

# Twenty-First-Century Changes in U.S. Regional Heavy Precipitation Frequency Based on Resolved Atmospheric Patterns

XIANG GAO AND C. ADAM SCHLOSSER

*Massachusetts Institute of Technology Joint Program on the Science and Policy of Global Change,  
Massachusetts Institute of Technology, Cambridge, Massachusetts*

PAUL A. O'GORMAN

*Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology,  
Cambridge, Massachusetts*

ERWAN MONIER

*Massachusetts Institute of Technology Joint Program on the Science and Policy of Global Change,  
Massachusetts Institute of Technology, Cambridge, Massachusetts*

DARA ENTEKHABI

*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Manuscript received 25 July 2016, in final form 15 November 2016)

## ABSTRACT

Precipitation-gauge observations and atmospheric reanalysis are combined to develop an analogue method for detecting heavy precipitation events based on prevailing large-scale atmospheric conditions. Combinations of atmospheric variables for circulation (geopotential height and wind vector) and moisture (surface specific humidity, column and up to 500-hPa precipitable water) are examined to construct analogue schemes for the winter [December–February (DJF)] of the “Pacific Coast California” (PCCA) region and the summer [June–August (JJA)] of the Midwestern United States (MWST). The detection diagnostics of analogue schemes are calibrated with 1979–2005 and validated with 2006–14 NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA). All analogue schemes are found to significantly improve upon MERRA precipitation in characterizing the occurrence and interannual variations of observed heavy precipitation events in the MWST. When evaluated with the late twentieth-century climate model simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5), all analogue schemes produce model medians of heavy precipitation frequency that are more consistent with observations and have smaller intermodel discrepancies than model-based precipitation. Under the representative concentration pathways (RCP) 4.5 and 8.5 scenarios, the CMIP5-based analogue schemes produce trends in heavy precipitation occurrence through the twenty-first century that are consistent with model-based precipitation, but with smaller intermodel disparity. The median trends in heavy precipitation frequency are positive for DJF over PCCA but are slightly negative for JJA over MWST. Overall, the analyses highlight the potential of the analogue as a powerful diagnostic tool for model deficiencies and its complementarity to an evaluation of heavy precipitation frequency based on model precipitation alone.

## 1. Introduction

Over the last two decades, the analysis of extreme precipitation events has attracted much attention because of their significant impacts on natural and human systems. In particular, many studies have shown that

extreme precipitation events are likely to respond substantially to anthropogenically enhanced greenhouse forcing with changes in their frequency and intensity (Wehner 2005; Kharin et al. 2007; Sun et al. 2007; Kao and Ganguly 2011; Min et al. 2011; Pall et al. 2011; Dominguez et al. 2012; Kharin et al. 2013; Sillmann et al. 2013; Monier and Gao 2015). Such shifts could have dramatic ecological, economic, and sociological consequences (IPCC 2012).

---

*Corresponding author e-mail:* Xiang Gao, xgao304@mit.edu

DOI: 10.1175/JCLI-D-16-0544.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

Understanding how extreme precipitation events will change in the future and enabling consistent and robust projections is therefore important for the public and policy makers as we prepare for consequences of climate change.

Simulations with global coupled ocean–atmosphere general circulation models (GCMs) forced with projected greenhouse gas and aerosol emissions are the primary tools for assessing possible future changes in climate extremes (Kharin et al. 2007; Sun et al. 2007; Kharin et al. 2013; Sillmann et al. 2013). However, previous studies have shown that climate models generally do not correctly reproduce the frequency and intensity distribution of present-day precipitation (Dai 2006; Sun et al. 2006; Wilcox and Donner 2007; DeAngelis et al. 2013). In future projections with comprehensive climate models, studies find that there can be a wide disagreement about the sign of change or the rate of increase in precipitation extremes among models, particularly in the tropics (Sillmann et al. 2013; O’Gorman 2012; Kharin et al. 2007; Sun et al. 2007; Kharin et al. 2013). These results suggest that model differences appear to be the main source of uncertainty in the projected changes in precipitation extremes (Kharin et al. 2007). Lack of skill in climate models’ regional distributions of precipitation is largely attributed to the bulk description of poorly understood processes such as moist convection and of topographical features at the subgrid scale (1–10 km). How such processes and features are parameterized or represented with typical coarse spatial resolution of climate models (~100 km or more) varies considerably among models and this can have a large effect on the precipitation intensity distribution (e.g., Wilcox and Donner 2007).

On the other hand, it has been shown that climate models simulate fairly realistic large-scale atmospheric circulation features associated with heavy precipitation events compared to observations. DeAngelis et al. (2013) found that climate models from phase 3 of the Coupled Model Intercomparison Project (CMIP3) capture realistically the large-scale physical mechanisms linked to extreme precipitation over North America, although there exist biases in intensity of heavy and extreme precipitation among the models. Kawzoe and Gutowski (2013) showed that the climate models from phase 5 of the Coupled Model Intercomparison Project (CMIP5) produce very heavy precipitation in the upper Mississippi region under the same synoptic conditions seen in the observations. Based on regional climate model simulations of contemporary and future climates, Gutowski et al. (2008) assessed the synoptic circulations conducive to the extreme cold-season precipitation in the central United States. They showed that the model

reproduces the observed synoptic conditions for extremes even though it exhibits difficulty in simulating the precipitation intensity, and such circulation behavior is rather robust in the face of climate change. These results suggest that we can place more confidence in the circulation features associated with extreme precipitation than in the precipitation amount simulated from GCMs. In other words, analyses of model-simulated atmospheric circulation features accompanying extreme events may give more robust indication or projections of their occurrence and changes. This has, in fact, been illustrated in several studies. Hewitson and Crane (2006) demonstrated that precipitation downscaled from synoptic-scale atmospheric circulation changes in multiple GCMs can provide a more consistent projection of precipitation change than the GCMs’ precipitation. More recently, Gao et al. (2014) developed an “analogue method” to detect the occurrence of heavy precipitation events over the United States. The method employs composites to identify prevailing large-scale atmospheric conditions associated with widespread, heavy precipitation events at local scale. They found that the method, when applied to an ensemble of CMIP5 twentieth-century climate model simulations, produces heavy precipitation frequencies that are more consistent with observations in the multimodel median and that have smaller intermodel spreads as opposed to using model-simulated precipitation.

This study is a continuation of the previous work on the development and evaluation of analogue method for detecting heavy precipitation events under contemporary climate conditions (Gao et al. 2014). The motivations of this study are to answer questions such as the following: Is the superior performance of the analogue method exemplified in Gao et al. (2014) specific to certain large-scale atmospheric variables or robust across choices of alternative variables? How does the method apply for projecting heavy precipitation frequency in the future? Here we expand upon the analogue method presented in Gao et al. (2014) with additional atmospheric fields and examine the performances of the augmented methods in quantifying the present-day heavy precipitation frequency and their projected changes in response to different anthropogenic forcing scenarios using CMIP5 model simulations. In Gao et al. (2014), the analogue detection diagnostics for heavy precipitation are constructed using a combination of 500-hPa geopotential height and vertical motion as well as total precipitable water. Preliminary examination of CMIP5 model simulations under future emission scenarios indicates that the overall increasing trend of geopotential height associated with climate warming is superimposed on the anomalous dipole structure [see

Figs. 3 and 4 in Gao et al. (2014)] seen in the contemporary climate. This makes the use of geopotential-height anomalies problematic within the analogue framework for future climates (shown in section 3b). Furthermore, while the increases in precipitation extremes as the climate warms have been widely found to be associated with atmospheric water vapor content increase (Allen and Ingram 2002; Pall et al. 2007), O’Gorman and Schneider (2009) examined the scaling of the total condensation rate in extreme precipitation events and found that the amount of near-surface or low-level water vapor may be more relevant to precipitation extremes than the total column water vapor. Given these considerations, herein we evaluate how the performance of the analogue scheme constructed with 500-hPa horizontal wind vectors compares to that of the analogue scheme constructed with 500-hPa geopotential height anomalies. We are also interested in whether the analogue scheme is sensitive to the use of different variables to represent atmospheric water vapor content relevant to heavy precipitation as the climate warms, such as near-surface specific humidity, lower-tropospheric precipitable water as represented by precipitable water up to certain level (500 hPa is used here due to high orography in some regions), and total precipitable water.

The paper is organized as follows. Section 2 describes the datasets (observations, reanalysis, and climate model simulations). The development, calibration, and validation of the expanded analogue schemes are given in section 3. The evaluation of the expanded analogue schemes with the CMIP5 late twentieth-century historical climate experiment is discussed in section 4. Section 5 presents comparisons of the projected changes in heavy precipitation frequency under two CMIP5 radiative forcing scenarios based on the augmented analogue schemes and model-simulated precipitation. A summary and discussion are provided in section 6.

## 2. Datasets

### a. Observed precipitation

Daily precipitation observations were obtained from the NOAA Climate Prediction Center (CPC) unified rain gauge-based analysis (Higgins et al. 2000b). These observations, spanning from 1948 to present, are confined to the continental United States land areas and gridded to a  $0.25^\circ \times 0.25^\circ$  resolution from roughly 10 000 daily station reports. The analysis was produced using an optimal interpolation scheme and went through several types of quality control including “duplicate station” and “buddy” checks, among others. Previous assessments of gridded analyses and station observations over

the United States have shown that gridded analyses are reliable for studies of fluctuations in daily precipitation as long as the station coverage is sufficiently dense and rigorous quality control procedures are applied to the daily data (Higgins et al. 2007).

### b. NASA-MERRA reanalysis

We use Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011) to analyze the large-scale atmospheric circulations associated with the heavy precipitation, and to calibrate and validate the analogue schemes. MERRA uses the GEOS-5 atmospheric circulation model, the Catchment land surface model, and an enhanced three-dimensional variational data assimilation (3DVar) analysis algorithm. The data assimilation system of GEOS-5 implements an “incremental analysis updates” (IAU) procedure in which the analysis correction is applied to the forecast model states gradually. This has ameliorated the spindown problem with precipitation and greatly improved aspects of the stratospheric circulation. MERRA’s physical parameterizations have also been enhanced so that the shock of adjusting the model system to the assimilated data is reduced. In addition, MERRA incorporates observations from NASA’s Earth Observing Systems (EOS) satellites, particularly those from EOS/Aqua, in its assimilation framework. MERRA is updated in real time, spanning the period from 1979 to the present. The three-dimensional 3-hourly atmospheric diagnostics on 42 pressure levels are available at a  $1.25^\circ$  resolution.

### c. Climate model simulations

We use the climate model simulations from the CMIP5 historical experiment (years 1850–2005) and experiments for the twenty-first century (years 2006–2100) employing two different radiative forcing scenarios. The historical runs were forced with observed temporal variations of anthropogenic and natural forcings and, for the first time, time-evolving land cover (Taylor et al. 2012). The future scenarios, called representative concentration pathways (RCPs; Moss et al. 2010), are designed to accommodate a wide range of possibilities in social and economic development consistent with specific radiative forcing paths. The estimated radiative forcing values by year 2100 are 4.5 and  $8.5 \text{ W m}^{-2}$  in the two experiments considered here, namely RCP4.5 and RCP8.5. In comparison with Table 1 of Gao et al. (2014), model CMCC-CM and MIROC-ESM do not provide the near-surface specific humidity and vertical velocity in two RCP experiments. Removal of these two results in a total of 18 models that provide all the essential meteorological variables for the analogue schemes across the three experiments

considered here. The models are ACCESS1.0, ACCESS1.3, BCC-CSM1.1, BCC-CSM1.1-m, BNU-ESM, CanESM2, CCSM4, CNRM-CM5, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC5, MIROC-ESM-CHEM, MRI-CGCM3, and NorESM1-M (expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>). In this study, only one ensemble member from each model is analyzed.

#### *d. Data processing*

The same set of meteorological variables are assembled or derived from both the MERRA reanalysis and climate model simulations, including 500-hPa geopotential height, 500-hPa vector winds, 500-hPa vertical velocity, near-surface specific humidity, total precipitable water, precipitable water up to 500 hPa, and vertically integrated water vapor flux vector up to 500 hPa. Precipitable water up to 500 hPa is used to represent lower-level moisture; vertical integration is performed up to 500 hPa instead of, say, 850 hPa to allow for regions of high orography. The vertically integrated water vapor flux is employed here to illustrate the moisture transport feeding the heavy precipitation events in local areas (but is not used in the development of analogue schemes). The more relevant diagnostic is vapor convergence, but its estimation based on reanalysis is problematic due to the required total mass balance correction.

