations, are slightly weaker in magnitude, and have peak anomalies at somewhat different times than occurred in nature (similar to the four-member GEOS SubX results in Fig. 4). Thus, the GEOS forecasts of the atmospheric state during the summer of 2012 are robust, with a 35-member ensemble showing overall skillful predictions of the warm and dry conditions that transpired.

a. Land initialization

What effect does land initialization have on the skill of the GEOS hindcasts? To answer this, we compare the GEOS CTL hindcasts with parallel runs in which the land initial conditions are taken from alternative years

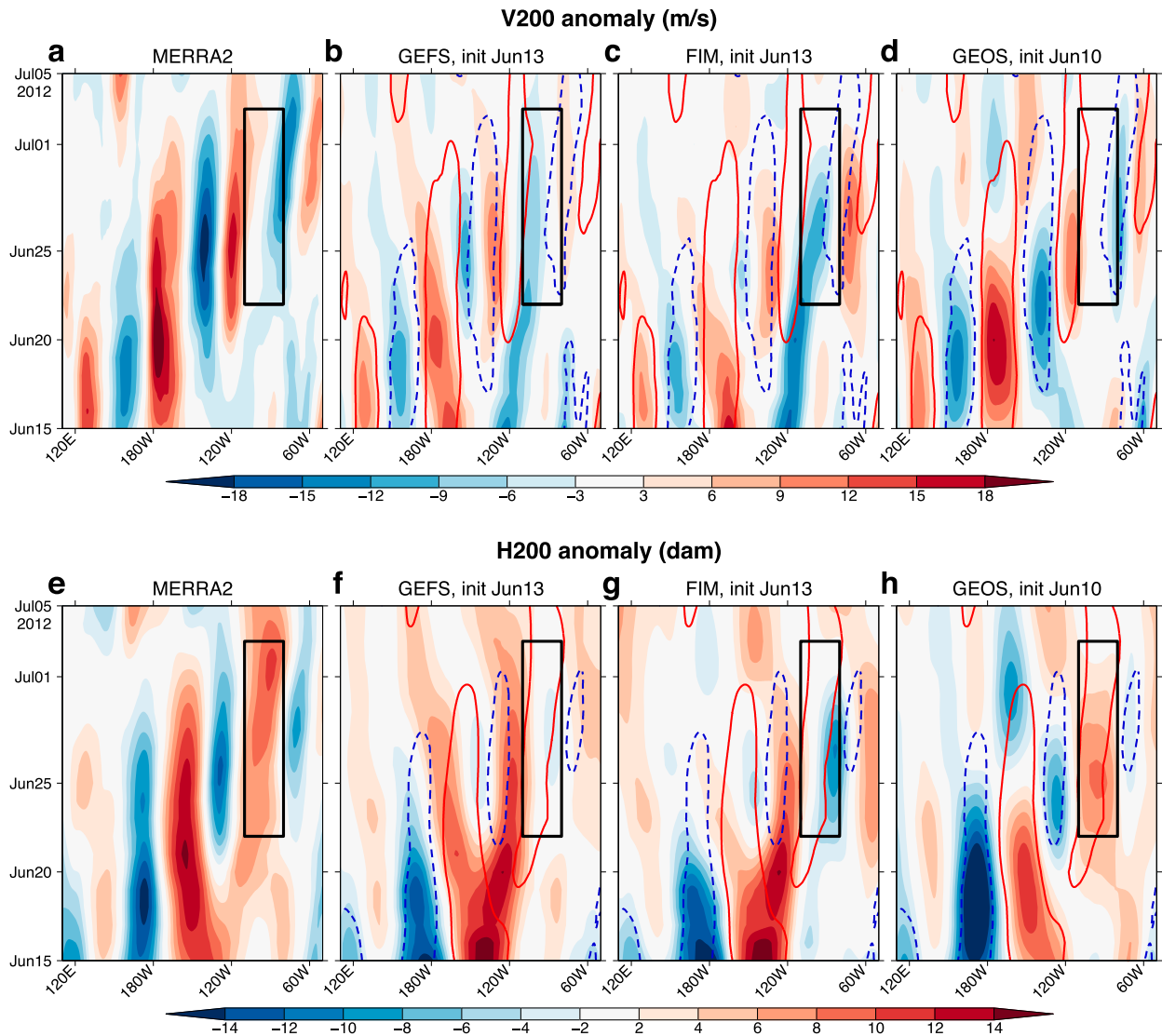


FIG. 7. The quasi-stationary Rossby wave train of June 2012 in observations and example SubX hindcasts. Hovmöller diagrams show anomalies in (top) V200 and (bottom) H200 averaged over 33° – 50° N for a domain covering the western North Pacific to the western North Atlantic for (a),(e) MERRA-2, (b),(f) GEFS initialized 13 Jun, (c),(g) FIM initialized 13 Jun, and (d),(h) GEOS initialized 10 Jun. The MERRA-2 positive and negative 6 m s^{-1} V200 anomaly contours are plotted in (b)–(d) as solid red and dashed blue lines, respectively. Similarly, MERRA-2 positive and negative 6-dam H200 anomaly contours are plotted in (f)–(h). A 10-day running mean has been applied in all panels for smoother plotting. The black rectangle indicates the central United States (110° – 80° W) from 22 Jun through 3 Jul (the approximate timing of the development and intensification of a ridge over North America).

(i.e., the noSFC hindcasts described in section 2b, also shown in Fig. 8). In the noSFC hindcasts, predicted anomalies of P and T_{2m} generally reduce to near zero approximately two weeks after the hindcasts are initialized. Anomalies of appreciable magnitude that occur in the first two weeks are the consequence of atmospheric initial conditions, which are unperturbed in the noSFC runs. Beyond two weeks, the atmospheric states contributing to P and T_{2m} anomalies become so divergent across the ensemble (a consequence of error

growth imposed by the shuffled land initial conditions) that their average is no different from climatology. The results thus show that with some exceptions (highlighted later), atmospheric initial conditions alone were insufficient to predict anomalies of P and T_{2m} during the summer of 2012 on a subseasonal time scale.

To further stress the above findings, Figs. 9 and 10 display the predicted spatial patterns of P and T_{2m} , respectively, for the peak drought period of 20 June–9 July 2012. When initialized on 26 May, for example, the

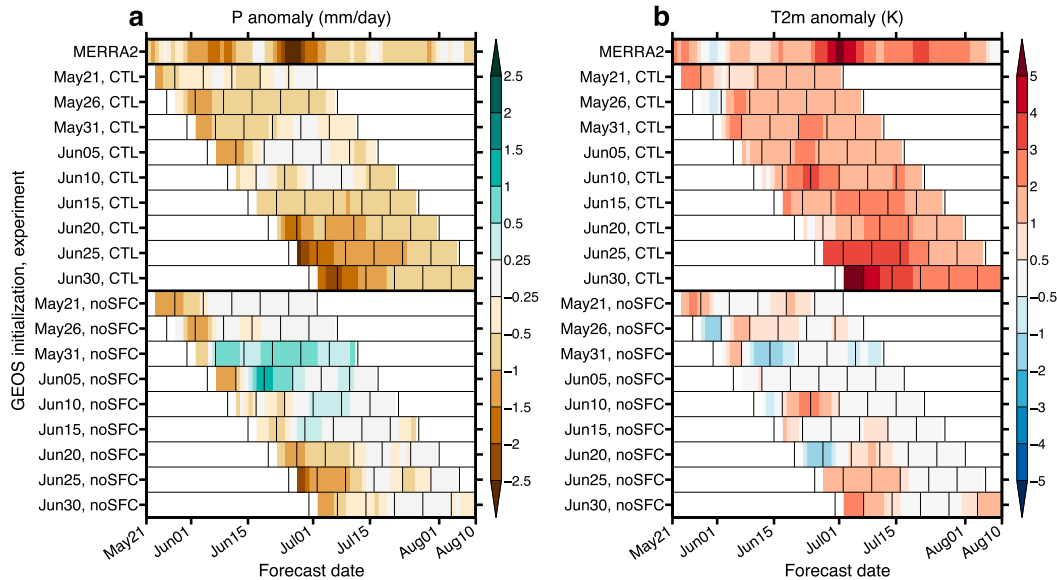


FIG. 8. Prediction skill of GEOS 35-member hindcasts for the 2012 flash drought. The 7-day running mean anomaly (averaged over the central United States, 105° – 83° W, 33° – 50° N) of (a) P (mm day^{-1}) and (b) $T2m$ (K) is shown for MERRA-2 and the ensemble mean of each GEOS hindcast initialized between 21 May and 30 Jun 2012. Results are shown for the control (CTL) and scrambled surface (noSFC) experiments (see section 2b for experiment details). Vertical lines are drawn at the initialization date and each week after initialization for each of the GEOS hindcasts.

