

# Identifying weather regimes for regional-scale stochastic weather generators

Nasser Najibi  | Sudarshana Mukhopadhyay | Scott Steinschneider

Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York

## Correspondence

Nasser Najibi, Department of Biological and Environmental Engineering, Cornell University, 111 Wing Drive, Riley-Robb Hall, Ithaca, NY, 14853.  
Email: nn289@cornell.edu

## Funding information

National Science Foundation, Grant/Award Number: AGS1702273; U.S. Department of Agriculture, Grant/Award Number: 2019-67019-30122

## Abstract

Weather regime based stochastic weather generators (WR-SWGs) have recently been proposed as a tool to better understand multi-sector vulnerability to deeply uncertain climate change. WR-SWGs can distinguish and simulate different types of climate change that have varying degrees of uncertainty in future projections, including thermodynamic changes (e.g., rising temperatures, Clausius-Clapeyron scaling of extreme precipitation) and dynamic changes (e.g., shifting circulation and storm tracks). These models require the accurate identification of WRs that are representative of both historical and plausible future patterns of atmospheric circulation, while preserving the complex space–time variability of weather processes. This study proposes a novel framework to identify such WRs based on WR-SWG performance over a broad geographic area and applies this framework to a case study in California. We test two components of WR-SWG design, including the method used for WR identification (Hidden Markov Models (HMMs) vs. K-means clustering) and the number of WRs. For different combinations of these components, we assess performance of a multi-site WR-SWG using 14 metrics across 13 major California river basins during the cold season. Results show that performance is best using a small number of WRs (4–5) identified using an HMM. We then juxtapose the number of WRs selected based on WR-SWG performance against the number of regimes identified using metastability analysis of atmospheric fields. Results show strong agreement in the number of regimes between the two approaches, suggesting that the use of metastable regimes could inform WR-SWG design. We conclude with a discussion of the potential to expand this framework for additional WR-SWG design parameters and spatial scales.

## KEY WORDS

hidden Markov models, metastability, stochastic weather generator, weather regime

## 1 | INTRODUCTION

Stochastic weather generators (SWGs) are statistical models that simulate synthetic time series of meteorological variables (e.g., precipitation, temperature, wind speeds) based on the statistical characteristics of observed

weather (Richardson, 1981; Wilks, 1998). These computationally inexpensive models are widely used to generate long synthetic meteorological records in lieu of limited observational data (e.g., Khazaei *et al.*, 2012; Forsythe *et al.*, 2014; Breinl *et al.*, 2017). Such records can help characterize past and future climate vulnerabilities of

water resources, agricultural, or other environmental systems (Wilks, 1992; Mavromatis and Jones, 1998; Wilks and Wilby, 1999; Ivanov *et al.*, 2007; Birt *et al.*, 2010; Chen and Brissette, 2014; Schlabing *et al.*, 2014).

Recently, weather-regime (WR) based SWGs (WR-SWGs) have emerged as a tool to develop large ensembles of physically-based scenarios for climate change impact assessments over large spatial domains (Kavvas *et al.*, 1988; Bardossy and Plate, 1991; Hay *et al.*, 1991; Corte-Real *et al.*, 1999; Ailliot *et al.*, 2015; Steinschneider *et al.*, 2019). WRs are defined as recurring, large-scale patterns of atmospheric flow (e.g., upper-level, quasi-stationary blocks and troughs) that appear at fixed geographic locations, persist for days-to-weeks within a season (typically beyond the lifetime of individual storms), and organize high-frequency weather systems (Reinhold *et al.*, 1982; Legras *et al.*, 1985; Robertson and Ghil, 1999). These regimes can be interpreted as quasi-stationary attractors within the state space of the atmosphere; the trajectory of the atmospheric state intermittently collapses towards and persists around these attractors before transitioning into another regime (Hannachi *et al.*, 2017). The recurrence and persistence of WRs regulate the frequency of dry and wet periods, as well as the likelihood of extreme events over multiple space-time scales (Michelangeli *et al.*, 1995; Agel *et al.*, 2018; Zhang and Villarini, 2019). Accordingly, WR-SWGs utilize simulations of WRs, often based on a Markovian process, to organize the stochastic simulation of local weather, thus establishing a hierarchy within the simulation scheme that links low-frequency, global-scale components of atmospheric circulation with higher-frequency weather systems. This is particularly beneficial when developing future climate change scenarios over large spatial domains (i.e., regional to national scales) for which global climate models (GCMs) are most credible (Johnson and Sharma, 2009; Farnham *et al.*, 2018). GCM-based climate projections can be used to infer plausible shifts in large-scale climate dynamics (e.g., variations in the frequency of different WRs), which can be parameterized and propagated into a WR-SWG to generate scenarios of future climate change over a region.