The 3-hourly MERRA atmospheric diagnostics are first averaged into daily values. All the daily fields, including the precipitation observation as well as the precipitation and meteorological fields from MERRA reanalysis and each CMIP5 climate model, are then regridded to the common  $2.5^\circ \times 2^\circ$  resolution via area averaging. Such a conservative regridding procedure has been shown to especially improve agreement between observed and simulated extreme precipitation metrics (Chen and Knutson 2008). The period with the greatest overlap among the CPC observations (1948–present), MERRA reanalysis (1979–present), and the CMIP5 historical experiment (1850–2005) is 1 January 1979–31 December 2005. So at each grid cell, we convert the meteorological fields of each data source to normalized anomalies based on their respective seasonal climatological mean and standard deviation of this 27-yr period. The same seasonal climatological means and standard deviations are also employed to obtain the normalized anomalies for the meteorological fields of MERRA reanalysis from 2006 to 2014 and CMIP5 two RCP experiments from 2006 to 2100.

We use the CPC observed precipitation to identify the heavy precipitation events, while the MERRA reanalysis is employed to construct the large-scale composites of

atmospheric patterns associated with identified heavy precipitation events, and to calibrate and validate the analogue schemes. The presented analogue approach allows for the characterization of the heavy precipitation frequency only. Because of the limits of deterministic predictability of weather, the reproduction of the exact heavy precipitation date is not expected when this method is applied to the CMIP5 historical simulations. Rather, our intent is to examine the collective performances of the CMIP5 models in detecting the cumulative occurrence of the heavy precipitation events under contemporary climate, to document their potential changes as climate warms—over a given spatial and temporal domain of interest—based on prevailing large-scale physical mechanisms, and to evaluate how such analogue approach compares with observations and more conventional model-simulated precipitation.

### **3. Calibration and validation of analogue method**

There is no universally appropriate definition of heavy or extreme precipitation, and Gao et al. (2014) discussed three different methods commonly used in the previous literature to identify heavy precipitation events. In this study, we follow the same definition as was used in Gao et al. (2014): a precipitation event is a daily amount above  $1 \text{ mm day}^{-1}$  at one observational or model grid at  $2.5^\circ \times 2^\circ$ , and a heavy precipitation event occurs at any grid cell when the daily amount exceeds the 95th percentile of all precipitation events at that grid cell during a specific period (season). The 95th percentile of the distribution from the precipitation observation based on contemporary climate (1979–2005) is used to extract the heavy precipitation events for MERRA reanalysis from 1979 to 2014 as well as for CMIP5 model simulations of historical experiment from 1979 to 2005 and RCP experiments from 2006 to 2100. We then pool all extracted events at all data grid cells within the regions of our interest from the observations, MERRA reanalysis and CMIP5 model simulations separately. It should be noted that at  $2.5^\circ \times 2^\circ$  grid resolution, we do not account for the “widespread” heavy precipitation events on any particular day as we did at  $0.25^\circ \times 0.25^\circ$  grid resolution in Gao et al. (2014). The MERRA reanalysis large-scale atmospheric fields from 1979 to 2005 will be used to develop and calibrate the analogue schemes, and from 2006 to 2014 to validate them.

Gao et al. (2014) demonstrated the application of analogue scheme for several regions of the United States, including the south-central United States, which is susceptible to heavy rainfall. In this study, we focus our analyses on two of those regions: the “Pacific Coast California” (PCCA) region where heavy precipitation

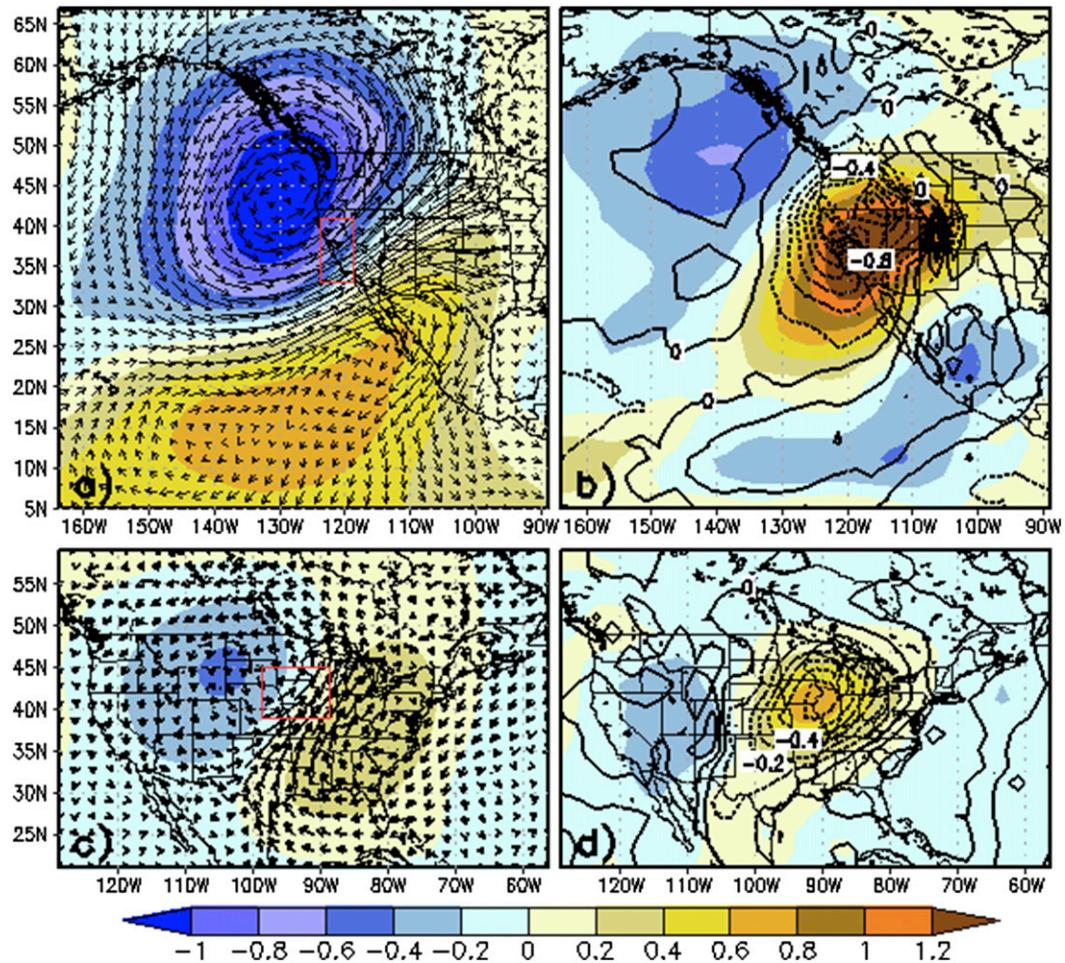


FIG. 1. Composite fields as normalized anomalies for the Pacific Coast California (PCCA) in DJF. (a) 500-hPa geopotential height (shaded,  $h_{500}$ ) and the vertical integrated water vapor flux vector up to 500 hPa (arrow) based on 165 heavy precipitation events at  $2.5^\circ \times 2^\circ$ . (b) 500-hPa vertical velocity (contour,  $\omega_{500}$ ) and total precipitable water (tpw; shaded). (c),(d) As in (a),(b), but for the Midwestern United States (MWST) in JJA based on 566 heavy precipitation events at  $2.5^\circ \times 2^\circ$ . The red rectangles depict our study regions.

events occur most frequently in the winter season [December–February (DJF)] and the Midwestern United States (MWST) where heavy precipitation events dominate mostly in the summer season [June–August (JJA); Gao et al. 2014; their Fig. 1]. PCCA, a domain bounded by  $33^\circ$ – $41^\circ$ N and  $123.75^\circ$ – $118.75^\circ$ W at  $2.5^\circ \times 2^\circ$  resolution (red rectangle in Fig. 1a), is a typical region where large-scale flows and complex topography may contribute to the occurrence of heavy precipitation events. Because of the missing values along the land–sea boundary, we use 8 grid cells out of a total of 15 grid cells in the red rectangle. For MWST, we focus on the northern U.S. Great Plains, a region bounded by  $39^\circ$ – $45^\circ$ N and  $98.75^\circ$ – $88.75^\circ$ W at the  $2.5^\circ \times 2^\circ$  resolution (20 grid cells shown as red rectangle in Fig. 1c), including the states of Kansas, Missouri, Nebraska, Iowa, Illinois, South Dakota, Minnesota, and Wisconsin. This region is

shown to be representative of an area of relatively high summer precipitation variance compared to elsewhere over the continent (Dirmeyer and Kinter 2010). Outstanding recent cases of large-scale flooding in this region include those of late spring and summer of 1993 and 2008.

#### a. Synoptic condition composites

We extract the 165 and 566 heavy precipitation events from the observations of 1979–2005 at  $2.5^\circ \times 2^\circ$  for the DJF season of PCCA and JJA season of MWST, respectively. We examine various atmospheric fields, which provide insight into the preferred synoptic conditions conducive to heavy precipitation events. Figure 1 shows the composites as standardized anomalies for two regions, produced by averaging the MERRA Reanalysis across the observed event days.

For the PCCA region, the composite shows heavy events occurring when a deep trough develops around the eastern North Pacific Ocean and an anomalous cyclonic circulation center is located to the south, promoting a southwesterly flow of moist air from near Hawaii to the West Coast of the United States (Fig. 1a). Also evident are moister air and strong upward motion centered over the northern California and Nevada, but extending into the interior of the western United States (Fig. 1b). Studies have demonstrated that major winter precipitation events along the Pacific Coast are mostly associated with the “Pineapple Express” (Higgins et al. 2000a; Warner et al. 2012). Compared with the Figs. 1a and 1b, the standardized anomalies of all the meteorological fields are weaker for the Midwestern United States. Nevertheless, the presence of lower heights to the west and higher heights to the east of the analysis region is still evident (Fig. 1c). A key ingredient for heavy precipitation in the region is the transport of warm, moist air from the Gulf of Mexico north-northeastward across the north-central United States, mainly by the general circulation as the period is not dominated by intense tropical cyclone activity (Dirmeyer and Kinter 2010). The origins of this moisture plume may extend farther south and east toward the Caribbean Sea. The composites exhibit characteristics of the “Maya Express” that fetches moisture from the subtropics or tropics, originating as evaporation from the Gulf of Mexico, eastern Mexico, or in particular the Caribbean Sea, and links into the Great Plains low-level jet, creating a much longer “atmospheric river” of moisture (Dirmeyer and Kinter 2010). Moister air and strong upward motion are also clearly observed, centered on our study region (Fig. 1d). Over both regions, these major features exhibited by various composite anomaly fields are statistically significant at the 0.05 level.