CTL experiment predicts negative P and positive $T2m$ anomalies in the U.S. Great Plains and Midwest that broadly agree with MERRA-2 (Figs. 9a,b and 10a,b). However, the magnitude of the predicted anomalies, especially for P , is weaker than observed. Without land initial conditions (noSFC experiment), these P and $T2m$ anomalies are essentially absent (Figs. 9c and 10c). Furthermore, the epicenter of the predicted $T2m$ anomalies for 20 June–9 July, when initialized 26 May (CTL experiment), closely corresponds to the epicenter of anomalously dry soil moisture conditions in the summer of 2012 (cf. Figs. 1h and 10b). These findings provide strong evidence that the extreme anomalies associated with the peak of the 2012 flash drought were contributed, in part, by preceding large soil moisture deficits over the central United States.

b. Rossby wave train

Although atmospheric initial conditions alone do not generally contribute to prediction skill beyond two weeks, one noteworthy exception is with the 10 June initialization from the GEOS model. Figures 9 and 10 also show the spatial pattern of the 20 June–9 July forecasts of P and $T2m$ when GEOS is initialized 10 June. Compared to the 26 May initialization, the predicted $T2m$ anomaly from the CTL hindcast initialized 10 June is larger in both magnitude and spatial

extent, agreeing better with MERRA-2 (Figs. 10a,d). Interestingly, when surface initial conditions are shuffled for the 10 June initialization (noSFC hindcast), a potent $T2m$ anomaly remains across the northern United States and southern Canada (Fig. 10e). Similarly, the predicted P anomaly for 20 June–9 July is only marginally reduced in magnitude in the noSFC hindcast compared to the CTL hindcast (Figs. 9d,e). It appears, therefore, that both land and atmospheric initial conditions on 10 June contribute separately to the predicted anomalies of P and $T2m$ that occur roughly 2–4 weeks later.

A closer inspection of the V200 and H200 fields from the 35-member CTL hindcast initialized 10 June reveals that the quasi-stationary wave train in mid-late June is adequately predicted (Figs. 11–12), in agreement with the four-member results shown in Fig. 7. In particular, the development of waves between Eurasia and the western North Pacific in mid-June, eastward propagation of the waves, and associated development of a ridge over central North America in late June are all reasonably captured (Figs. 11a,b and 12a,b). The results are very similar in the noSFC hindcast (not shown), confirming that atmospheric initial conditions alone are sufficient to generate the wave train. Remarkably, the intraensemble spread during the first two weeks of the forecast, the key period of wave train development, is

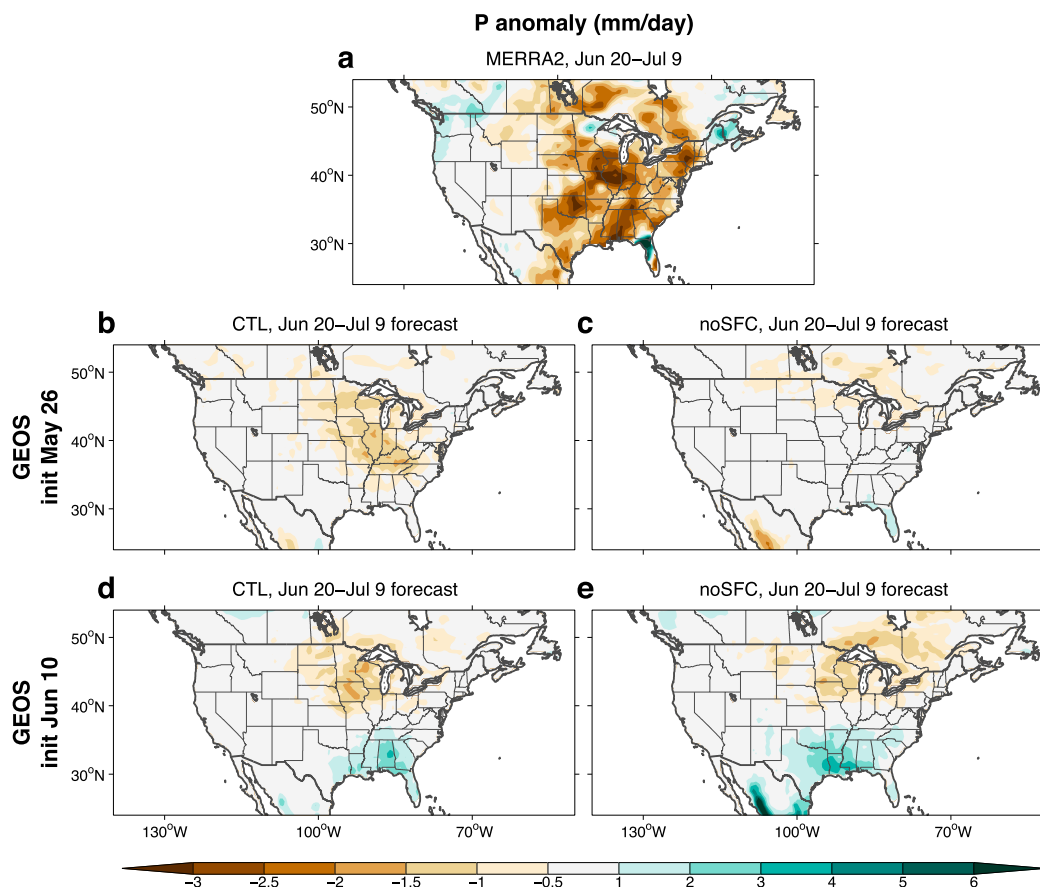


FIG. 9. Prediction skill of precipitation during the 2012 flash drought in selected GEOS 35-member hindcasts. (a) MERRA-2 P anomaly (mm day^{-1}) for 20 Jun–9 Jul 2012. GEOS ensemble-mean forecasts of the P anomaly for 20 Jun–9 Jul when initialized 26 May for the (b) CTL and (c) noSFC experiments. GEOS ensemble-mean forecasts of the P anomaly for 20 Jun–9 Jul when initialized 10 Jun for the (d) CTL and (e) noSFC experiments.

quite small (Figs. 11c and 12c), demonstrating the robustness of the wave train development in GEOS. We further find that the wave train development is adequately predicted when GEOS is initialized on or after 9 June (not shown). Additional research is currently underway to better understand the physical basis of the robust wave train prediction in GEOS. In short, the upper-level ridge over North America induced by the predicted wave train substantially enhances the warming that would occur from dry initial soil conditions alone (Figs. 10b,d). The GEOS's prediction of the wave train thus leads to a more skillful forecast of the peak of the 2012 drought, particularly with regard to the magnitude of T2m anomalies. The consequence of this is a rather skillful multiweek forecast for conditions that would favor drought intensification and subsequent land–atmosphere feedbacks, including reduced atmospheric moisture, increased SH, and reduced LH (Fig. 13).