There are two primary strategies to identify WRs used to drive WR-SWGs. The first takes a bottom-up approach that classifies WRs as latent processes that are associated with regional patterns of surface weather. This approach—commonly employed via Hidden Markov Models (HMMs) or their non-homogeneous counterparts (NHMMs)—uses space-time patterns of precipitation or other meteorological variables across a local geographic unit of interest (e.g., a river basin) to classify the historical record into a small set of latent states (i.e., WRs) (Hughes and Guttorm, 1994; Katz and Parlange, 1995;

Hughes *et al.*, 1999; Sansom and Thomson, 2001; Robertson *et al.*, 2004; Holsclaw *et al.*, 2016; Kim and Lee, 2017). The second strategy takes a top-down approach in which WRs are identified directly through the classification of large-scale circulation fields (e.g., geopotential heights, integrated vapour transport) (Bardossy and Plate, 1991, 1992; Wilson *et al.*, 1992; Steinschneider *et al.*, 2019). This can be accomplished using a variety of approaches (Hannachi *et al.*, 2017), such as *K*-means clustering (Diday and Simon, 1976; Michelangeli *et al.*, 1995; Robertson *et al.*, 2015), self-organizing maps (Hewitson and Crane, 2002; Huth *et al.*, 2008), or HMMs applied to the large-scale atmospheric data (Majda *et al.*, 2006; Franzke *et al.*, 2008).

The bottom-up and the top-down approaches have unique strengths and weaknesses, especially for WR-SWG applications to climate vulnerability assessments. Bottom-up approaches have the benefit of identifying WRs tailored for the local weather of a geographic unit of interest. However, because the WRs are based on space-time patterns of surface meteorology over a relatively small spatial domain (very often a river basin), the diversity of WRs can be limited. For instance, most HMM applications to precipitation in a river basin identify a small set of WRs linked to dry conditions over the entire basin, wet conditions over the entire basin, and a few spatial gradients of precipitation across the basin (Robertson *et al.*, 2004; Betrò *et al.*, 2008; Khalil *et al.*, 2010; Verbist *et al.*, 2010; Holsclaw *et al.*, 2016; Pineda and Willems, 2016; Cioffi *et al.*, 2017; Guo *et al.*, 2018). In reality, many different large-scale atmospheric states could lead to any of those patterns of precipitation. Thus, the bottom-up approach often mixes all of those distinct atmospheric patterns into a single WR, hindering the development of scenarios where some of those atmospheric states are projected to become more likely and some less likely (Ghil and Robertson, 2002).

WRs identified through a top-down approach are likely less myopic, because they are inferred from atmospheric fields over a much larger spatial domain (e.g., the entire Pacific/North American sector). This raises the likelihood that these WRs would map better to those projected by GCMs (Haines and Hannachi, 1995; Muñoz *et al.*, 2017), supporting the generation of process-oriented future climate scenarios. However, the WRs in a top-down approach are often identified using methods designed to explain variability in the atmospheric fields, rather than variability in the local meteorology. Therefore, the WRs identified in a top-down approach may not be well suited for reproducing local weather statistics through a WR-SWG.