### b. Analogue detection diagnostics

In Gao et al. (2014), 500-hPa geopotential height ( $h_{500}$ ), 500-hPa vertical velocity ( $\omega_{500}$ ), and total-column precipitable water (tpw) in combination have been used to construct the analogue scheme for detecting the occurrence of heavy precipitation events. Examination of CMIP5 model simulations under future emission scenarios indicates that the overall increasing trend of geopotential height associated with climate warming disrupts the anomalous dipole structure seen in contemporary climate conditions, making its application in analogue method for future climates problematic (Fig. 2). In contrast, the distinct patterns of composite horizontal wind vector components over the study region are fairly well preserved between the contemporary and future climates. Here we examine the alternative

analogue scheme constructed with 500-hPa horizontal winds ( $uv_{500}$ ) in place of geopotential height. Besides the total precipitable water, we also assess the performance of analogue schemes based on two other atmospheric water vapor content variables relevant to heavy precipitation, namely near-surface specific humidity ( $q_{2m}$ ) and precipitable water up to 500 hPa ( $tpw_{500}$ ). The synoptic behavior exhibited by the composites of 500-hPa vertical velocity is also found to be fairly consistent between the contemporary and projected climates (not shown). This suggests that there are no apparent shifts in circulation regimes of these atmospheric variables (except for  $h_{500}$ ) associated with heavy precipitation, and can thus be applied for assessing the heavy precipitation frequency changes in a future climate. In the summer season, the influence of large-scale atmospheric dynamics is generally weaker and the role of small-scale convective processes may be greater in comparison with the winter season. It is likely that employment of atmospheric variables other than described above for analogue schemes of summer season may result in better performance. However, the aim of our study is not to find a specific analogue scheme with the best performance for each region and season examined here through an exhaustive exercise. Instead, we are interested in whether the same set of analogue schemes can perform well across different regions and seasons. Therefore, we mainly focus on the key resolved large-scale atmospheric variables associated with heavy precipitation that are widely documented in the previous literature (i.e., moisture supply, upward motion, flow of air, etc.). Then in total, we examine six combinations of atmospheric variables to construct the analogue schemes for both regions/seasons, hereafter referred to as follows:

- hw<sub>500</sub>q<sub>2m</sub> = 500-hPa height and vertical wind, as well as near-surface specific humidity
- hw<sub>500</sub>tpw<sub>500</sub> = 500-hPa height and vertical wind, as well as total precipitable water to 500 hPa
- hw<sub>500</sub>tpw = 500-hPa height and vertical wind, as well as total-column precipitable water
- uvw<sub>500</sub>q<sub>2m</sub> = 500-hPa horizontal and vertical winds, as well as near-surface specific humidity
- uvw<sub>500</sub>tpw<sub>500</sub> = 500-hPa horizontal and vertical winds, and total precipitable water to 500 hPa
- uvw<sub>500</sub>tpw = 500 hPa horizontal and vertical winds, as well as total-column precipitable water

We employ two metrics, the “hotspot” and the spatial anomaly correlation coefficient (SACC), to characterize the distinct synoptic conditions conducive to heavy precipitation events shown in composites (Gao et al. 2014). The hotspot metric diagnoses the extent to which

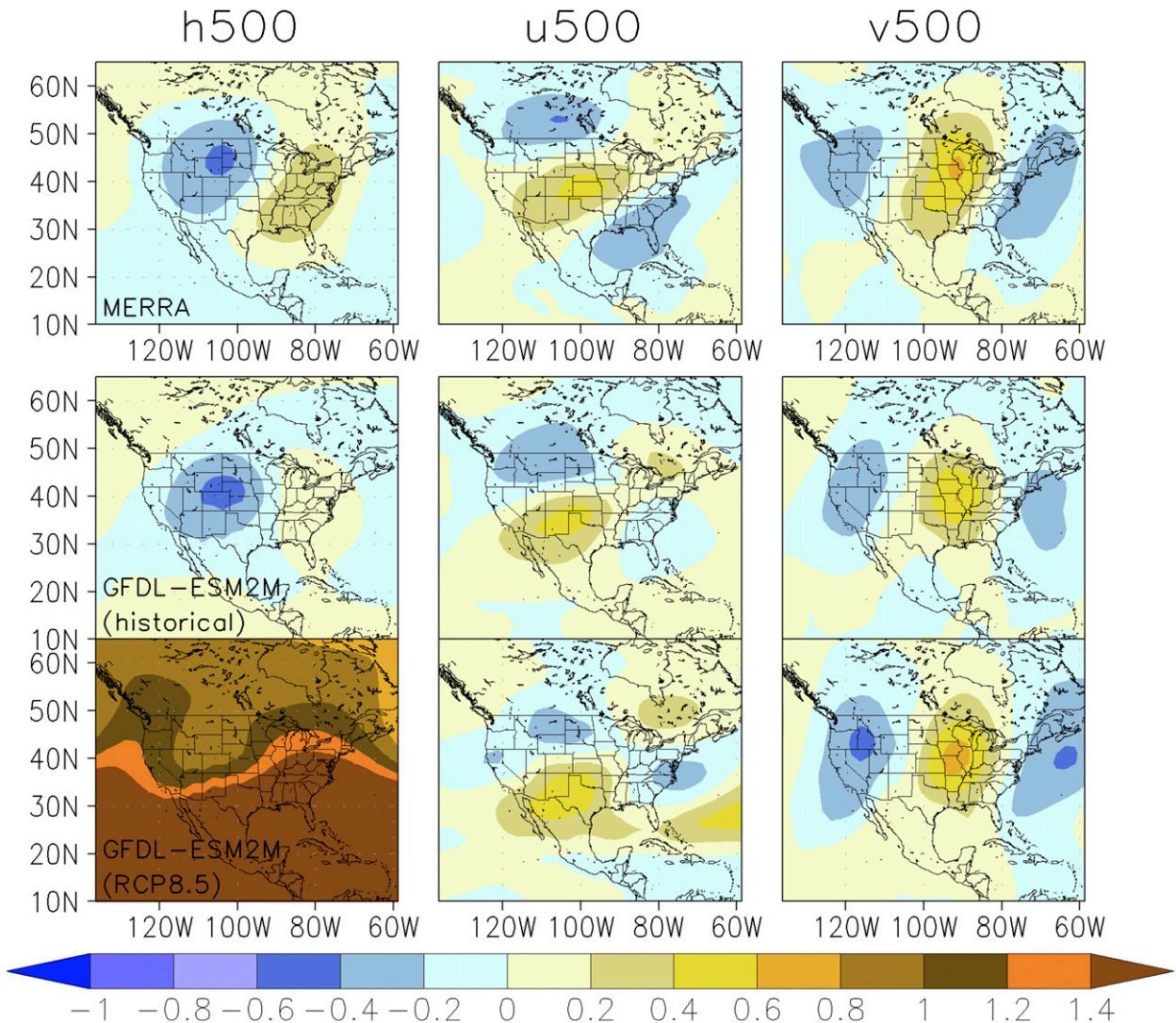


FIG. 2. Comparison of composite fields of 500-hPa (left) geopotential height ( $h_{500}$ ), and (middle) zonal ( $u_{500}$ ) and (right) meridional ( $v_{500}$ ) wind as normalized anomalies from (top) MERRA reanalysis based on 566 heavy precipitation events, and from an example of CMIP5 models (GFDL-ESM2M) based on extracted heavy precipitation events of (middle) 1979 to 2005 (historical) and (bottom) 2067 to 2093 (RCP8.5 scenario) using the model's 95th percentiles of 1979–2005 for the MWST region in JJA.

the composite of each atmospheric field is representative of any individual event. It involves the calculation of sign count at each grid cell by recording the number of individual events whose standardized anomalies have consistent sign with the composite. Hotspots are identified as the grid cells where the events used to construct the composites exhibit strong sign consistency with the composite itself (i.e., the larger sign counts). SACC is calculated between the MERRA atmospheric fields and the corresponding composites for each day of DJF or JJA from 1979 to 2005. The exact region used for SACC calculation is arbitrary, but its boundaries are chosen such that the coherent structures of the composite fields are captured and centered. We then assess 10 ranges of

SACC thresholds from 0.0 to 1.0 with an interval of 0.1. We tested the SACC calculations for regions with small differences in their size and aspect ratio, but find that the resulting optimal thresholds (described later) are insensitive to these differences for all the analogue combinations examined.

We follow the same “criteria of detection” for detection of heavy precipitation events as was used for the analogue scheme  $hw_{500tpw}$  in Gao et al. (2014), but we adapt them to the use of horizontal vector winds and other water vapor content variables, simply by treating two horizontal wind components as two variables corresponding to the trough and ridge of geopotential height. The criteria are that 1) at least three out of four

variables have consistent signs with the corresponding composites over the selected hotspot grid cells; 2) at least one out of three variables has a SACC value larger than the determined thresholds; and 3) all the SACC values have to be positive. This last criterion is only applied for DJF of PCCA as we find that it is too strict for JJA of MWST [resulting in too few heavy precipitation events in calibration; this is likely attributable to the relatively weaker strength of all the composite anomaly fields in comparison with DJF of PCCA (Fig. 1), and this is also consistent with a lower degree of consistency over the hotspots].

### c. Calibration and validation

For each of the six analogue schemes, we employ automatic calibration to determine the cutoff values for the number of hotspots and thresholds for SACC of all relevant atmospheric fields simultaneously (e.g.,  $h$ ,  $\omega$ , and  $tpw$ ). The calibration is performed by running different combinations of the number of hotspots and ranges of SACC values across all relevant atmospheric fields, and assessing the daily MERRA atmospheric fields in DJF or JJA from 1979 to 2005 to determine whether the criteria of detection described above are met for that day. If so, the day is considered as having a heavy precipitation event occurring. We use the “confusion matrix” commonly employed in the binary classification as goodness-of-fit criteria to evaluate how well the analogue schemes reproduce the observed heavy precipitation events. The same measures are also employed to assess how well the analogue schemes with optimized threshold values apply to the validation period from year 2006 to 2014, and how well the analogues perform compared to MERRA precipitation.

Confusion matrix features four values, namely, the number of true positives (TP), false positives (FP; type I error), true negatives (TN), and false negatives (FN; type II error). We employ five more metrics as performance measures derived from these four numbers:

- 1) *True positive rate (TPR)* measures the proportion of positives (i.e., extremes) that are correctly identified as such:

$$TPR = TP/(TP + FN).$$

- 2) *False positive rate (FPR)* measures the proportion of negatives (i.e., nonextremes) that are incorrectly identified as positives (i.e., extremes):

$$FPR = FP/(FP + TN).$$

- 3) *Precision or positive predictive value (PPV)* is the ratio of true positives to combined true and false positives:

$$PPV = TP/(TP + FP).$$

- 4) *Accuracy (ACC)* is the ratio of combined true positives and negatives to total population:

$$ACC = (TP + TN)/(TP + FP + TN + FN).$$

- 5) *F1 score*, a single measure of performance for the positive class, is the harmonic mean of precision and true positive rate and is calculated as shown:

$$F1 = 2TP/(2TP + FP + FN).$$

Accuracy, although widely used to evaluate the robustness of a model for making predictions, is not a reliable metric for the real performance of a classifier because it will yield misleading results if the dataset is unbalanced (i.e., when the number of samples in different classes vary greatly), just like the case of extreme versus nonextreme events. The additional meaningful measures to evaluate such a classifier are precision and true positive rate, which can be thought of as measures of a classifier exactness and completeness, respectively. A low precision and low true positive rate indicate a large number of false positives and false negatives, respectively. F1 score conveys the balance between the precision and the true positive rate.

In our study, the optimal cutoff values for the number of hotspots and thresholds for SACC are chosen as the combination of values and thresholds that produce the observed number of heavy precipitation events (equal to TP + FP) with the best TPR. In this case, FP is equal to FN, and the F1 score is equal to PPV and TPR.

Table 1 shows performance measures of using various analogue schemes to detect heavy precipitation events in DJF of PCCA during calibration (1979–2005) and validation (2006–14) periods. MERRA precipitation has better performance metrics than the analogue schemes, with higher TPRs, PPVs, and F1 scores, slightly higher ACCs, and slightly lower FPRs. The TPRs, PPVs, F1 scores, ACCs, and FPRs during the calibration period are 53%–58%, 53%–58%, 53%–58%, 94%, and 3% across analogue schemes in comparison with 58%, 66%, 62%, 95%, and 2% for MERRA precipitation. Performances during the validation period are worse than those during the calibration period for both MERRA precipitation and analogue schemes, with lower TPRs, PPVs, and F1 scores. The FPRs and ACCs are fairly insensitive measures with only minor changes. The TPRs, PPVs, F1 scores, ACCs, and FPRs are 35%–40%, 43%–51%, 39%–44%, 94%, and 3% across analogue schemes in comparison with 42%, 53%, 47%, 94%, and 2% for MERRA precipitation. Small changes in ACC values

TABLE 1. Calibration and validation statistics with different combinations of atmospheric variables to construct analogue diagnostics for DJF of PCCA. FNR and TNR are not included in the table as they can be simply derived from TPR and FPR, respectively. The numbers in bold indicate better performance in analogues than in MERRA precipitation. The numbers in parentheses indicate the total number of observed heavy precipitation events. The numbers in italics indicate the statistics from MERRA reanalysis.