Two important conclusions can be drawn here: 1) quasi-stationary Rossby waves are a potentially valuable

source of flash drought prediction skill; without the adequate prediction of such waves, the onset and/or true intensity of a flash drought may not be properly predicted, and 2) the development and propagation of Rossby wave trains, and their associated drought-inducing quasi-stationary circulation anomalies, can be sufficiently predicted in a subseasonal forecast model as far as 2–3 weeks in advance. While previous studies have also emphasized the importance of Rossby waves for flash drought prediction skill (e.g., Teng et al. 2013; Wang et al. 2017), our study is the first to demonstrate this importance with a state-of-the-art subseasonal forecast model.

c. Wave train–enhanced soil moisture deficits

Once quasi-stationary atmospheric circulation anomalies are established over the central United States, they can induce local extreme precipitation deficits and heat waves that dry the soil. The dry soil helps to further sustain the drought through land–atmosphere

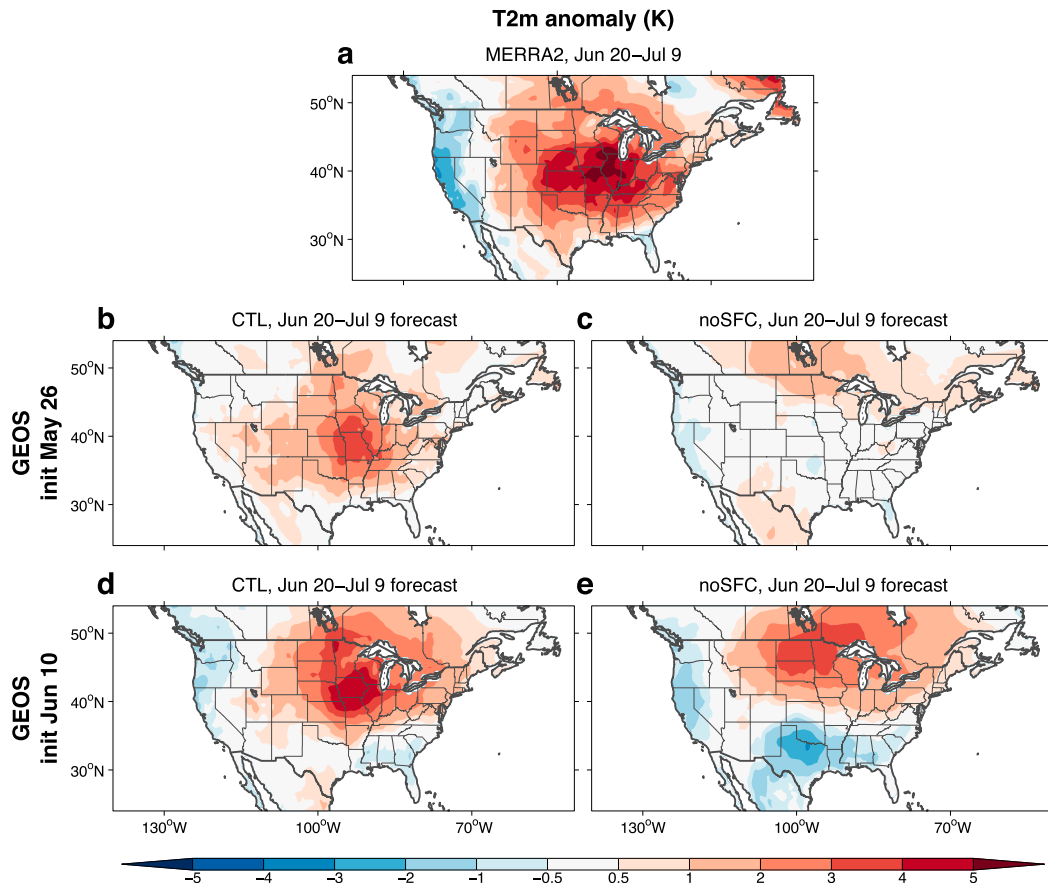


FIG. 10. As in Fig. 9, but for the T2m anomaly (K) for 20 Jun–9 Jul 2012.

feedbacks, and these feedbacks can extend subseasonal prediction skill.

An example of the above is provided with the 25 June initialization. The day of 25 June coincides with a strong positive H200 anomaly over North America associated with the arrival of the quasi-stationary wave train (Figs. 2h and 14a,b). When GEOS is initialized with realistic land and atmospheric states on this date, the strong initial ridge induces warm and dry atmospheric conditions that act to maintain the initial soil moisture deficit throughout the 45-day forecast (Figs. 14c–e, red curves). The parallel noSFC hindcast indicates that the extreme initial atmospheric conditions, alone, contribute to substantial drying of the soil within the first two weeks, mostly due to a dearth of precipitation during that time (Figs. 14c,d, blue curves). The noSFC hindcast subsequently predicts negative P and positive T2m anomalies beyond week 2 that are roughly half the magnitude of those in the CTL hindcast (Figs. 14d,e). Since no upper-level circulation anomalies are present after week 2 of the noSFC hindcast (not shown), the P and T2m anomalies then are likely the response to the

dry soil. The results imply that soil moisture anomalies induced by quasi-stationary circulations can be substantial and can later induce P and T2m anomalies beyond the weather prediction time frame, potentially enhancing/extending subseasonal forecast skill. This highlights the importance of accurate atmospheric initial conditions, a global forecast system that is skillful for 1–2-week forecasts, and a realistic representation of land–atmosphere dynamics for subseasonal prediction.

5. Summary and discussion

In this paper, we examine the subseasonal prediction skill of the 2012 Great Plains flash drought in the SubX ensemble. Overall, the prediction skill is highly variable among models. With further analysis of the SubX ensemble and additional controlled hindcasts with the GEOS model, we are able to deduce robust sources of prediction skill emanating separately from land and atmospheric initial conditions. Although the models do not adequately predict the onset or rapid worsening of dry soils in 2012, the more skillful hindcasts exhibit