This study focuses on top-down identification of WRs specifically for WR-SWG applications. There is a growing

literature that has examined design components for WR identification with the specific purpose of explaining variability in local weather. For example, COST Action 733 (Harmonization and Applications of Weather Types Classifications for European Regions) was an initiative to establish a common reference for classifications of atmospheric circulation over Europe (Philipp *et al.*, 2016; Tveito and Huth, 2016). Studies within this initiative examined how the choice of space and time domain, atmospheric variables, number of regimes, and classification algorithm influenced the resulting WRs and how well they characterized surface temperature and precipitation distributions (Beck *et al.*, 2016; Huth *et al.*, 2016) and trends in those variables (Cahynová and Huth, 2016). However, to the authors' knowledge, this body of work has focused exclusively on identifying WRs that effectively separate the marginal distributions of precipitation and temperature. It remains unclear how the design parameters in WR identification impact the fidelity of WR-SWG simulations, which must reproduce a wide array of statistical characteristics of weather (e.g., moments, persistence, extremes, spatial coherence) across multiple temporal scales.

In addition, another body of work has explored the identification of WRs based on their interpretability, parsimony, and persistence. For instance, the concept of metastability has been introduced to characterize quasi-persistent atmospheric states consistent with extended regime-like structure (Majda *et al.*, 2006; Franzke *et al.*, 2008; Kwasniok, 2008). Here we distinguish metastable regimes (MRs) from WRs based on their characteristic timescales and transition probabilities. By definition, MRs are long-lived and exhibit relatively infrequent transitions between each other (Franzke *et al.*, 2008). These characteristics generally limit the number of MRs that can be identified in the atmosphere. WRs can be defined to have these characteristics (i.e., WRs can be MRs), but they can also be defined to be shorter-lived, more numerous, and exhibit preferential transition patterns (Huth *et al.*, 2016). Further, a larger set of WRs can be separable into a smaller number of MRs, where sets of WRs within a given MR exhibit “fast” (i.e., frequent) transitions, while sets of WRs across different MRs tend to exhibit “slow” (i.e., infrequent) transitions (Franzke *et al.*, 2008). For example, two WRs can be defined to represent different spatial perturbations or intensities of the negative phase of the Northern Hemisphere annular mode (NAM) that transition between each other often before

transitioning into other WRs reflective of neutral or positive phases of the NAM (Franzke *et al.*, 2008). These two WRs compose one MR that represents the negative NAM. It remains unclear whether a parsimonious selection of WRs that exhibit metastable behaviour is also capable of supporting high-quality simulations of local weather with a WR-SWG, or if more numerous and shorter-lived WRs are required to support stochastic simulation.

Based on the knowledge gaps identified above, this work addresses two main research questions regarding WR-SWG development:

1. How should atmospheric data be clustered in a top-down approach to identify WRs that are suitable for WR-SWGs?
2. How many WRs should be used for WR-SWGs applied at regional (state to national) scales?

To answer these questions, we utilize a previous WR-SWG tailored for the Western United States (Steinschneider *et al.*, 2019) and explore its ability to replicate a range of local weather characteristics in 13 river basins across the state of California. We assess WR-SWG performance in reproducing precipitation intensity, variability, persistence, and extremes using two top-down approaches for WR identification (HMMs vs. *K*-means clustering), and different numbers of WRs defined for the Pacific/North America sector. This analysis is supplemented by a metastability analysis for a range of WR classifications, which helps to ensure interpretability in the final WRs selected. These analyses, taken together, advance a new approach for designing WR-SWGs for application to large spatial scales, which may be preferable so that a single modelling framework can be employed to generate internally consistent future climate scenarios across state or national scales.

## 2 | DATA AND METHODS

### 2.1 | Climate data and case study basins in California

We collected daily time-series of gridded ( $2.5^\circ \times 2.5^\circ$ ) 500-hPa geopotential heights from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis (NCEP/NCAR Reanalysis 1) dataset (Kalnay *et al.*, 1996) between 1950 and 2017 (68 years) for the cold season (November–March). There are  $t = 10,134$  days in this timespan. We extracted these gridded data for the region between

180 and 100°W and 30–60°N to quantify patterns of large-scale atmospheric flow over the Pacific/North American sector that govern moisture transport and storm tracks impacting the Western United States (Milrad, 2018; Steinschneider *et al.*, 2019).