Scheme	TPR	FPR	ACC	PPV	F1 score	Total events
1979–2005 (165)						
MERRA	<i>0.582</i>	<i>0.022</i>	<i>0.951</i>	<i>0.658</i>	<i>0.618</i>	<i>146</i>
hw <sub>500</sub> q <sub>2m</sub>	0.539	0.033	0.938	0.539	0.539	165
hw <sub>500</sub> tpw <sub>500</sub>	0.564	0.032	0.941	0.564	0.564	165
hw <sub>500</sub> tpw	0.570	0.031	0.942	0.570	0.570	165
uvw <sub>500</sub> q <sub>2m</sub>	0.533	0.034	0.937	0.533	0.533	165
uvw <sub>500</sub> tpw <sub>500</sub>	0.564	0.032	0.941	0.564	0.564	165
uvw <sub>500</sub> tpw	0.576	0.031	0.943	0.576	0.576	165
2006–14 (48)						
MERRA	<i>0.417</i>	<i>0.024</i>	<i>0.943</i>	<i>0.526</i>	<i>0.465</i>	<i>38</i>
hw <sub>500</sub> q <sub>2m</sub>	0.375	<b>0.022</b>	0.942	0.514	0.434	44
hw <sub>500</sub> tpw <sub>500</sub>	0.375	0.029	0.936	0.450	0.409	40
hw <sub>500</sub> tpw	0.354	0.030	0.933	0.425	0.386	40
uvw <sub>500</sub> q <sub>2m</sub>	0.396	0.026	0.940	0.487	0.437	39
uvw <sub>500</sub> tpw <sub>500</sub>	0.354	0.026	0.937	0.459	0.400	37
uvw <sub>500</sub> tpw	0.375	0.025	0.940	0.486	0.423	37

across two periods and two analyses (MERRA precipitation vs analogue schemes) are mostly attributed to our unbalanced dataset with nonextreme events (and thus TN) occupying the large portion, whereas small changes in FPR values are associated with both the dominance of TN and the same order of magnitude of detected total events (and thus FP) by the two analyses. Among the three water vapor content analogues, there is no clearly superior choice in terms of performance. During the calibration period, the schemes with tpw and tpw<sub>500</sub> perform similarly and slightly better than those with q<sub>2m</sub>. During the validation period, the schemes with q<sub>2m</sub> display a marginal improvement over those with tpw and tpw<sub>500</sub>. Furthermore, the analogue schemes with uv<sub>500</sub> have comparable performances to their geopotential height counterparts during both periods.

Table 2 shows similar statistics to Table 1, but for JJA of MWST. Immediately evident is the poorer performance of MERRA precipitation for MWST than for PCCA during both periods, with much lower TPRs (35% and 26% decrease for calibration and validation, respectively), ACCs (14% and 20% decrease), and F1 scores (27% and 20% decrease). However, PPVs are higher because they are mostly associated with the partition of predicted heavy precipitation events between TP and FP. Note that MERRA precipitation gives a much lower number of heavy precipitation events (30% and 20%) in comparison with the observation. Nevertheless, the relatively larger portion of TP results in

TABLE 2. As in Table 1, but for JJA of MWST.

Scheme	TPR	FPR	ACC	PPV	F1 score	Total events
1979–2005 (566)						
MERRA	<i>0.226</i>	<i>0.021</i>	<i>0.807</i>	<i>0.757</i>	<i>0.348</i>	<i>169</i>
hw <sub>500</sub> q <sub>2m</sub>	<b>0.549</b>	0.133	0.795	0.549	<b>0.549</b>	566
hw <sub>500</sub> tpw <sub>500</sub>	<b>0.564</b>	0.129	0.801	0.564	<b>0.564</b>	566
hw <sub>500</sub> tpw	<b>0.560</b>	0.130	0.800	0.560	<b>0.560</b>	566
uvw <sub>500</sub> q <sub>2m</sub>	<b>0.572</b>	0.126	0.805	0.572	<b>0.572</b>	566
uvw <sub>500</sub> tpw <sub>500</sub>	<b>0.571</b>	0.127	0.804	0.571	<b>0.571</b>	566
uvw <sub>500</sub> tpw	<b>0.567</b>	0.128	0.803	0.567	<b>0.567</b>	566
2006–14 (244)						
MERRA	<i>0.16</i>	<i>0.023</i>	<i>0.739</i>	<i>0.74</i>	<i>0.265</i>	<i>50</i>
hw <sub>500</sub> q <sub>2m</sub>	<b>0.410</b>	0.132	0.733	0.565	<b>0.475</b>	177
hw <sub>500</sub> tpw <sub>500</sub>	<b>0.451</b>	0.137	<b>0.742</b>	0.579	<b>0.507</b>	190
hw <sub>500</sub> tpw	<b>0.459</b>	0.142	<b>0.740</b>	0.574	<b>0.510</b>	195
uvw <sub>500</sub> q <sub>2m</sub>	<b>0.434</b>	0.130	<b>0.742</b>	0.582	<b>0.497</b>	182
uvw <sub>500</sub> tpw <sub>500</sub>	<b>0.484</b>	0.158	0.737	0.562	<b>0.520</b>	210
uvw <sub>500</sub> tpw	<b>0.475</b>	0.154	0.737	0.563	<b>0.515</b>	206

higher PPVs. Bosilovich (2013) examined the interannual variations of MERRA summertime precipitation over the United States and found that the Midwest is one of the weakest regions where significant biases exist for the seasonal mean. In contrast, the analogue schemes appear fairly robust across the two regions in terms of TPRs, PPVs, and F1 scores, with comparable and better values for MWST than for PCCA during the calibration and validation period, respectively. For MWST, the analogue schemes also tend to underestimate the number of heavy precipitation events during the validation period, but to a much lesser extent than MERRA precipitation. Both analogue schemes and MERRA precipitation exhibit performance degradation during the validation period, with lower TPRs, ACCs, PPVs, and F1 scores, but higher FPRs than those during the calibration period. All analogue schemes outperform MERRA precipitation during both periods in terms of TPRs and F1 scores. However, FPRs are higher due to the larger FP from the analogues than from MERRA precipitation, associated with the large difference in their detected total events (566 vs 169 for calibration and 50 vs 177–210 for validation). As the number of the “tagged” occurrences increases, both TPR and FPR are expected to increase accordingly. The ACCs remain fairly comparable between two analyses as they are largely dominated by TN. Similarly, there is no clearly superior choice of analogues associated with three water vapor content representations in terms of various performance measures. The analogue group with uv<sub>500</sub> shows marginal improvements over the group with h<sub>500</sub> during both periods based on most of the performance measures, but the overall differences in the performance metrics among all analogue schemes are relatively small.

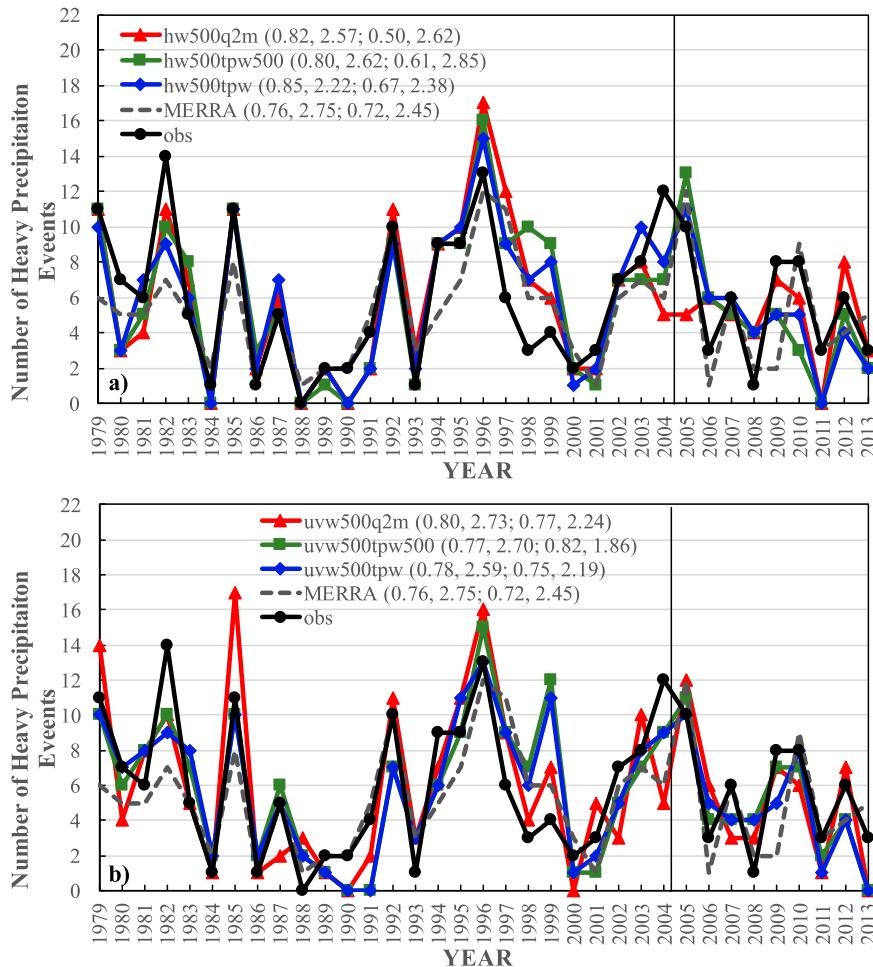


FIG. 3. Comparisons of interannual variations of seasonal heavy precipitation frequency obtained from various analogue schemes, MERRA precipitation (MERRA), and the observation (obs) for DJF of PCCA during the calibration (1979–2005) and validation (2006–14) periods: (a) the analogue scheme group with geopotential height ( $hw_{500}^*$ ) and (b) the analogue scheme group with horizontal vector winds ( $uvw_{500}^*$ ). Also shown in the parentheses of figure legend are temporal correlations and RMSE between various schemes and observation during two periods.

We also examine the performances of various analogue schemes in depicting the interannual variations of seasonal heavy precipitation frequency from 1979 to 2005 (calibration) and 2006 to 2014 (validation) as compared to the observations and MERRA precipitation over two study regions (Figs. 3 and 4). For the DJF season, the number of heavy precipitation events for each “year” is computed based on the numbers in December of the current year and the numbers in January and February of the subsequent year (thus, the results for January and February of 1979 and in December of 2014 are not included). So December 1979–February 1980 is labeled on our graphs as 1979, and so on. For PCCA, the analogue schemes and MERRA precipitation reproduce the observed interannual

variations of winter heavy precipitation frequencies reasonably well with the temporal correlation above 0.75 and a root-mean-square error (RMSE) of less than 3 days during the calibration period (Figs. 3a,b). All the analogue schemes outperform MERRA precipitation with higher correlations and smaller RMSEs. During the validation period, the analogue group with  $h_{500}$  exhibit some degradation in these statistics and do not perform as well as MERRA precipitation, whereas the analogue group with  $uv_{500}$  consistently shows better performance than MERRA precipitation (however, the difference between the correlations of the calibration and validation periods are not statistically significant at the 0.05 level for both analyses). More specifically, we find that both MERRA precipitation

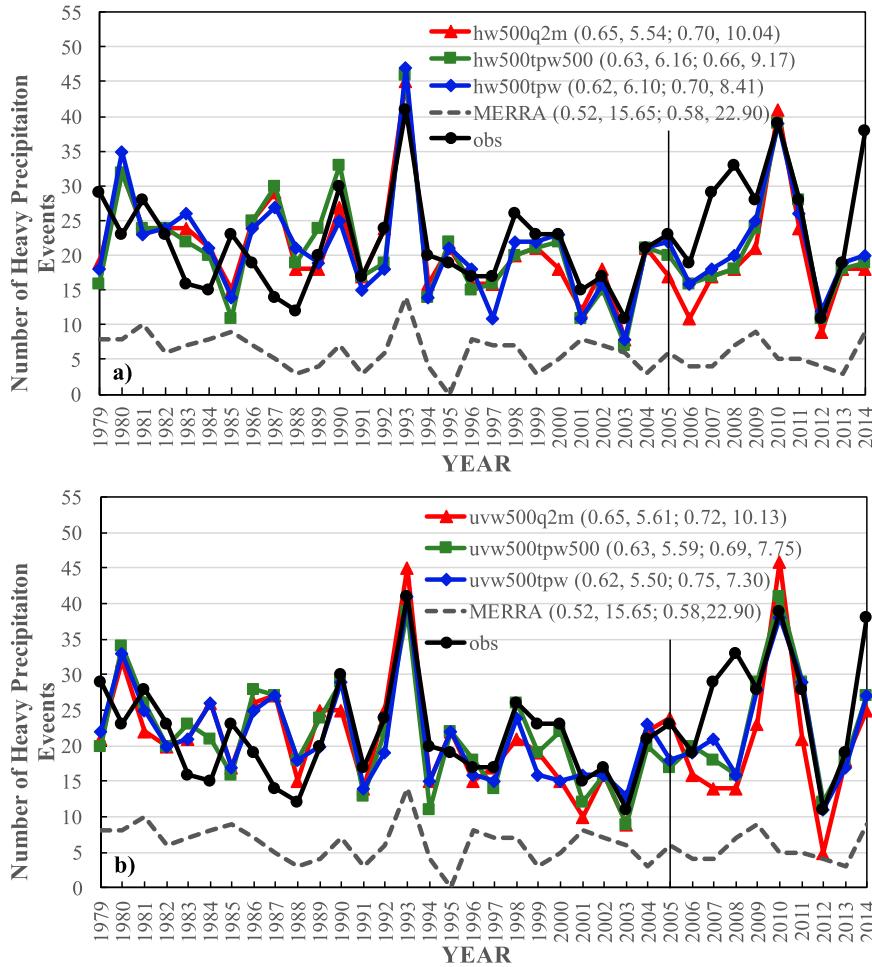


FIG. 4. As in Fig. 3, but for JJA of MWST.

and all or some analogue schemes capture peaks, such as the heavy precipitation that occurred during February 1986 and winter 1992/93, 1996/97, 2005/06, and 2010/11 as well as valleys for winter 1984/85, 1986/87, 1988/89, 1993/94, 2000/01, and 2008/09. Both analyses strongly underestimate the observed number of events for winter 1982/83 (a very strong ENSO year) and winter 2004/05 but overestimate it for winter 1997/98. MERRA precipitation also significantly underestimate the observed number of events for winter 1979/80, 1994/95, and 2009/10.