V200 anomaly in observations and the GEOS hindcast initialized June 10, 2012

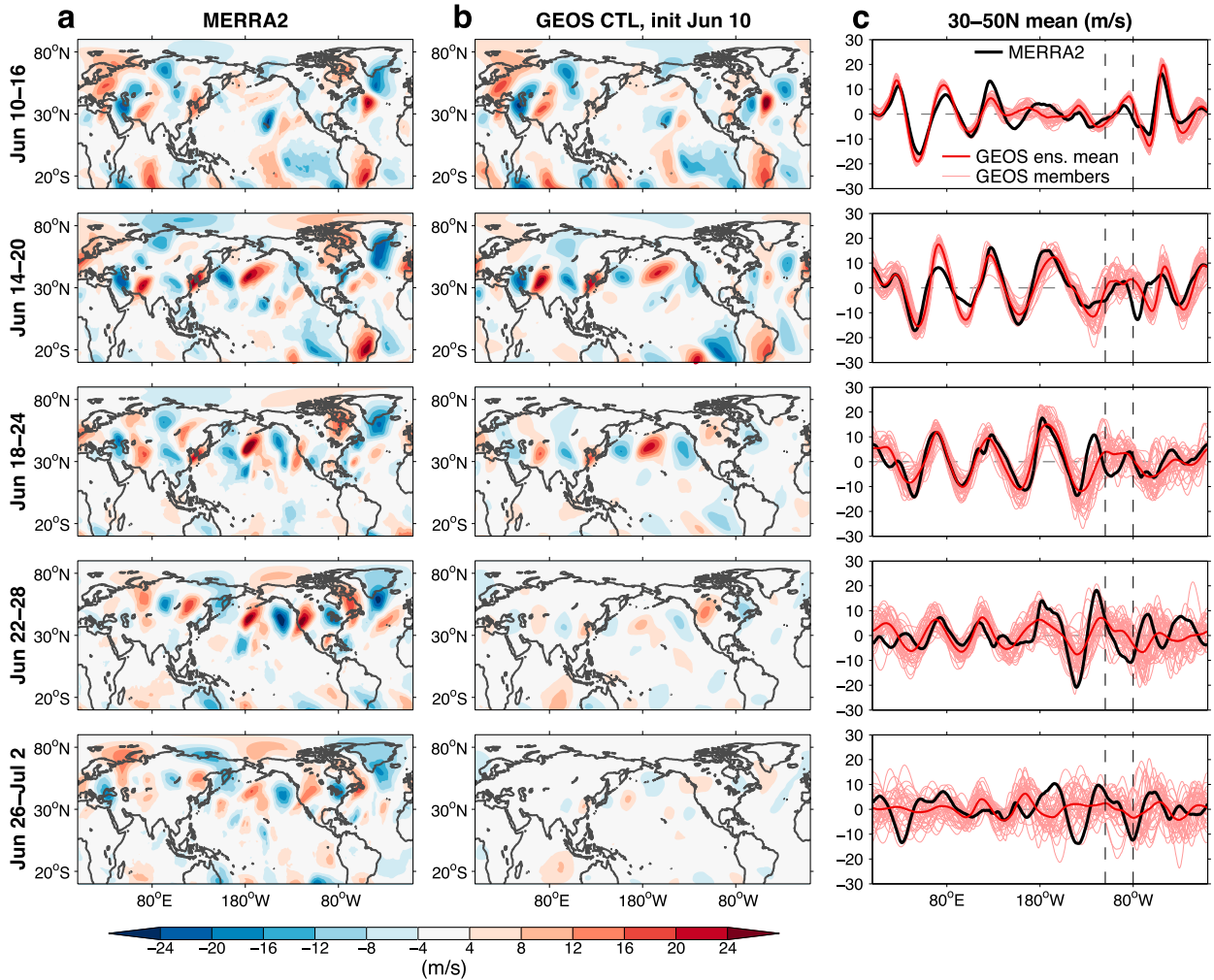


FIG. 11. Prediction of the June 2012 quasi-stationary wave train in the GEOS hindcast initialized 10 Jun. (a) Temporal evolution of the V200 anomaly (m s^{-1}) in MERRA-2 from 10 Jun through 2 Jul 2012. Each row corresponds to the 7-day average printed along the left side of the figure. (b) Corresponding temporal evolution of the V200 anomaly in the GEOS 35-member ensemble mean initialized 10 Jun for the CTL experiment. (c) Meridional mean of the observed and predicted V200 anomalies shown in (a) and (b) over 30°–50°N. In (c), the ensemble-mean of the GEOS hindcast is indicated with a thick red line and individual members with thin pink lines. Vertical dashed lines in (c) indicate the central United States (110°–80°W).

useful predictive guidance with regard to continuing or worsening drought conditions throughout the summer.

The first source of prediction skill stems from the accurate representation of the soil physical state at the start of a forecast. Our results strongly demonstrate, based on two separate findings, that accurate soil moisture initialization improves subseasonal forecasts of the 2012 flash drought and other similar events: 1) biases in predicted P and T2m anomalies during central U.S. droughts are smaller when initial soil moisture anomalies are more accurate in the SubX ensemble (Figs. 5 and 6), and 2) the adequate prediction of precipitation and temperature anomalies associated with land–atmosphere

coupling during the 2012 drought relies on the accurate initialization of both the land and atmosphere in the GEOS model (Figs. 8–10).

This has promising implications. The dry soil moisture conditions in the late spring of 2012 predisposed the atmosphere to be warm and dry in subsequent weeks to months due to land–atmosphere feedbacks. In the absence of strong circulation anomalies that could counteract this, accurately initialized forecast models could have predicted that warm and dry conditions would prevail in the summer of 2012. Note, however, that negative P and positive T2m anomalies in 2012 are only forecasted at subseasonal time scales once initial soil

H200 anomaly in observations and the GEOS hindcast initialized June 10, 2012

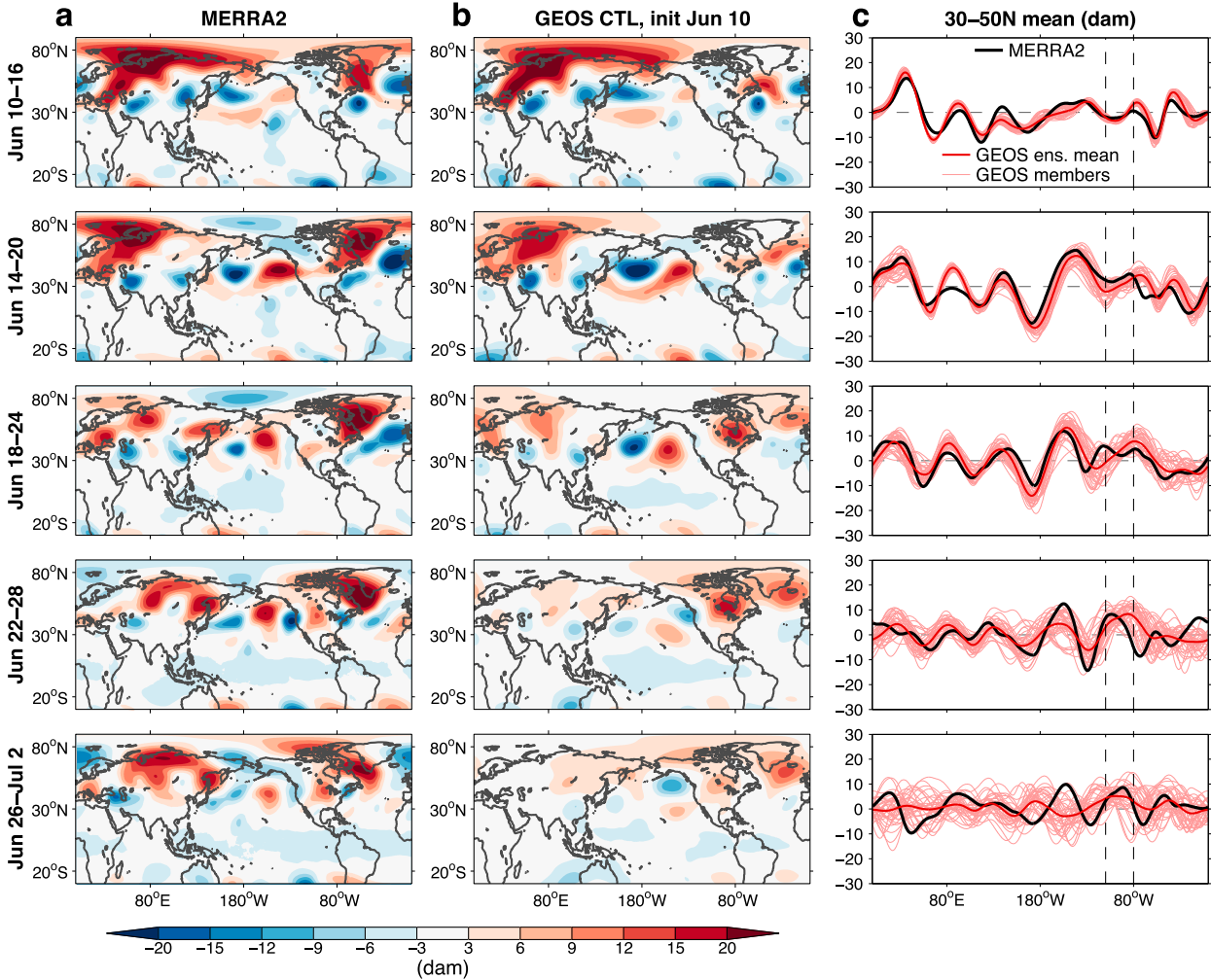


FIG. 12. As in Fig. 11, but for the temporal evolution of the H200 anomaly (dam) from 10 Jun through 2 Jul 2012.

conditions become sufficiently dry (Figs. 2, 4c,d, and 8). Without an antecedent soil moisture anomaly, subseasonal predictions depend largely on atmospheric initial conditions, which in many cases do not provide prediction skill beyond the first two forecast weeks. This explains why the precipitation deficits and warm temperatures in early May 2012, which led to dry soil conditions in the first place (Figs. 2a–c), are not well predicted on 3–4-week lead times in the SubX ensemble (Figs. 4c,d).