We also calculated daily vertically integrated water vapour transport (IVT,  $\text{kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1}$ ) from the same daily NCEP/NCAR Reanalysis 1 database using specific

humidity ( $q$ ,  $\text{kg}\cdot\text{kg}^{-1}$ ), zonal and meridional wind speeds ( $u$  and  $v$ ,  $\text{m}\cdot\text{s}^{-1}$ ), and pressure levels ( $p$ ) between 1,000 and 300-hPa:

$$\begin{aligned} \text{IVT} &= \frac{1}{g} \int_{1,000}^{300} q \\ \cdot \mathbf{V} dp &= \frac{1}{g} \sqrt{\left( \int_{1,000}^{300} q \cdot u dp \right)^2 + \left( \int_{1,000}^{300} q \cdot v dp \right)^2} \end{aligned} \quad (1)$$

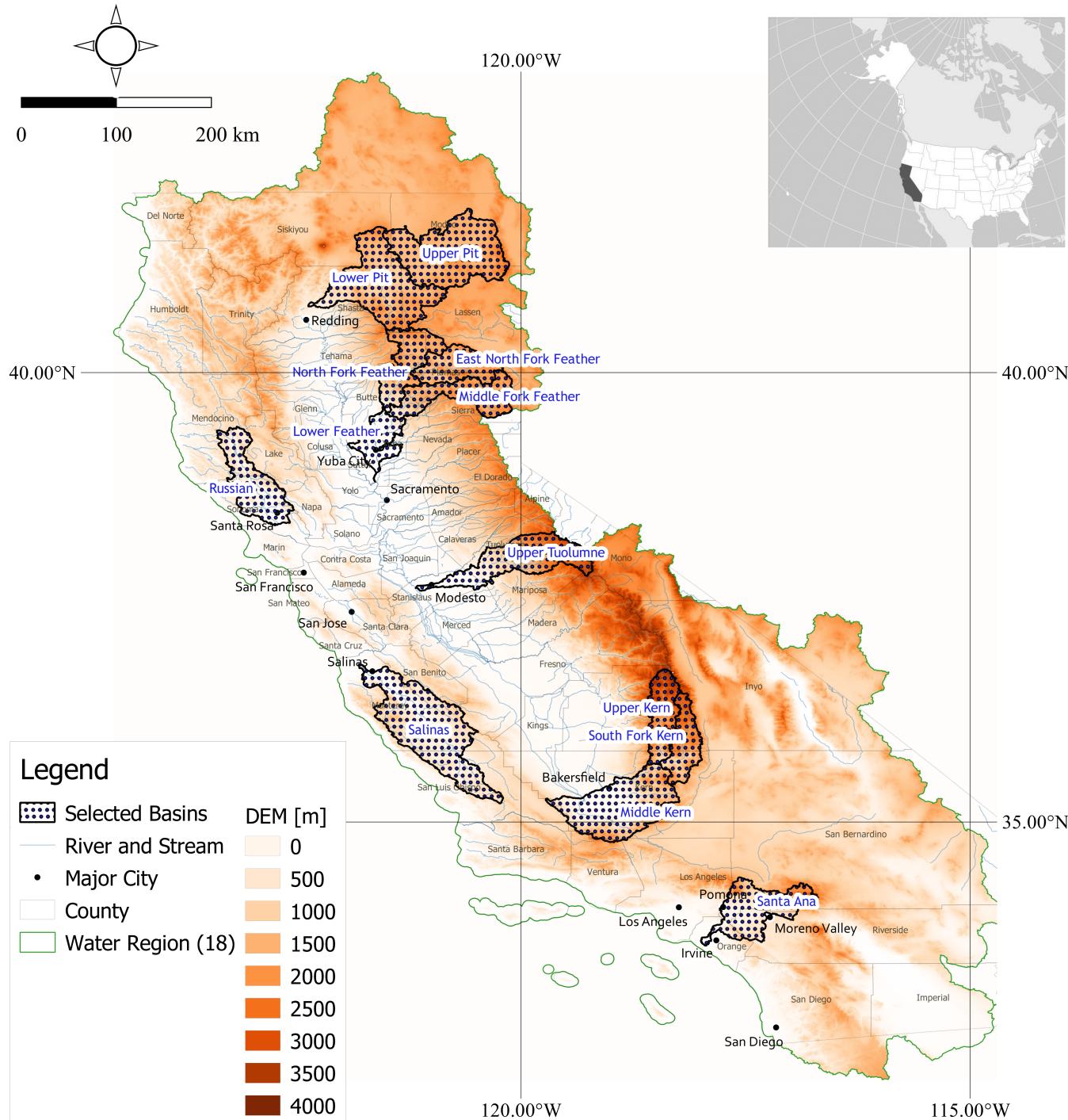


FIGURE 1 Map of the selected 13 river basins in California [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Here,  $\mathbf{V}$  is the wind vector [ $\text{m}\cdot\text{s}^{-1}$ ],  $dp$  refers to the pressure difference between two adjacent vertical levels, and  $g$  indicates the acceleration induced by gravity ( $9.81 \text{ [m}\cdot\text{s}^{-2}]$ ).

For regional hydrometeorological measurements, we accessed observed daily precipitation [mm] time-series during 1950–2013 (64 years) from the gridded meteorological dataset for North America (south of  $53^\circ\text{N}$ ) provided by Livneh *et al.* (2015) at a  $0.0625^\circ \times 0.0625^\circ$  ( $\sim 6$ -by-6 km) resolution. This gridded dataset applies the SYMAP algorithm to station data in the gridding process, in addition to an orographic scaling based on the PRISM algorithm (see more in Shepard, 1984 and Daly *et al.*, 1994). Daily temperature [C] data were taken from the same dataset.

We selected 13 river basins (at 8-digit Hydrologic Unit Code: HUC8 level) that together constitute seven major river systems across California (Figure 1). These basins have drainage areas ranging from 2,000 to over 8,000  $\text{km}^2$ . They are also distributed over different regions of California (northern, central, southern, and