In comparison with the PCCA, MERRA precipitation in the MWST exhibits rather poor performance in tracking year-to-year variations of heavy events with lower temporal correlation (0.52 vs 0.76 for validation and 0.58 vs 0.72 for calibration) and much larger RMSE (15.65 vs 2.75 days and 22.90 vs 2.45 days). Immediately evident is its significant underestimation of heavy events throughout the entire 27-yr period. The performances of various analogue schemes are slightly

worse than for the PCCA with lower correlations (0.62–0.75) and larger RMSEs (6–10 days). The performances of the MERRA precipitation and analogue schemes degrade during the validation period in representing the magnitude of heavy precipitation frequency with much larger RMSEs than during the calibration period, but capture rather well the observed interannual variability with higher correlations (however, the difference of the correlations between the calibration and validation periods is not statistically significant at the 0.05 level). We see that various analogue schemes (especially with  $uv_{500}$ ) capture the heavy precipitation of 1990, 1993, and 2010 as well as years with relatively low frequency of events such as 1988, 1991, 1997, 2003, and 2012. The analogue schemes significantly underestimate the observed number of events for 2007/08 and 2014, but overestimate the 1980 and 1987 number of events. Nevertheless, all the analogue schemes greatly improve upon the MERRA precipitation with higher correlations and much lower RMSEs across the calibration and validation.

#### 4. Simulated late twentieth-century heavy precipitation frequency

Next we apply various analogue schemes to the CMIP5 late twentieth-century model simulations. We examine the capabilities of current state-of-the-art climate models to realistically replicate the “resolved” large-scale atmospheric conditions associated with heavy precipitation events. Validating the circulation behaviors linked to these events in climate models can ensure the assessment of their future changes with greater confidence. This is achieved by judging the CMIP5 model-simulated daily meteorological conditions of 1979 to 2005 against the constructed composites (e.g., Fig. 1) for their similarity in terms of the established criteria of detection (described in section 3b). In this way, any day when the criteria of detection are met would be considered as a heavy precipitation event. We then compare the results of the analogue schemes with the heavy precipitation events identified from the observations, MERRA precipitation, and the CMIP5 model precipitation (all at  $2.5^\circ \times 2^\circ$  resolution).

Figure 5 displays the comparisons of the number of 1979–2005 winter heavy precipitation events obtained from the CMIP5 model precipitation and various analogue schemes across 18 climate models for the PCCA region. Also included are the numbers of heavy precipitation events estimated from the observations and MERRA precipitation. We can see that the precipitation-based analyses (the “pr” whisker plot) from all the models strongly overestimate the number of heavy precipitation events, with the observation far below the minimum. Wet biases over the West Coast of the United States were also observed for the CMIP3 twentieth-century annual precipitation of all the 22 participating models against the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) (Xie and Arkin 1997) observation-based climatology (IPCC 2007; see Fig. S8.9b in the supplemental material therein). However, different models exhibit a varying degree of overestimation and the resulting heavy precipitation frequencies demonstrate a wide interquartile range (IQR;  $\sim 200$  days) and intermodel spread ( $\sim 400$  days). In contrast, the results from all the analogue schemes produce more consistent multimodel medians with the observation as well as largely reduced IQRs (25–50 days) and intermodel ranges ( $\sim 100$  days). Overall, the central tendencies of various analogue schemes are to overestimate the number of heavy precipitation events, with the observation generally falling in the first or second quartiles. Among three water vapor content representations, the analogue schemes with  $q_{2m}$  have the largest IQRs. There are no

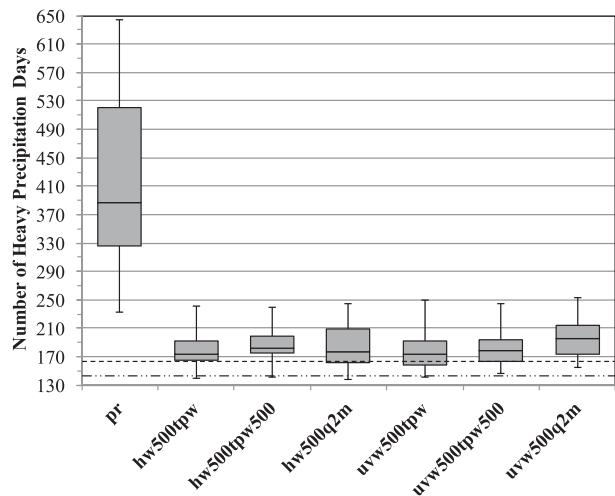


FIG. 5. Comparisons of the number of winter season (DJF) heavy precipitation events in PCCA estimated from CMIP5 model-simulated precipitation and various analogue schemes applied to CMIP5 model-simulated atmospheric synoptic conditions during the period of 1979 to 2005. The whisker plot shows the minimum, the lower and upper quartile, median, and the maximum across 18 CMIP5 models. The dashed and dash-dotted lines indicate the number of heavy precipitation events identified from the Higgins et al. (2000b) gridded observations and MERRA precipitation at  $2.5^\circ \times 2^\circ$ , respectively. The 95th percentile of 1979–2005 observed precipitation is used to extract precipitation-based heavy precipitation events from CMIP5 and MERRA.

salient differences between the performances of the analogue schemes with  $h_{500}$  versus  $uv_{500}$  in terms of the multimodel medians. MERRA precipitation is found to slightly underestimate the number of events.

Both model precipitation and analogue schemes display larger intermodel discrepancies for MWST than for PCCA (Fig. 6). In the MWST region, recycling ratios increase during summer and thus increase the dependence of precipitation on the boundary layer parameterization and the land model (through its representation of evaporation). The weaker performances of the analogue schemes are likely associated with the weaker influence of large-scale atmospheric dynamics in the summer and the greater role of convective processes. This does not necessarily indicate a poor choice of atmospheric variables for analogue schemes in MWST. Instead, the improved performance of the analogue schemes compared to MERRA summer precipitation as shown in Table 2 and Fig. 4 demonstrates their potential even for the season when the influence of the large-scale atmospheric circulation is weaker. We can see that precipitation from all 18 models and MERRA reanalysis underestimates the number of heavy precipitation events with the deviations ranging from 4 to 506 days. Such dry biases over the Midwest are consistent with the CMIP3 twentieth-century annual precipitation from a majority of

models and the multimodel mean (IPCC 2007, Fig. S8.9b therein). The analogue schemes based on  $h_{500}$  underestimate the heavy precipitation frequencies with the observation close to upper quartile, while those based on  $uv_{500}$  show slightly better performances with the observed frequency closer to median values. Nevertheless, the model medians of all analogue schemes are more consistent with the observed number of events than model-simulated precipitation and the results are also less uncertain with smaller IQRs and intermodel ranges. The analogue schemes with  $q_{2m}$  contain the largest intermodel spread, while those with  $tpw$  and  $tpw_{500}$  perform similarly.

Overall, all analogue schemes improve upon the model precipitation in terms of their assessment of late twentieth-century heavy precipitation frequency from the perspectives of both accuracy (consistencies of multimodel medians with observation) and precision (intermodel spreads) over two study regions, regardless of water vapor content variables chosen to construct the analogue scheme. This clearly suggests that current state-of-the-art climate models are capable of realistically simulating the atmospheric synoptic conditions associated with heavy precipitation events with reasonable frequencies. Accordingly, the analogue schemes based on resolved large-scale circulation features can provide more useful skill in detecting heavy precipitation events. The largest intermodel spread from the  $q_{2m}$ -based analogue scheme indicates that climate models may not be well constrained in simulating  $q_{2m}$  compared with  $tpw$  and  $tpw_{500}$ , mostly because the surface humidity in the climate models is usually controlled by a number of processes, including vertical mixing, surface evaporation (which is affected by wind speed), soil moisture, solar heating, and other factors. Similar performances between  $tpw$ -based and  $tpw_{500}$ -based analogue schemes as well as  $h_{500}$ -based and  $uv_{500}$ -based are somewhat expected as simulations of these counterparts in climate models are based on the essentially identical or similar numerical ingredients.

We further examine the consistency between the heavy precipitation frequency from the model precipitation and from all the analogue schemes on a per model basis for both study regions. Here we only show  $uv_{500}$ -based analogue schemes (Fig. 7) as their  $h_{500}$ -based counterparts give very similar results. Immediately evident is that climate models exhibit a wide range of different levels of consistency between precipitation-based and analogue-based results as well as among various analogue results over both regions. One caveat in our analyses is that unforced variability is likely responsible for some of the differences between climate models (for both precipitation and analogues) as well as between models and observations. Nevertheless, Sriver

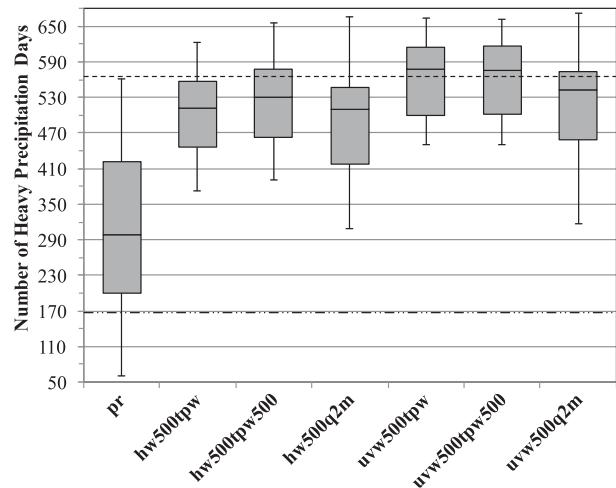


FIG. 6. As in Fig. 5, but for summer season (JJA) of MWST.

et al. (2015) demonstrated that 34 CMIP5 models yield a considerable larger spread in representing local-scale daily summer precipitation maxima than the 50 Community Earth System Model (CESM) ensemble simulations with different initial conditions—and therefore implying that intermodel biases among CMIP models still possess a larger source of discrepancy than that from internal variability. We assess the uncertainty of observed heavy precipitation frequency by performing a block bootstrap with each year as a block (nonoverlapping). Using  $\pm 2$  standard errors of observed heavy precipitation frequency calculated from 500 bootstrap samples (about 18 days for PCCA and 27 days for MWST) as thresholds for evaluation of model performance, we divide the climate models into four groups. The blue area represents the climate models that are capable of realistically simulating precipitation and large-scale circulation conditions conducive to the heavy precipitation events, while the white area is characteristic of those that are rather poor in both regards. The purple area represents climate models with realistically simulated synoptic conditions but not precipitation, while the pink is opposite to the purple. For both study regions, none of the climate models fall into the blue area, while several fall into the white region with neither precipitation nor any of analogue-based frequencies close to the observations. A majority of models fall into the purple region with some or all analogue-based frequencies consistent with observation. An extreme case of this group is the climate model A, which shows strong consistency and robustness in simulating three atmospheric water vapor content variables, reasonably frequent and realistically simulated atmospheric synoptic conditions linked to heavy precipitation events, and an apparent disconnection between model precipitation and large-scale