In addition to soil moisture, cross-Pacific stationary Rossby wave trains, and their associated drought-inducing circulation anomalies, are shown here to be a key source of subseasonal prediction skill for extreme events such as the 2012 flash drought. In particular, our simulations with the GEOS model show that the adequate prediction of the 2012 wave train and subsequent

North America ridge leads to a more realistic forecast. However, a limitation of this study is that we do not fully understand why the GEOS model predicts the wave train better than other models or why certain initialization dates (e.g., 10 June) produce better forecasts. Addressing these unresolved questions requires a better understanding of quasi-stationary wave trains, their physical sources, and the characteristics of forecast models that enable their adequate prediction. Studies have shown that subseasonal convective variability in the East Asian monsoon region can trigger drought-inducing wave trains such as that in 2012 (Jiang and Lau 2008; Zhu and Li 2016; Lopez et al. 2019). Furthermore, it has been demonstrated that a realistic representation of the climatological summer jet stream in the North Pacific is necessary for the successful simulation of wave train propagation (Wang et al. 2017). Additional

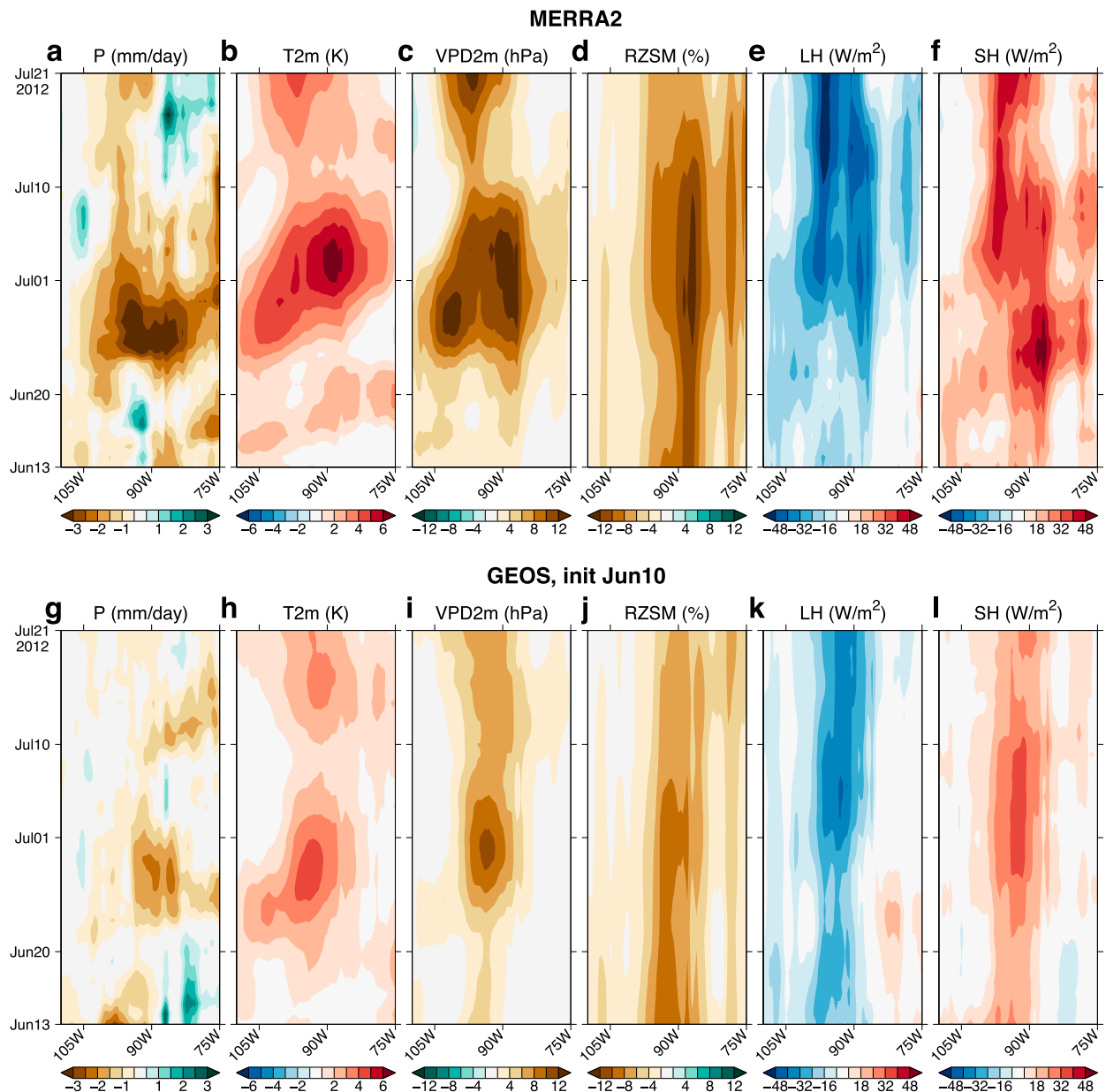


FIG. 13. Rapid intensification of the 2012 Great Plains drought in MERRA2 and the GEOS SubX hindcast initialized 10 Jun. Shown in the top panels are Hovmöller diagrams of anomalies in (a) P (mm day^{-1}), (b) T_{2m} (K), (c) VPD_{2m} (hPa) (d) $RZSM$ (%), (e) LH (W m^{-2}), and (f) SH (W m^{-2}) from MERRA-2 for 13 Jun through 21 Jul. Corresponding anomalies from the GEOS hindcast initialized 10 Jun are shown in the bottom panels. (g)–(l) The average of the first four GEOS ensemble members (i.e., the original SubX members) is shown in order to match the ensemble size available from operational SubX forecasts. All anomalies are averaged over 33° – 50° N (see box in Fig. 1). A 7-day running mean has been applied for smoother plotting.

research on these topics should provide a better overall understanding of forecast system performance and help identify areas of the forecast systems to improve in order to improve subseasonal prediction of extreme events.

The key findings of this study, which are largely based on the 2012 Great Plains flash drought, presumably apply to central U.S. summer flash droughts in general.

One reason is that strong land–atmosphere interaction is intrinsic to the central United States during summer (e.g., Koster et al. 2004a, 2009a); our results based on flash drought events in three different years support this (Fig. 6). Another reason is that stationary Rossby waves, which are known to induce short-term warm-season droughts in the United States, are often a manifestation