coastal areas) with varying climatic conditions and precipitation regimes. Gridded precipitation observations were extracted within and along the boundary for these 13 basins during the cold season ( $t = 9,529$  days). Table 1 provides detailed information about these basins and river systems, including hydrologic characteristics and dimensions of the compiled gridded data points.

## 2.2 | Stochastic weather generator

We use a recently developed WR-SWG (Steinschneider *et al.*, 2019) to investigate the proposed research questions discussed in Section 1. The model is a semiparametric, multivariate, and multisite weather generator that was previously tested in the Upper Tuolumne River Basin (basin #11 in Table 1). The model conditions simulations of local weather variables (precipitation, temperature) on simulated WRs. The weather generator features a hierarchical structure with three core modules including (1) the identification of historic WRs and simulation of new WRs using a Markov Chain; (2) simulation of local weather conditioned on the WRs using a block bootstrapping technique, followed by the addition of noise to perturb simulated data beyond their historic range; and (3) perturbations to the simulation schemes in (1) and (2) that impose different scenarios of dynamical and thermodynamic climate change. The block bootstrap and WR simulation scheme help to preserve complex space-time structure in resulting weather over multiple river basins across a broad geographic area, while the perturbation scheme is designed to embed different, process-oriented hypotheses of climate change into the resulting simulations. We refer the reader to Steinschneider *et al.* (2019) for additional detail on the model structure. In addition, we note that we implemented two minor changes to the block bootstrapping technique in module 2 to improve the properties of the resampled weather, which are detailed in the Supplementary Materials (see Text S1).