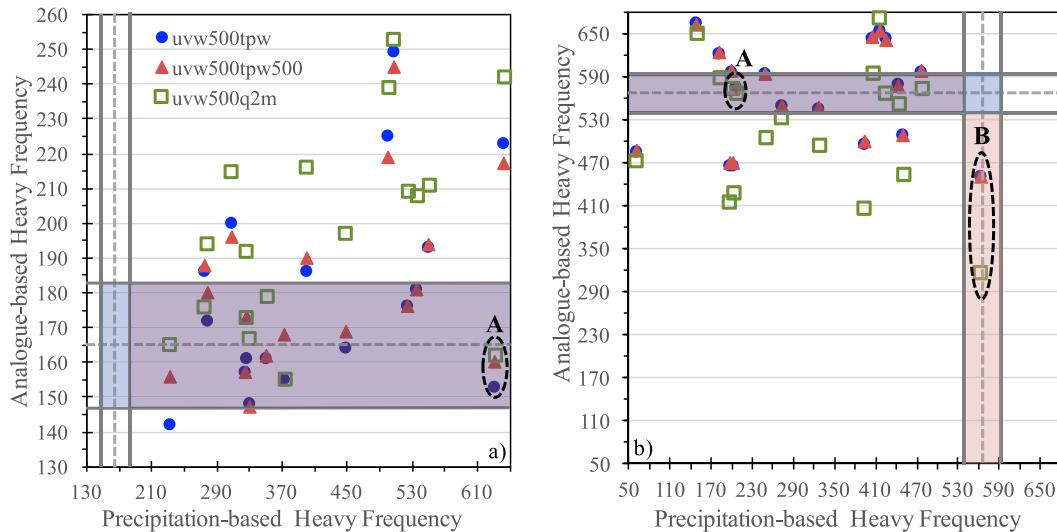


FIG. 7. Scatterplots of late twentieth-century CMIP5 heavy precipitation frequency based on model precipitation and three  $uv_{500}$  analogue schemes for (a) DJF of PCCA and (b) JJA of MWST. The two dashed gray lines represent observed heavy precipitation frequencies. The solid gray lines represent  $\pm 2$  standard errors of the observed heavy precipitation frequency calculated using 500 bootstrap samples.

circulation features. For model A the heavy precipitation frequencies from the three analogue schemes match well with the observations, but there exists a large bias in precipitation-based frequency. The large portion of climate models in this group further emphasize the need to better understand the influence of processes such as moist convection and topographical features at the sub-grid scales and to improve their parameterizations for precipitation calculation in climate models. Only one model (model B) falls into the pink area in the MWST; it has correctly simulated heavy precipitation frequency but the three atmospheric water vapor content variables are not consistent with each other or with the observations and model precipitation. Furthermore, regardless of what region the climate models lie in, the consistency among different atmospheric water vapor content variables is not always guaranteed. As expected, tpw and  $tpw_{500}$  are more consistent with each other in comparison with  $q_{2m}$ , especially in the MWST. In summary, various climate models demonstrate different skills in reproducing precipitation and large-scale circulation features, and therefore choices of analogue schemes based on different atmospheric variables can lead to different skills in detecting heavy precipitation events. Through such analyses, the analogue method can be potentially employed as a powerful diagnostic tool to evaluate the representation of heavy precipitation events in climate models, and the diagnosed model deficiencies can further provide useful insights into model development and improvement.

Given the comparable performances of the analogue schemes based on  $uv_{500}$  to those based on  $h_{500}$  and the

aforementioned complication of geopotential height changes under warming climate, we will employ only the  $uv_{500}$ -based analogue schemes to assess the projected changes in heavy precipitation frequency in the next section.

## 5. Projected future changes in heavy precipitation frequency

We use the 95th percentile values of the 1979–2005 seasonal precipitation observations to extract the heavy precipitation events of RCP experiments from 2006 to 2100. The use of fixed thresholds is one of the ways to examine how the predefined events (i.e., heavy or extreme precipitation) migrate in a changing climate. We convert the CMIP5 model-simulated daily meteorological fields from 2006 to 2100 to normalized anomalies relative to the seasonal climatological means and standard deviations of each model from the CMIP5 historical simulations (1979–2005). We analyze the projected changes in heavy precipitation frequency during seven 27-yr periods centered at the years 2020, 2030, 2040, 2050, 2060, 2070, and 2080, respectively. So the first period spans from 2007 to 2033, and so on. The change of each model is calculated relative to its respective seasonal heavy precipitation frequency from 1979 to 2005 and expressed as number of events per year. This is done for both model-based precipitation and the three analogue schemes based on  $uv_{500}$ .

Figure 8 displays the general evolution of the changes in heavy precipitation frequency estimated from an

ensemble of model precipitation and the analogue scheme  $uvw_{500tpw}$  under the RCP8.5 and RCP4.5 scenarios for DJF of PCCA. Under the RCP8.5 scenario, the multimodel medians of both analyses indicate pronounced increases in heavy precipitation frequency, with medians of precipitation and analogue results showing 1.3–2.7 and 1.3–3.1 more events per year throughout the examined periods, respectively (Fig. 8a). There is an upward trend in the medians with the largest increases occurring near or at the end of the century. The medians of the analogue results are generally larger (indicative of stronger increases) than those of the corresponding model precipitation. Both analyses show some disagreements in the sign of change, with the majority of models indicating increases in the frequency. However, the analogue results demonstrate reduced disagreements in the sign of change in comparison with model precipitation, with all the models consistently showing the increases in the frequency during five out of seven periods (including the last three). Intermodel disagreements in the magnitude of change remain larger for model precipitation than for analogue results, ranging from decrease of 3.5 to increase of 8.5 events per year and decrease of 1 to increase of 7.5 events per year throughout the entire period, respectively. Especially during the middle to late periods, the model precipitation results exhibit rather marked increases in both IQRs and intermodel spreads compared with the early periods. In contrast, IQRs and intermodel spreads in the analogue results remain fairly consistent throughout the entire period.

As expected, the increases in the frequency from both analyses are less pronounced under the lower emission scenario RCP4.5, with multimodel medians showing 0.2 fewer to 2.2 more events per year for precipitation and 0.7–2 more events per year for the analogue scheme throughout the entire period and with the larger increases occurring in the late periods (Fig. 8b). Likewise, during most of the periods, the medians of analogue results exhibit slightly stronger increases than the corresponding precipitation results. The emissions mitigation tends to shift not only the multimodel medians but also the entire distributions toward the smaller increases in heavy precipitation frequency across all the periods. As a result, both analyses show stronger disagreements in the sign of change than under RCP8.5 scenario, with more models showing decreases in frequency, especially in the early periods. However, intermodel disagreements in the magnitude of change are slightly reduced due to the smaller radiative forcing, ranging from decrease of 5 to increase of 5 events per year for precipitation and decrease of 2 to increase of 6 events per year for analogue across the entire period. Overall, the

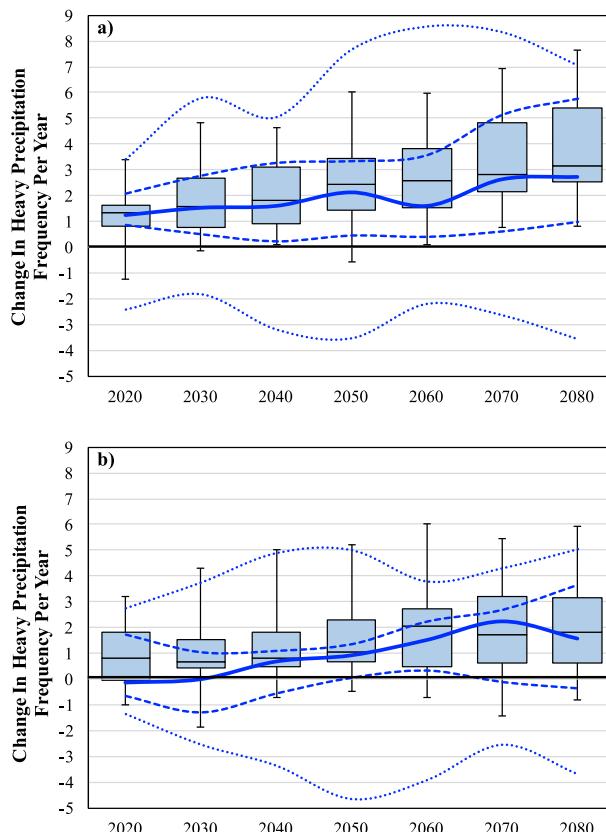


FIG. 8. The changes in heavy precipitation frequency estimated from an ensemble of CMIP5 model precipitation (blue lines) and the analogue scheme  $uvw_{500tpw}$  (whisker bar) under the (a) RCP8.5 and (b) RCP4.5 scenarios for DJF of PCCA across the periods centered at year 2020, 2030, 2040, 2050, 2060, 2070, and 2080, respectively. The solid, dashed, and dotted blue lines represent median, Q1 and Q3, and minimum and maximum values, respectively.

analogue scheme  $uvw_{500tpw}$  produces smaller intermodel spreads as compared with model precipitation during all the periods, especially under RCP8.5 scenario.

Evolutions of frequency changes from the analogue schemes  $uvw_{500tpw_{500}}$  and  $uvw_{500q_{2m}}$  illustrate very similar features to those from the analogue scheme  $uvw_{500tpw}$ , except that the multimodel medians of  $uvw_{500q_{2m}}$  demonstrate stronger increases of 1.4–3.8 and 0.8–2.0 events per year under the RCP8.5 and RCP4.5 scenarios, respectively (not shown). Figure 9 displays the comparison of frequency changes from model precipitation and three analogue schemes during the period of years 2067 to 2093 (centered at year 2080) under both RCP scenarios. All the analogue schemes improve upon model precipitation by producing reduced disagreements in the sign of frequency changes and smaller intermodel spreads, especially under the higher-emission RCP8.5 scenario. The mitigation effect of lower emissions (RCP4.5) is evident with smaller increases consistently for both analyses. Among the three analogue

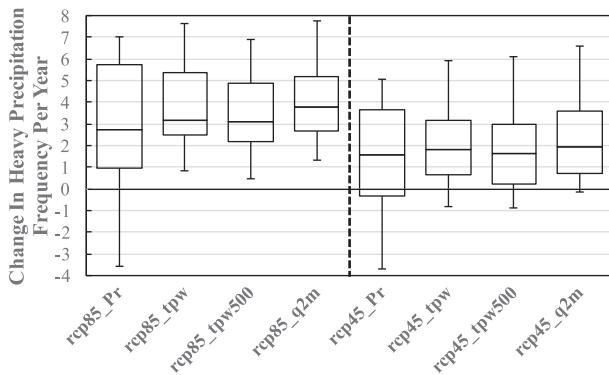


FIG. 9. The changes in heavy precipitation frequency during the period 2067–93 (centered at 2080) estimated from an ensemble of CMIP5 model precipitation and synoptic conditions employed by various analogue schemes under RCP8.5 and RCP4.5 scenarios for DJF of PCCA.

schemes, no scheme is clearly superior in consistently producing the smallest intermodel spreads under both scenarios, and this is observed during other periods as well (not shown).

The general evolution of the changes in JJA MWST heavy precipitation frequency estimated from an ensemble of model-simulated precipitation and the analogue scheme  $uvw_{500tpw}$  is displayed in Fig. 10 under the RCP8.5 and RCP4.5 scenarios. Immediately evident and distinctively different from DJF of PCCA is that the multimodel medians of both analyses generally exhibit small decreases in heavy precipitation frequency throughout the examined periods under both RCP scenarios. Wehner (2013) also reported decreases in projected midcentury summer precipitation extremes over large parts of North America based on NARCCAP regional climate model simulations. Under the RCP8.5 scenario, the multimodel medians show 0.0–0.6 fewer events per year for precipitation and 0.3–0.9 fewer events per year for the analogue scheme throughout the periods (Fig. 10a). There is no evident downward trend in the medians. Both analyses exhibit wide disagreements in the sign of change with about 50%–75% of the models showing decreases in frequency during different periods. It is worth noting the distinctively large intermodel discrepancies in the magnitude of change from model precipitation during the middle to late periods, which are more than doubled those in the early periods. By the end of the century, the discrepancies can range from an increase of six to a decrease of seven events. In contrast, the intermodel discrepancies from analogue scheme  $uvw_{500tpw}$  remain fairly constant and consistently smaller than those from model precipitation across the periods. Both analyses also produce IQRs rather consistent throughout the entire period.