GEOS prediction skill when initialized June 25, 2012

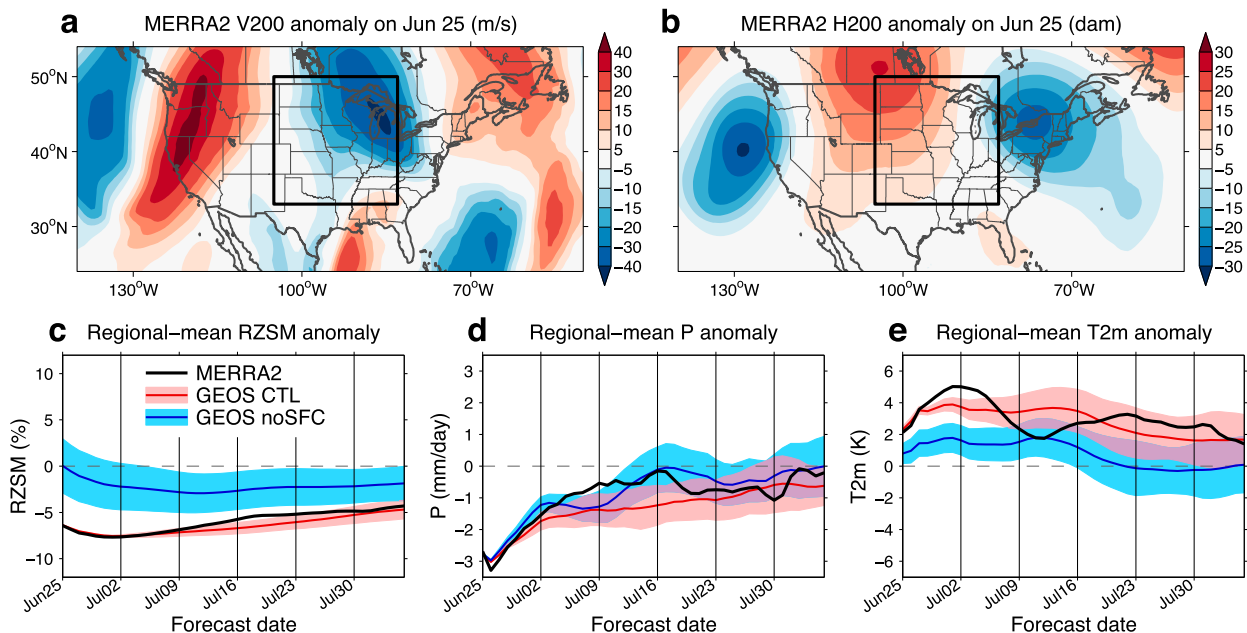


FIG. 14. Closer examination of the prediction skill of GEOS hindcasts initialized 25 Jun 2012. (top) Observed anomaly of (a) V200 (m s^{-1}) and (b) H200 (dam) on 25 Jun 2012, from MERRA-2. (bottom) Regional-mean (105° – 83° W, 33° – 50° N, indicated with a box in the top panels) anomalies of (c) RZSM (%), (d) P (mm day^{-1}), and (e) T2m (K) in MERRA-2 (black), the GEOS CTL hindcast initialized 25 Jun (red), and the GEOS noSFC hindcast initialized 25 Jun (blue). The 7-day running mean anomaly is shown. In (c)–(e), the 35-member GEOS ensemble mean is shown as a line and ensemble spread (± 1 standard deviation) is shown as shading. Vertical lines and corresponding abscissa labels in (c)–(e) mark the initialization date and every 7 days thereafter.

of recurring modes of the subseasonal atmospheric circulation (e.g., Schubert et al. 2011; Wang et al. 2017).

In summary, our results highlight key sources of subseasonal prediction skill of flash droughts in the central United States and identify areas where further model development and research are needed. One area where model development may improve subseasonal prediction is land initialization. As shown here, subseasonal forecasts are more skillful in models that use the same LSM for their soil moisture initialization and to generate forecasts. Not all models follow this approach; if they did, subseasonal prediction of central U.S. flash droughts would likely improve in a multimodel ensemble. In addition to land initialization, the simulation of land–atmosphere coupling in models can have an important influence on prediction skill. A limitation of our study is the use of a single reanalysis dataset (MERRA-2) for forecast verification, which makes a robust assessment of land–atmosphere dynamics challenging. Future research on the surface energy budget, planetary boundary layer dynamics, and evaporative demand in models, using a variety of verification datasets, could prove useful for better understanding model biases as they relate to flash drought prediction.

Finally, to accurately predict drought intensification on subseasonal time scales, precipitation must be

accurately predicted (Livneh and Hoerling 2016; Koster et al. 2019). Our results show that the subseasonal prediction skill of precipitation is limited, even in cases for which soil moisture is accurately initialized and the 2012 wave train is sufficiently predicted (e.g., Figs. 4, 9–10, 13). This undoubtedly hinders the ability of any SubX model to adequately predict soil moisture declines associated with flash drought onset or intensification (e.g., Fig. 13). Improvement in precipitation prediction is therefore crucial for meaningful flash drought forecasting. Additional research on the impacts of model biases (e.g., Klein et al. 2006; Zhang et al. 2018; Chang et al. 2019), atmospheric resolution (e.g., Zheng et al. 2019), and convective parameterizations (e.g., Wang and Hsu 2019; Xie et al. 2019), for example, on the prediction skill of precipitation is accordingly warranted.

Acknowledgments. This work was supported by the NOAA MAPP (NA14OAR4310221) and NASA MAP (NNG17HP01C and WBS 802678.02.17.01.33) programs. We acknowledge the agencies that support the SubX system, and we thank the climate modeling groups (listed in Table 1) for producing and making available their model output. NOAA MAPP, ONR, NASA MAP, and NOAA NWS jointly provided coordinating support

and led development of the SubX system. SubX output may be obtained from the Data Library of the International Research Institute for Climate and Society, Columbia University (<http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/>). MERRA-2 data were developed by the Global Modeling and Assimilation Office at NASA GSFC under funding by the NASA MAP program. We thank the partnering NLDAS groups for producing the NLDAS-2 model output. MERRA-2 data and NLDAS-2 output are disseminated through the Goddard Earth Science Data and Information Services Center (GES DISC). CPC global temperature data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, and are available from their website (<https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html>).

APPENDIX

Computation of SubX Hindcast Anomalies and MMM

For the SubX hindcasts, daily anomalies are computed for a given initialization date by first averaging all hindcasts initialized on that calendar day over the 17 years of available output, generating a lead-dependent climatology, then removing that climatology from each of the individual hindcasts. This calculation is straightforward for the models that initialize hindcasts on unique calendar dates each year (GEM, GEOS, CFSv2, and CCSM4). However, several models (46LCEM1, GEFS, FIM, and ESPC) initialize hindcasts on a “same day of week” schedule, such that different calendar dates are initialized each year (Table 1). For these models, we define a unique set of 52 initialization dates from which to compute lead-dependent climatologies. The unique set of dates begins with 4 January, ends with 27 December, and has an increment of 7 days and excludes the leap day (29 February). For each unique date, the hindcasts initialized closest to that date every year are averaged over the 1999–2015 period to compute an approximate lead-dependent climatology corresponding to that unique initialization date. The computed climatologies are then subtracted from the hindcasts making up each climatology to generate anomalies.

Prior to computing climatologies and anomalies for the CFSv2 and ESPC models, additional steps are employed due to the atypical initialization frequencies of these models (see Table 1). CFSv2 was initialized four times every day at 0000, 0600, 1200, and 1800 UTC. To reduce the number of CFSv2 initializations to be more consistent with the other SubX models, we first average CFSv2 hindcasts over consecutive 7-day periods

beginning with the 1 January initialization, aligning forecast verification dates for the hindcasts that are averaged. A single average is computed for each 7-day period and for each unique hour of the day that CFSv2 hindcasts were initialized. Each unique initialization hour is then regarded as a separate ensemble member for CFSv2. In a similar fashion, the ESPC hindcast output is averaged over the four consecutive dates that were initialized, resulting in a single mean hindcast each week.

Due to differences in the dates and frequencies of the model initializations, the computation of the MMM is not trivial. Here, we compute the MMM for a unique set of predefined initialization dates. These unique dates are the same as described above for the computation of climatologies (i.e., 4 January, 11 January, . . . , 27 December). For each unique date, the ensemble-mean hindcasts from each model that are initialized closest to that date are averaged, while aligning forecast verification dates. In most cases, only one hindcast from each model contributes to the MMM for a given unique initialization date. In cases where two hindcasts contribute, the hindcasts from both nearby initializations are first averaged for the given model.

REFERENCES

- Ambrizzi, T., B. J. Hoskins, and H.-H. Hsu, 1995: Rossby wave propagation and teleconnection patterns in the austral winter. *J. Atmos. Sci.*, **52**, 3661–3672, [https://doi.org/10.1175/1520-0469\(1995\)052<3661:RWPATP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<3661:RWPATP>2.0.CO;2).
- Basara, J. B., J. I. Christian, R. A. Wakefield, J. A. Otkin, E. H. Hunt, and D. P. Brown, 2019: The evolution, propagation, and spread of flash drought in the central United States during 2012. *Environ. Res. Lett.*, **14**, 084025, <https://doi.org/10.1088/1748-9326/ab2cc0>.
- 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>.
- Bleck, R., and Coauthors, 2015: A vertically flow-following icosahedral grid model for medium-range and seasonal prediction. Part I: Model description. *Mon. Wea. Rev.*, **143**, 2386–2403, <https://doi.org/10.1175/MWR-D-14-00300.1>.
- Bosilovich, M. G., and Coauthors, 2015: MERRA-2: Initial evaluation of the climate. NASA Tech. Memo. NASA/TM-2015-104606, Vol. 43, 145 pp., <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>.
- , F. R. Robertson, L. Takacs, A. Molod, and D. Mocko, 2017: Atmospheric water balance and variability in the MERRA-2 reanalysis. *J. Climate*, **30**, 1177–1196, <https://doi.org/10.1175/JCLI-D-16-0338.1>.
- Burnash, R. J. C., 1995: The NWS river forecast system—Catchment modeling. *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., Water Resources Publications, 311–366.
- Chang, Y., S. D. Schubert, R. D. Koster, A. M. Molod, and H. Wang, 2019: Tendency bias correction in coupled and uncoupled global climate models with a focus on impacts over North America. *J. Climate*, **32**, 639–661, <https://doi.org/10.1175/JCLI-D-18-0598.1>.
- Chen, L. G., J. Gottschalk, A. Hartman, D. Miskus, R. Tinker, and A. Artusa, 2019: Flash drought characteristics based on U.S.