**TABLE 1** List of river systems and river basins selected for this study

Basin number	River system number	Drainage area [ $\text{km}^2$ ]	States	HUC8	Basin Name	Number of gridded locations
1	1	3,846	CA	18010110	Russian	102
2	2	6,944	CA	18020002	Upper Pit	192
3	2	6,832	CA	18020003	Lower Pit	183
4	3	3,136	CA	18020121	North Fork Feather	85
5	3	2,663	CA	18020122	East Branch North Fork Feather	73
6	3	3,535	CA	18020123	Middle Fork Feather	95
7	3	2,005	CA	18020159	Honcut Headwaters-Lower Feather	56
8	4	2,828	CA	18030001	Upper Kern	72
9	4	2,541	CA	18030002	South Fork Kern	66
10	4	6,778	CA	18030003	Middle Kern-Upper Tehachapi-Grapevine	168
11	5	4,851	CA	18040009	Upper Tuolumne	123
12	6	8,622	CA	18060005	Salinas	225
13	7	4,387	CA	18070203	Santa Ana	114

The identification and simulation of the WRs (module 1) play a critical role in the performance of the weather generator, and it is also the module through which dynamic climate changes (i.e., perturbations to future atmospheric circulation) are embedded in the model simulations. This work tests two design components for WR identification (Section 2.3) and assesses their impact on WR-SWG fidelity using a variety of performance measures (Section 2.4). We then explore an approach to further diagnose the structure, temporal dynamics, and metastability of the identified regimes (Section 2.5).

## 2.3 | Design components for weather regime identification

Akin to the experiments in Huth *et al.* (2016), we explore the impact of design components for WR identification on an end-point application, which unique to this study is WR-SWG simulation performance. First, we compare two methods to define WRs in the Pacific/North American sector. The baseline approach used in Steinschneider *et al.* (2019) is based on a  $K$ -means clustering algorithm (Diday and Simon, 1976). In that approach, 500-hPa geopotential height anomalies are projected onto their first  $J$  empirical orthogonal functions (EOFs), where  $J$  is chosen using a screen test to ensure that the selected EOFs explain the majority (e.g., > 90%) of the variance in the data.  $K$ -means clustering is then conducted on the first  $J$  principal components (PCs) to partition each day in the record into one of  $K$  separate WRs.

While  $K$ -means clustering is a commonly used and computationally inexpensive approach, but it does not consider WR persistence when partitioning days into individual WRs (Hannachi *et al.*, 2017). Therefore, we consider an HMM as an alternative approach to identify WRs. HMMs are nonlinear statistical models that use latent variables to identify clusters in state-space while simultaneously accounting for the distribution and temporal dynamics of observed data. In this application, we evaluate a first-order homogenous HMM on the same  $J$  PCs of geopotential height discussed above to assign each day in the record to one of  $K$  separate WRs. Importantly, days are classified into WRs in a way that explicitly considers WR persistence, which we hypothesize will lead to more persistent weather (i.e., longer dry and wet spells) in the WR-SWG simulations. We present a brief description of the HMM below, and refer readers to standard references for additional details (e.g., Rabiner, 1989; MacDonald and Zucchini, 1997; Ghahramani, 2001; Zucchini *et al.*, 2017).

We adopt similar notation to that in Kwasniok (2014) and let  $r_t$  take integer values from 1 to  $K$ , representing

the presence of one of  $K$  latent WRs on day  $t$ . The vector  $x_t$  is equal to the values of the  $J$  PCs of geopotential height anomalies on day  $t$ . The probability of observing  $x_t$  conditional on being in regime  $r_t = i$  follows a multivariate Gaussian distribution and is given by:

$$P(x_t|r_t=i) \sim \text{MVN}(\mu_i, \Sigma_i) \quad (2)$$

where  $\mu_i$  is a  $J$ -vector of means for the  $i$ th regime and,  $\Sigma_i$  is a  $J \times J$  diagonal matrix of variances. The sequence of regimes  $r_{1,\dots,N}$  is assumed to follow a Markovian process with a time-invariant,  $K$ -by- $K$  transition probability matrix (t.p.m.)  $A$  defined with elements:

$$A_{ij} = P(r_{t+1}=j|r_t=i) \quad (3)$$

$A$  is a row-stochastic matrix that verifies  $\sum_j A_{ij} = 1$ , and  $A_{ij} \geq 0$ .