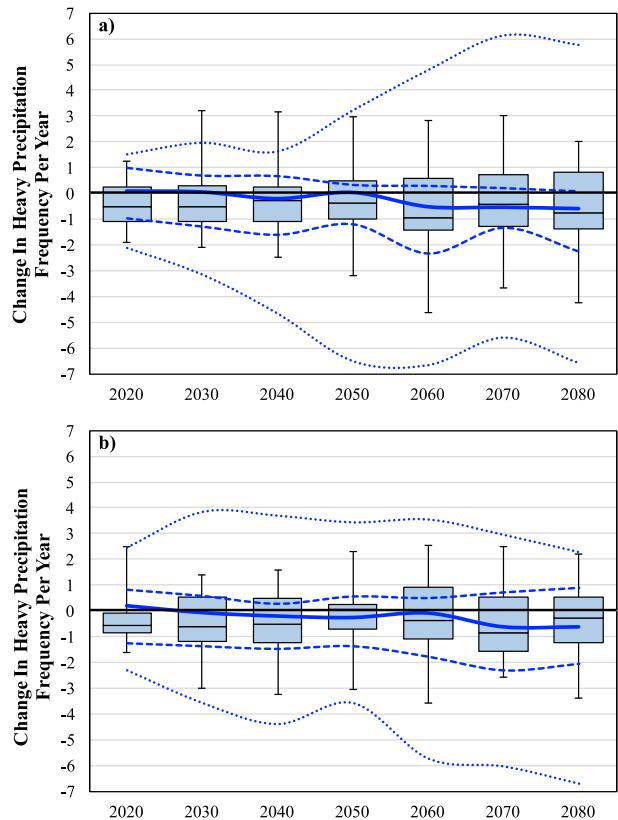


FIG. 10. As in Fig. 8, but for JJA of MWST.

The mitigation effect with the lower emissions (RCP4.5) is rather weak except that the intermodel spreads are much reduced in the middle to late periods for precipitation and in most of the periods for the analogue results. The magnitudes of change throughout the entire period range from an increase of 4 to a decrease of 7 events per year and an increase of 2.5 to a decrease of 3.5 events per year for precipitation and analogue scheme, respectively. The multimodel medians and disagreements in the sign of change from both analyses remain fairly similar to the corresponding counterparts under the RCP8.5 scenario throughout the period (Fig. 10b). Overall, the analogue scheme  $uvw_{500tpw}$  produces much smaller intermodel spreads than model precipitation during all the periods under both RCP scenarios.

We see similar characteristics in evolutions of frequency changes from the analogue schemes  $uvw_{500tpw500}$  and  $uvw_{500q2m}$  to those from the analogue scheme  $uvw_{500tpw}$ , except that their multimodel medians can show slightly stronger or slightly weaker decreases during different periods (not shown). The comparison of frequency changes from model precipitation and three analogue schemes are displayed in Fig. 11 for the last period (centered at year 2080) under both RCP scenarios. All the analogue schemes are superior to model precipitation by

producing smaller intermodel spreads of frequency changes, especially under the higher-emission RCP8.5 scenario. The mitigation of lower emission is not evident, except that the intermodel spreads are reduced for both analyses. Among three analogue schemes,  $uvw_{500q_{2m}}$  exhibits the largest intermodel discrepancies under both scenarios, which are also observed during other periods (not shown).

The correspondence between precipitation-based and each of analogue-based frequency changes on a per model basis is also examined in the last period under the RCP8.5 scenario for two study regions (Fig. 12). The degree of divergence across all the models is assessed with root-mean-squared deviation (RMSD). Over the PCCA, 16 out of 18 climate models consistently show the increases in the frequency changes from both analyses (Fig. 12a). The overall degree of divergence is 2.4, 2.3, and 2.7 events per year between precipitation-based and each of analogue-based ( $uvw_{500tpw}$ ,  $uvw_{500tpw_{500}}$ , and  $uvw_{500q_{2m}}$ ) frequency changes, respectively. The sign of the heavy frequency change is the same (positive) for all three analogues in all the models (Fig. 12a), but different models demonstrate a varying degree of consistency in the magnitude of the change with the divergence for a given model ranging from 0.1 to 2.3 events per year. The overall degree of divergence is 0.4, 0.6, and 0.9 events per year for pairs of analogue schemes  $uvw_{500tpw}$  and  $uvw_{500tpw_{500}}$ ,  $uvw_{500tpw}$  and  $uvw_{500q_{2m}}$ , and  $uvw_{500tpw_{500}}$  and  $uvw_{500q_{2m}}$ , respectively. Over the MWST, fewer climate models show the same sign of change between precipitation-based and analogue-based frequency changes. Furthermore, the sign can be opposite for different models although more models indicate decreases in precipitation-based frequency changes than increases. The overall degree of divergence is 2.5, 2.7, and 3.0 events per year between precipitation-based and each of analogue-based ( $uvw_{500tpw}$ ,  $uvw_{500tpw_{500}}$ , and  $uvw_{500q_{2m}}$ ) frequency changes, respectively, slightly larger than the corresponding values over the PCCA. We also see that, compared with the PCCA, more models show inconsistency in the sign of the frequency change for the three analogues (dashed circles in Fig. 12b). The divergences in the magnitudes of the change for a given model ranges from 0.2 to 2.7 events per year and the overall degree of divergence are 1.2, 1.3, and 1.5 events per year for pairs of analogue schemes  $uvw_{500tpw}$  and  $uvw_{500tpw_{500}}$ ,  $uvw_{500tpw}$  and  $uvw_{500q_{2m}}$ , and  $uvw_{500tpw_{500}}$  and  $uvw_{500q_{2m}}$ , respectively, slightly larger than the corresponding PCCA values as well.

In summary, the performance of model precipitation in the projected heavy precipitation frequency changes is inferior for the summer of MWST to that for

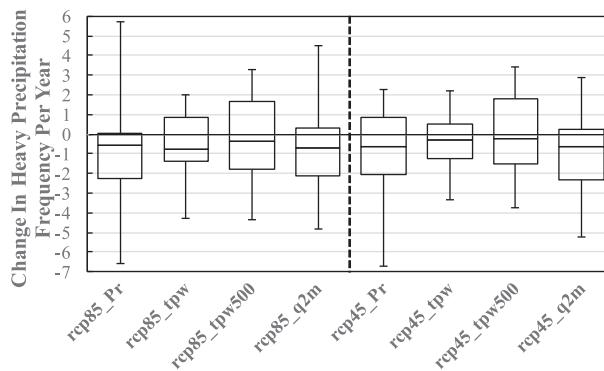


FIG. 11. As in Fig. 9, but for JJA of MWST.

the winter of PCCA in terms of larger intermodel spreads in the late periods under both RCP scenarios. Additionally, more models exhibit an inconsistent sign between precipitation-based and each of analogue-based frequency changes and the overall degree of divergences are larger. This is likely due to the regional and seasonal differences in the nature of heavy precipitation. During summer in the MWST region, land–atmosphere interactions and unresolved convection are important, leading to significant differences in model skill. Seeley and Romps (2015) also found that CMIP5 ensemble’s future changes in the frequency of environments favorable for severe thunderstorms in the central United States under RCP8.5 forcing are considerably more diverse in summer than in spring, and the disagreement on the sign of changes is closely tied to changes in boundary layer humidity. Together with the largest intermodel discrepancies exhibited by  $uvw_{500q_{2m}}$  (in comparison with  $uvw_{500tpw}$  and  $uvw_{500tpw_{500}}$ ) for the summer of MWST under both scenarios, this suggests that improving the representation of low-level humidification processes, such as the influence of soil moisture or water vapor advection from the Gulf of Mexico into the Great Plains, is likely an important step toward further constraining the climate models in assessing future heavy precipitation frequency changes, regardless of whether model precipitation or analogue scheme  $uvw_{500q_{2m}}$  is employed. Overall, the performances of various analogue schemes remain fairly consistent and robust across two seasons (regions) and RCP scenarios. The analogue-based projections improve upon precipitation-based results in terms of generally smaller intermodel discrepancies, especially under the higher-emission RCP8.5 scenario.

## 6. Summary and discussion

In this study, gridded precipitation gauge observations and atmospheric reanalysis are combined to develop an

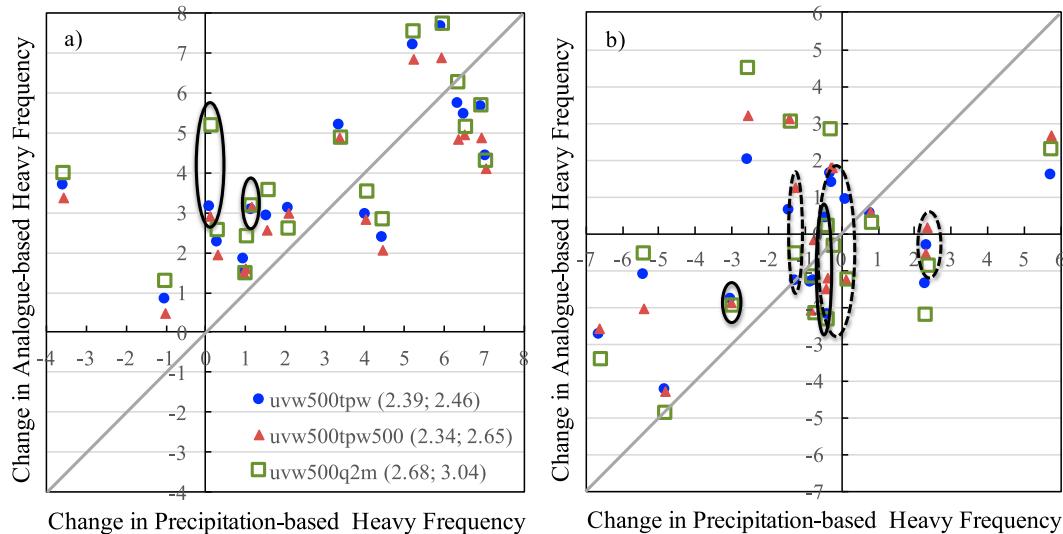


FIG. 12. Scatterplots of changes in heavy precipitation frequencies per year during 2067–93 with respect to 1979–2005 based on CMIP5 model precipitation and synoptic features employed by three  $uv_{500}$  analogue schemes for (a) the DJF of PCCA and (b) the JJA of MWST. The solid gray line represents the 1:1 line. The solid circles represent the minimum and maximum divergences in the magnitudes of frequency changes from three analogue schemes of specific climate model. The dashed circles represent the models with the inconsistency in the sign of frequency changes among three analogue schemes. The parenthesis shows the RMSDs between the precipitation-based and each of analogue-based frequency changes for the PCCA (first number) and MWST (second number).

analogue method for detecting the occurrence of heavy precipitation event based on the prevailing large-scale atmospheric conditions (“composites”). The composites are constructed for the winter season of the “Pacific Coast California” (PCCA) region and for the summer season of the Midwestern United States (MWST), where heavy precipitation exhibits typical “Pineapple Express” and “Maya Express” characteristics, respectively. The identified synoptic regimes demonstrate interactions between flow fields and regional moisture supply. Composites in both regions feature the presence of an upper-level dipole pattern associated with a trough and a ridge over a much larger spatial scale, strong flow as well as moist air and strong synoptic-scale upward motion directly over the study regions.

We examine the combinations of different atmospheric circulation variables (geopotential height and horizontal wind vectors) and water vapor content variables (near-surface specific humidity, column precipitable water, and precipitable water up to 500 hPa) to construct the analogue schemes. The detection diagnostics of various analogue schemes are first calibrated with 27-yr (1979–2005) and then validated with 9-yr (2006–14) MERRA reanalysis. The performance of MERRA precipitation in detecting the observed number of heavy precipitation events are weaker in the MWST than in the PCCA with much lower TPRs, ACCs, and F1 scores during both calibration and validation periods. In contrast, the performances of various

analogue schemes remain fairly consistent across two regions with comparable or even better TPRs, PPVs, and F1 scores in the MWST during both periods, although at the expense of FPR and ACCs. Both analyses show regional differences in representing the observed interannual variations of heavy precipitation frequencies, especially during the validation period, with lower temporal correlation but much higher RMSE against the observation in the MWST than in the PCCA. Nevertheless, various analogue schemes are found to significantly outperform MERRA precipitation in characterizing the observed number and interannual variability of heavy precipitation events in the MWST which is one of the weakest regions for MERRA summer precipitation. Among three water vapor content variables considered for the analogues, there was no superior choice. In addition, the analogue schemes based on 500-hPa horizontal wind vector ( $uv_{500}$ ) are fairly comparable to those based on 500-hPa geopotential height ( $h_{500}$ ).