- Drought Monitor. *Atmosphere*, **10**, 498, <https://doi.org/10.3390/atmos10090498>.
- Chen, P., and M. Newman, 1998: Rossby wave propagation and the rapid development of upper-level anomalous anticyclones during the 1988 U.S. drought. *J. Climate*, **11**, 2491–2504, [https://doi.org/10.1175/1520-0442\(1998\)011<2491:RWPATR>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2491:RWPATR>2.0.CO;2).
- Choat, B., T. J. Brodribb, C. R. Brodersen, R. A. Duursma, R. López, and B. E. Medlyn, 2018: Triggers of tree mortality under drought. *Nature*, **558**, 531–539, <https://doi.org/10.1038/s41586-018-0240-x>.
- Christian, J. I., J. B. Basara, J. A. Otkin, E. D. Hunt, R. A. Wakefield, P. X. Flanagan, and X. Xiao, 2019: A methodology for flash drought identification: Application of flash drought frequency across the United States. *J. Hydrometeorol.*, **20**, 833–846, <https://doi.org/10.1175/JHM-D-18-0198.1>.
- Dirmeyer, P. A., S. Halder, and R. Bombardi, 2018: On the harvest of predictability from land states in a global forecast model. *J. Geophys. Res. Atmos.*, **123**, 13 111–13 127, <https://doi.org/10.1029/2018JD029103>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rodgers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, <https://doi.org/10.1029/2002JD003296>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Griffies, S. M., 2012: Elements of the Modular Ocean Model (MOM). GFDL Ocean Group Tech. Rep. 7, 618 pp., http://mdl-mom5.herokuapp.com/web/docs/project/MOM5_elements.pdf.
- , and Coauthors, 2005: Formulation of an ocean model for global climate simulations. *Ocean Sci.*, **1**, 45–79, <https://doi.org/10.5194/os-1-45-2005>.
- He, M., J. S. Kimball, Y. Yi, S. Running, K. Guan, K. Jensco, B. Maxwell, and M. Maneta, 2019: Impacts of the 2017 flash drought in the US Northern plains informed by satellite-based evapotranspiration and solar-induced fluorescence. *Environ. Res. Lett.*, **14**, 074019, <https://doi.org/10.1088/1748-9326/ab22c3>.
- He, W., and Coauthors, 2018: Large-scale droughts responsible for dramatic reductions of terrestrial net carbon uptake over North America in 2011 and 2012. *J. Geophys. Res. Biogeosci.*, **123**, 2053–2071, <https://doi.org/10.1029/2018JG004520>.
- Herrera-Estrada, J. E., J. A. Martinez, F. Dominguez, K. L. Findell, E. F. Wood, and J. Sheffield, 2019: Reduced moisture transport linked to drought propagation across North America. *Geophys. Res. Lett.*, **46**, 5243–5253, <https://doi.org/10.1029/2019GL082475>.
- Hoerling, M. P., S. D. Schubert, and K. Mo, 2013: An interpretation of the origins of the 2012 central Great Plains drought. NOAA Drought Task Force Assessment Rep., 44 pp., <https://www.esrl.noaa.gov/psd/csi/factsheets/pdf/noaa-gp-drought-assessment-report.pdf>.
- , J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. D. Schubert, and R. Seager, 2014: Causes and predictability of the 2012 Great Plains drought. *Bull. Amer. Meteor. Soc.*, **95**, 269–282, <https://doi.org/10.1175/BAMS-D-13-00055.1>.
- Hoskins, B. J., and T. Ambrizzi, 1993: Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, **50**, 1661–1671, [https://doi.org/10.1175/1520-0469\(1993\)050<1661:RWPOAR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<1661:RWPOAR>2.0.CO;2).
- Hunke, E. C., and W. H. Lipscomb, 2010: CICE: The Los Alamos Sea Ice Model documentation and software user's manual, version 4.1. Los Alamos National Laboratory Doc. LA-CC-06-012, 76 pp., https://csdms.colorado.edu/mediawiki/images/archive/2012/01/26/014026!CICE_documentation_and_software_user's_manual.pdf.
- Jiang, X., and N.-C. Lau, 2008: Intraseasonal teleconnection between North American and western North Pacific monsoons with 20-day time scale. *J. Climate*, **21**, 2664–2679, <https://doi.org/10.1175/2007JCLI2024.1>.
- Kirtman, B. P., and Coauthors, 2014: The North American multi-model 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>.
- Klein, S. A., X. Jiang, J. Boyle, S. Malyshev, and S. Xie, 2006: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.*, **33**, L18805, <https://doi.org/10.1029/2006GL027567>.
- Koster, R. D., and M. J. Suarez, 1996: Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo. 104606, Vol. 9, 60 pp., <https://gmao.gsfc.nasa.gov/pubs/docs/Koster130.pdf>.
- , and —, 2001: Soil moisture memory in climate models. *J. Hydrometeorol.*, **2**, 558–570, [https://doi.org/10.1175/1525-7541\(2001\)002<0558:SMMMICM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0558:SMMMICM>2.0.CO;2).
- , and —, 2003: Impact of land surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeorol.*, **4**, 408–423, [https://doi.org/10.1175/1525-7541\(2003\)4<408:IOSIO>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)4<408:IOSIO>2.0.CO;2).
- , —, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, <https://doi.org/10.1029/2000JD900327>.
- , and Coauthors, 2004a: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, <https://doi.org/10.1126/science.1100217>.
- , and Coauthors, 2004b: Realistic initialization of land surface states: Impacts on subseasonal forecast skill. *J. Hydrometeorol.*, **5**, 1049–1063, <https://doi.org/10.1175/JHM-387.1>.
- , S. D. Schubert, and M. J. Suarez, 2009a: Analyzing the concurrence of meteorological droughts and warm periods, with implications for the determination of evaporative regime. *J. Climate*, **22**, 3331–3341, <https://doi.org/10.1175/2008JCLI2718.1>.
- , Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009b: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, <https://doi.org/10.1175/2009JCLI2832.1>.
- , and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37**, L02402, <https://doi.org/10.1029/2009GL041677>.
- , and Coauthors, 2011: The second phase of the Global Land–Atmosphere Coupling Experiment: Soil moisture contributions to subseasonal forecast skill. *J. Hydrometeorol.*, **12**, 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- , S. D. Schubert, H. Wang, S. P. Mahanama, and A. M. DeAngelis, 2019: Flash drought as captured by reanalysis data: Disentangling the contributions of precipitation deficit and excess evapotranspiration. *J. Hydrometeorol.*, **20**, 1241–1258, <https://doi.org/10.1175/JHM-D-18-0242.1>.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*, **99**, 14 415–14 428, <https://doi.org/10.1029/94JD00483>.