With regard to the late twentieth-century (1979–2005) heavy precipitation frequencies from an ensemble of CMIP5 models, precipitation from all the models tend to strongly overestimate the winter (DJF) frequencies in the PCCA, but underestimate the summer (JJA) frequencies in the MWST. In contrast, the results from all analogue schemes based on the calibrated optimal threshold values produce more consistent multimodel medians with the observations and also have smaller intermodel spreads. This clearly indicates that the

climate models are able to realistically simulate the large-scale atmospheric conditions associated with heavy precipitation events with reasonable frequencies. Both model precipitation and analogue results display much larger divergences in the MWST than in the PCCA, possibly attributed to the increased dependence of summer precipitation on the boundary layer parameterization and the land model as well as the greater role of convection and weaker control by synoptic forcing in summer. Likewise, the performances of the analogue schemes based on  $uv_{500}$  and  $h_{500}$  are comparable to each other. Among three water vapor content representations, the analogue schemes based on  $q_{2m}$  display the largest intermodel discrepancies, likely resulting from the low degree of consensus among climate models in representing low-level humidification processes over land.

The multimodel medians of both model precipitation and  $uv_{500}$ -based analogue schemes indicate strong increases and weak decreases in heavy precipitation frequency throughout the seven 27-yr periods for the PCCA and MWST, respectively. The increases in the PCCA are more pronounced under the higher-emission scenario RCP8.5 and the largest increases usually occur near or at the end of the century. The mitigation with the lower emission (RCP4.5) tends to shift the multimodel central tendency and distributions toward smaller increases, suggesting that the climate policies adopted in the coming decades will affect the occurrence of heavy precipitation in this region. Under the RCP8.5, both model precipitation and analogue schemes demonstrate reduced disagreements in the sign of change compared to the RCP4.5, while model precipitation shows increased discrepancies in the magnitude of change, especially during the middle to late periods. In the MWST, the mitigation effect is weak with multimodel medians and disagreements in the sign of change from both analyses remaining similar under both scenarios, except that the intermodel spreads are much reduced in the middle to late periods for precipitation and slightly reduced in most of the periods for the analogue results. Regardless of the RCP scenarios and study regions, all the analogue schemes exhibit similar characteristics to one another. In the PCCA no analogue scheme is clearly superior to another, while in the MWST  $q_{2m}$ -based analogue scheme exhibits the consistently largest intermodel discrepancies under both warming scenarios. Nevertheless, all the analogue schemes improve upon model precipitation in terms of having smaller intermodel spreads, especially under the RCP8.5 scenario.

The analogue method presented here can be potentially employed as a powerful diagnostic tool to evaluate the representation of heavy precipitation, consistency in

different large-scale ingredients of heavy precipitation, and the correspondence between precipitation and these ingredients in climate models. Our analyses indicate that current state-of-the-art climate models show varying degrees of skill with significant divergence in reproducing the observed heavy precipitation in the current climate, consistently representing the large-scale ingredients, and predicting the future heavy precipitation frequency changes. On a per-model basis, the performances of precipitation-based and analogue-based results can be remarkably different in various ways and the consistency among different atmospheric water vapor content variables is not guaranteed. Therefore, choices of analogue schemes based on different large-scale ingredients can lead to different skills in detecting heavy precipitation events as well. Regardless of precipitation or analogue schemes employed, the common feature is the weaker performances in characterizing heavy precipitation events for the summer in the MWST than for the winter in the PCCA, which is likely attributed to poorly constrained low-level humidification processes among climate models and the greater importance of smaller-scale convective events in the warmer months. Such diagnosed deficiencies can thus provide useful insights into model development and improvement and further constraining the climate models in assessing heavy precipitation frequencies and their changes. Furthermore, observed rainfall intensity has been previously found to scale with convective available potential energy (CAPE) (Lepore et al. 2014), and it would be interesting to assess whether also including measures of convective instability such as CAPE would improve the accuracy of the analogue schemes, especially for summertime precipitation.

The goals of this study are to expand our previously developed analogue scheme with additional atmospheric variables, to assess the abilities of these additional schemes in detecting late twentieth-century heavy precipitation events based on the resolved large-scale atmospheric ingredients from an ensemble of CMIP5 models, and to evaluate the resulting heavy precipitation frequency changes from increasing atmospheric greenhouse gas concentrations. The analogue schemes are found to perform significantly better than the MERRA precipitation in characterizing the observed number and interannual variations of summer heavy precipitation events. They also improve upon the CMIP5 model precipitation over both study regions by producing 1) more consistent multimodel medians of late twentieth-century heavy precipitation frequencies with the observation and 2) consistent median trends in future heavy precipitation frequency but with smaller intermodel discrepancies under both climate change scenarios. It is worth noting that

the analogue method is implemented under the supposition that large-scale atmospheric conditions play a dominant role. Thus, alterations of small-scale processes associated with climate change that are not captured by the analogue schemes may introduce a bias in our assessment. Nevertheless, our results indicate that the analogue schemes based on “resolved” large-scale atmospheric features provide skillful assessments of late twentieth-century heavy precipitation frequencies and more consistent future changes from climate models, and thus the analogues show promise as improved and value-added diagnoses as compared to an evaluation that considers model precipitation alone.

**Acknowledgments.** This work was funded by NASA Energy and Water Cycle Study Research Announcement (NNH07ZDA001N), MacroSystems Biology Program Grant (NSF-AES EF#1137306) from National Science Foundation, and An Integrated Framework for Climate Change Assessment (DE-FG02-94ER61937) from the Department of Energy. PAO’G acknowledges support from NSF-AGS-1552195. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI), and the WCRP’s Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP5 multimodel dataset. We thank the NOAA Climate Prediction Center for the global gridded precipitation observations and the NASA Global Modeling and Assimilation Office for the MERRA Reanalysis data.

#### REFERENCES

- Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrological cycle. *Nature*, **419**, 224–232, doi:10.1038/nature01092.
- Bosilovich, M. G., 2013: Regional climate and variability of NASA MERRA and recent reanalyses: US summertime precipitation and temperature. *J. Appl. Meteor. Climatol.*, **52**, 1939–1951, doi:10.1175/JAMC-D-12-0291.1.
- Chen, C., and T. Knutson, 2008: On the verification and comparison of extreme rainfall indices from climate models. *J. Climate*, **21**, 1605–1621, doi:10.1175/2007JCLI1494.1.
- Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630, doi:10.1175/JCLI3884.1.
- DeAngelis, A. M., A. J. Broccoli, and S. G. Decker, 2013: A comparison of CMIP3 simulations of precipitation over North America with observations: Daily statistics and circulation features accompanying extreme events. *J. Climate*, **26**, 3209–3230, doi:10.1175/JCLI-D-12-00374.1.
- Dirmeyer, P. A., and J. L. Kinter III, 2010: Floods over the U.S. Midwest: A regional water cycle perspective. *J. Hydrometeorol.*, **11**, 1172–1181, doi:10.1175/2010JHM1196.1.
- Dominguez, F., E. Rivera, D. P. Lettenmaier, and C. L. Castro, 2012: Changes in winter precipitation extremes for the western United States under a warmer climate as simulated by regional climate models. *Geophys. Res. Lett.*, **39**, L05803, doi:10.1029/2011GL050762.
- Gao, X., C. A. Schlosser, P. Xie, E. Monier, and D. Entekhabi, 2014: An analogue approach to identify heavy precipitation events: Evaluation and application to CMIP5 climate models in the United States. *J. Climate*, **27**, 5941–5963, doi:10.1175/JCLI-D-13-00598.1.
- Gutowski, W. J., Jr., S. S. Willis, J. C. Patton, B. R. J. Schwedler, R. W. Arritt, and E. S. Takle, 2008: Changes in extreme, cold-season synoptic precipitation events under global warming. *Geophys. Res. Lett.*, **35**, L20710, doi:10.1029/2008GL035516.
- Hewitson, B. C., and R. G. Crane, 2006: Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa. *Int. J. Climatol.*, **26**, 1315–1337, doi:10.1002/joc.1314.
- Higgins, R. W., J. K. E. Schemm, W. Shi, and A. Leetmaa, 2000a: Extreme precipitation events in the western United States related to tropical forcing. *J. Climate*, **13**, 793–820, doi:10.1175/1520-0442(2000)013<0793:EPEITW>2.0.CO;2.
- , W. Shi, E. Yarosh, and R. Joyce, 2000b: Improved United States Precipitation Quality Control System and Analysis. NCEP/Climate Prediction Center Atlas No. 7. [Available online at [http://www.cpc.ncep.noaa.gov/research\\_papers/ncep\\_cpc\\_atlas/7/index.html](http://www.cpc.ncep.noaa.gov/research_papers/ncep_cpc_atlas/7/index.html).]
- , V. Silva, W. Shi, and J. Larson, 2007: Relationships between climate variability and fluctuations in daily precipitation over the United States. *J. Climate*, **20**, 3561–3579, doi:10.1175/JCLI4196.1.
- IPCC, 2007: *Climate Change 2007: The Physical Science Basis*. S. Solomon et al., Eds., Cambridge University Press, 996 pp.
- , 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. C. B. Field et al., Eds., Cambridge University Press, 589 pp.
- Kao, S. C., and A. R. Ganguly, 2011: Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios. *J. Geophys. Res.*, **116**, D16119, doi:10.1029/2010JD015529.
- Kawazoe, S., and W. J. Gutowski Jr., 2013: Regional, very heavy daily precipitation in CMIP5 simulations. *J. Hydrometeorol.*, **14**, 1228–1242, doi:10.1175/JHM-D-12-0112.1.
- Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444, doi:10.1175/JCLI4066.1.
- , —, —, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, doi:10.1007/s10584-013-0705-8.
- Lepore, C., D. Veneziano, and A. Molini, 2014: Temperature and CAPE dependence of rainfall extremes in the eastern United States. *Geophys. Res. Lett.*, **42**, 74–83, doi:10.1002/2014GL062247.
- Min, S. K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, doi:10.1038/nature09763.
- Monier, E., and X. Gao, 2015: Climate change impacts on extreme events in the United States: An uncertainty analysis. *Climatic Change*, **131**, 67–81, doi:10.1007/s10584-013-1048-1.
- Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756, doi:10.1038/nature08823.
- O’Gorman, P. A., 2012: Sensitivity of tropical precipitation extremes to climate change. *Nat. Geosci.*, **5**, 697–700, doi:10.1038/ngeo1568.
- , and T. Schneider, 2009: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl. Acad. Sci. USA*, **106**, 14773–14777, doi:10.1073/pnas.0907610106.
- Pall, P., M. R. Allen, and D. A. Stone, 2007: Testing the Clausius–Clapeyron constraint on changes in extreme precipitation under CO<sub>2</sub> warming. *Climate Dyn.*, **28**, 351–363, doi:10.1007/s00382-006-0180-2.

- , T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, **470**, 382–385, doi:10.1038/nature09762.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.
- Seeley, J. T., and D. M. Romps, 2015: The effect of global warming on severe thunderstorms in the United States. *J. Climate*, **28**, 2443–2458, doi:10.1175/JCLI-D-14-00382.1.
- Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.*, **118**, 2473–2493, doi:10.1002/jgrd.50188.
- Sriver, R. L., C. E. Forest, and K. Keller, 2015: Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble. *Geophys. Res. Lett.*, **42**, 5468–5476, doi:10.1002/2015GL064546.
- Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, doi:10.1175/JCLI3672.1.
- , —, —, and —, 2007: How often will it rain? *J. Climate*, **20**, 4801–4818, doi:10.1175/JCLI4263.1.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Warner, M. D., C. F. Mass, and E. P. Salathé Jr., 2012: Wintertime extreme precipitation events along the Pacific Northwest coast: Climatology and synoptic evolution. *Mon. Wea. Rev.*, **140**, 2021–2043, doi:10.1175/MWR-D-11-00197.1.
- Wehner, M., 2005: Changes in daily precipitation and surface air temperature extremes in the IPCC AR4 models. *U.S. CLIVAR Variations*, Vol. 3, No. 3, U.S. Climate Variability and Predictability Program, Washington, D.C., 5–9.
- , 2013: Very extreme seasonal precipitation in the NARCCAP ensemble: Model performance and projections. *Climate Dyn.*, **40**, 59–80, doi:10.1007/s00382-012-1393-1.
- Wilcox, E. M. and L. J. Donner, 2007: The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. *J. Climate*, **20**, 53–69, doi:10.1175/JCLI3987.1.
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.