- Livneh, B., and M. P. Hoerling, 2016: The physics of drought in the U.S. central Great Plains. *J. Climate*, **29**, 6783–6804, <https://doi.org/10.1175/JCLI-D-15-0697.1>.
- Lopez, H., S.-K. Lee, S. Dong, G. Goni, B. Kirtman, R. Atlas, and A. Kumar, 2019: East Asian monsoon as a modulator of U.S. Great Plains heat waves. *J. Geophys. Res. Atmos.*, **124**, 6342–6358, <https://doi.org/10.1029/2018JD030151>.
- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513, <https://doi.org/10.3402/tellusa.v34i6.10836>.
- Lorenz, R., E. B. Jaeger, and S. I. Seneviratne, 2010: Persistence of heat waves and its link to soil moisture memory. *Geophys. Res. Lett.*, **37**, L09703, <https://doi.org/10.1029/2010GL042764>.
- Ma, H.-Y., and Coauthors, 2014: On the correspondence between mean forecast errors and climate errors in CMIP5 models. *J. Climate*, **27**, 1781–1798, <https://doi.org/10.1175/JCLI-D-13-00474.1>.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>.
- , and Coauthors, 2020: GEOS-S2S version 2: The GMAO high-resolution coupled model and assimilation system for seasonal prediction. *J. Geophys. Res. Atmos.*, **125**, e2019JD031767, <https://doi.org/10.1029/2019JD031767>.
- Moon, J.-Y., B. Wang, K.-J. Ha, and J.-Y. Lee, 2013: Teleconnections associated with Northern Hemisphere summer monsoon intraseasonal oscillation. *Climate Dyn.*, **40**, 2761–2774, <https://doi.org/10.1007/s00382-012-1394-0>.
- Morcrette, C. J., and Coauthors, 2018: Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the Southern Great Plains. *J. Geophys. Res.*, **123**, 2655–2683, <https://doi.org/10.1002/2017jd027199>.
- NCDC, 2019: U.S. billion-dollar weather and climate disasters 1980–2019. NOAA Rep., 16 pp., <https://www.ncdc.noaa.gov/billions/events.pdf>.
- Otkin, J. A., M. C. Anderson, C. Hain, I. E. Mladenova, J. B. Basara, and M. Svoboda, 2013: Examining rapid onset drought development using the thermal infrared-based evaporative stress index. *J. Hydrometeorol.*, **14**, 1057–1074, <https://doi.org/10.1175/JHM-D-12-0144.1>.
- , and Coauthors, 2016: Assessing the evolution of soil moisture and vegetation conditions during the 2012 United States flash drought. *Agric. For. Meteorol.*, **218–219**, 230–242, <https://doi.org/10.1016/j.agrformet.2015.12.065>.
- , M. Svoboda, E. D. Hunt, T. W. Ford, M. C. Anderson, C. Hain, and J. B. Basara, 2018: Flash droughts: A review and assessment of the challenges imposed by rapid-onset droughts in the United States. *Bull. Amer. Meteor. Soc.*, **99**, 911–919, <https://doi.org/10.1175/BAMS-D-17-0149.1>.
- 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>.
- Pendergrass, A. G., and Coauthors, 2020: Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat. Climate Change*, **10**, 191–199, <https://doi.org/10.1038/s41558-020-0709-0>.
- Reichle, R. H., and Q. Liu, 2014: Observation-corrected precipitation estimates in GEOS-5. NASA Tech. Rep. NASA/TM-2014-104606, Vol. 35, 18 pp., <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20150000725.pdf>.
- , —, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643–1664, <https://doi.org/10.1175/JCLI-D-16-0570.1>.
- Rippey, B. R., 2015: The U.S. drought of 2012. *Wea. Climate Extremes*, **10**, 57–64, <https://doi.org/10.1016/j.wace.2015.10.004>.
- Roy, T., J. A. Martinez, J. E. Herrera-Estrada, Y. Zhang, F. Dominguez, A. Berg, M. Ek, and E. F. Wood, 2019: Role of moisture transport and recycling in characterizing droughts: Perspectives from two recent U.S. droughts and the CFSv2 system. *J. Hydrometeorol.*, **20**, 139–154, <https://doi.org/10.1175/JHM-D-18-0159.1>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, <https://doi.org/10.1175/2010BAMS3001.1>.
- , and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Schubert, S. D., H. Wang, and M. J. Suarez, 2011: Warm season subseasonal variability and climate extremes in the Northern Hemisphere: The role of stationary Rossby waves. *J. Climate*, **24**, 4773–4792, <https://doi.org/10.1175/JCLI-D-10-05035.1>.
- , —, R. D. Koster, M. J. Suarez, and P. Ya. Groisman, 2014: Northern Eurasian heat waves and droughts. *J. Climate*, **27**, 3169–3207, <https://doi.org/10.1175/JCLI-D-13-00360.1>.
- , and Coauthors, 2016: Global meteorological drought: A synthesis of current understanding with a focus on SST drivers of precipitation deficits. *J. Climate*, **29**, 3989–4019, <https://doi.org/10.1175/JCLI-D-15-0452.1>.
- Seo, E., and Coauthors, 2019: Impact of soil moisture initialization on boreal summer subseasonal forecasts: Mid-latitude surface air temperature and heat wave events. *Climate Dyn.*, **52**, 1695–1709, <https://doi.org/10.1007/s00382-018-4221-4>.
- Smith, A. B., and R. W. Katz, 2013: US billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Nat. Hazards*, **67**, 387–410, <https://doi.org/10.1007/s11069-013-0566-5>.
- Sun, S., R. Bleck, S. G. Benjamin, B. W. Green, and G. A. Grell, 2018: Subseasonal forecasting with an icosahedral, vertically quasi-Lagrangian coupled model. Part I: Model overview and evaluation of systematic errors. *Mon. Wea. Rev.*, **146**, 1601–1617, <https://doi.org/10.1175/MWR-D-18-0006.1>.
- Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bull. Amer. Meteor. Soc.*, **83**, 1181–1190, <https://doi.org/10.1175/1520-0477-83.8.1181>.
- Teng, H., G. Branstator, H. Wang, G. A. Meehl, and W. M. Washington, 2013: Probability of US heat waves affected by a subseasonal planetary wave pattern. *Nat. Geosci.*, **6**, 1056–1061, <https://doi.org/10.1038/ngeo1988>.
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo Jr., 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dyn.*, **38**, 349–362, <https://doi.org/10.1007/s00382-010-0956-2>.
- Vogel, E., M. G. Donat, L. V. Alexander, M. Meinshausen, D. K. Ray, D. Karoly, N. Meinshausen, and K. Frieler, 2019: The effects of climate extremes on global agricultural yields. *Environ. Res. Lett.*, **14**, 054010, <https://doi.org/10.1088/1748-9326/ab154b>.
- Wang, H., S. D. Schubert, R. D. Koster, Y.-G. Ham, and M. J. Suarez, 2014: On the role of SST forcing in the 2011 and 2012

- extreme U.S. heat and drought: A study in contrasts. *J. Hydrometeorol.*, **15**, 1255–1273, <https://doi.org/10.1175/JHM-D-13-069.1>.
- , —, and —, 2017: North American drought and links to northern Eurasia: The role of stationary Rossby waves. *Climate Extremes: Patterns and Mechanisms*, *Geophys. Monogr.*, Vol. 226, Amer. Geophys. Union, 197–221.
- Wang, Y.-C., and H. H. Hsu, 2019: Improving diurnal rainfall phase over the Southern Great Plains in warm seasons by using a convective triggering design. *Int. J. Climatol.*, **39**, 5181–5190, <https://doi.org/10.1002/joc.6117>.
- Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, <https://doi.org/10.1029/2011JD016048>.
- Xie, S., and Coauthors, 2019: Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function. *J. Adv. Model. Earth Syst.*, **11**, 2290–2310, <https://doi.org/10.1029/2019MS001702>.
- Zhang, C., S. Xie, S. A. Klein, H.-Y. Ma, S. Tang, K. Van Weverberg, C. J. Morcrette, and J. Petch, 2018: CAUSES: Diagnosis of the summertime warm bias in CMIP5 climate models at the ARM Southern Great Plains site. *J. Geophys. Res. Atmos.*, **123**, 2968–2992, <https://doi.org/10.1002/2017JD027200>.
- Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76**, 1077–1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.
- Zheng, X., J.-C. Golaz, S. Xie, Q. Tang, W. Lin, M. Zhang, H.-Y. Ma, and E. L. Roesler, 2019: The summertime precipitation bias in E3SM atmosphere model version 1 over the central United States. *J. Geophys. Res. Atmos.*, **124**, 8935–8952, <https://doi.org/10.1029/2019JD030662>.
- Zhu, Z., and T. Li, 2016: A new paradigm for continental U.S. summer rainfall variability: Asia–North America teleconnection. *J. Climate*, **29**, 7313–7327, <https://doi.org/10.1175/JCLI-D-16-0137.1>.