The Expectation–Maximization algorithm (Dempster *et al.*, 1977) is used to estimate the parameters of the HMM (i.e.,  $\mu_i$  and  $\Sigma_i$  for each of the  $K$  states,  $P(r_1 = i)$  for initialization, and matrix  $A$ ). The expectation step is computed using the forward-backward algorithm (Baum and Petrie, 1966; Baum *et al.*, 1970), and the expectation–maximization steps are iterated until convergence. The most probable sequence of hidden states is computed using the Viterbi algorithm (Rabiner, 1989). In this work, we fit the HMM using the dependent mixture model frameworks R-package “depmixS4” (Visser and Speekenbrink, 2010).

In addition to the methods used to define WRs, we also test WR-SWG performance for different numbers of WRs across the two methods. Specifically, we develop WR-SWG simulations for both HMMs and  $K$ -means clustering using  $K = 2, 3, 4, 5, \dots, 12$  number of states. We initially tested additional numbers of states but found increasing declines in performance and therefore limited our final assessment to  $K \leq 12$ .

## 2.4 | Performance metrics

We assess WR-SWG simulation performance for daily precipitation across all 13 river basins and combinations of design components discussed in Section 2.3. For each combination of design components, we simulate an  $N = 200$  ensemble of 64-year simulations of daily precipitation for each basin. We consider 14 widely used statistical performance metrics, including: the mean of daily precipitation; standard deviation of cold-season total precipitation (to measure inter-annual variability); probability of dry and wet days; average and maximum duration of dry and wet spells; the cold-season daily precipitation

maximum; and the cold-season maximum of 3, 7, 10, 14, and 30-day moving averages of precipitation rates. We compute these metrics for all grid locations within a basin for both the observed data and each member of the simulated ensemble.

We summarize the discrepancy between observed and simulated values at the basin scale for each of the 14 metrics using two evaluation criteria (Willmott, 1981; Gupta *et al.*, 1998):

$$\text{Percent bias}[\%] = \frac{\tilde{P}_{s,i} - P_{o,i}}{P_{o,i}} \times 100\% \quad (4)$$

$$\text{Index of agreement}[-] = 1 - \frac{\sum_{i=1}^m (\tilde{P}_{s,i} - P_{o,i})^2}{\sum_{i=1}^m (|\tilde{P}_{s,i} - P_{\bar{o}}| + |P_{o,i} - P_{\bar{o}}|)^2} \quad (5)$$

Here, the percent bias for a given performance is a metric computed through taking their median across the  $m$  gridded locations within a basin.  $\tilde{P}_{s,i}$  indicates the median of that metric across the 200 ensemble members for site  $i$ , and  $P_{o,i}$  refers to the observed metric value for site  $i$ . In Equation (4), the tilde accent over the entire quotient indicates the median across sites. The percent bias criterion specifies systematic deviations between observed and simulated metrics across a basin. However, it does not provide information on random deviations within a basin (i.e., simulated metrics are overestimated for some grid cells but underestimated for others). To quantify this measure of skill, we use the index of agreement between the simulated and observed metrics (Equation 5). Here,  $P_{\bar{o}}$  is the average of all observed metric values across the  $m$  gridded locations. The index of agreement varies between 0 (for a complete lack of agreement) and 1 (for a full agreement). It provides information on both the bias and variance of simulated metric values across a basin. Note that in rare circumstances index of agreement may be indeterminate due to division by zero (Zambrano-Bigiarini, 2019), in which case the metric is assumed missing.

## 2.5 | Metastability of weather regimes

The analyses in Sections 2.3 and 2.4 provide insight on the number of WRs that lead to the best WR-SWG simulations across California. We juxtapose this number against the number of MRs across the Pacific/North American sector to determine whether a parsimonious

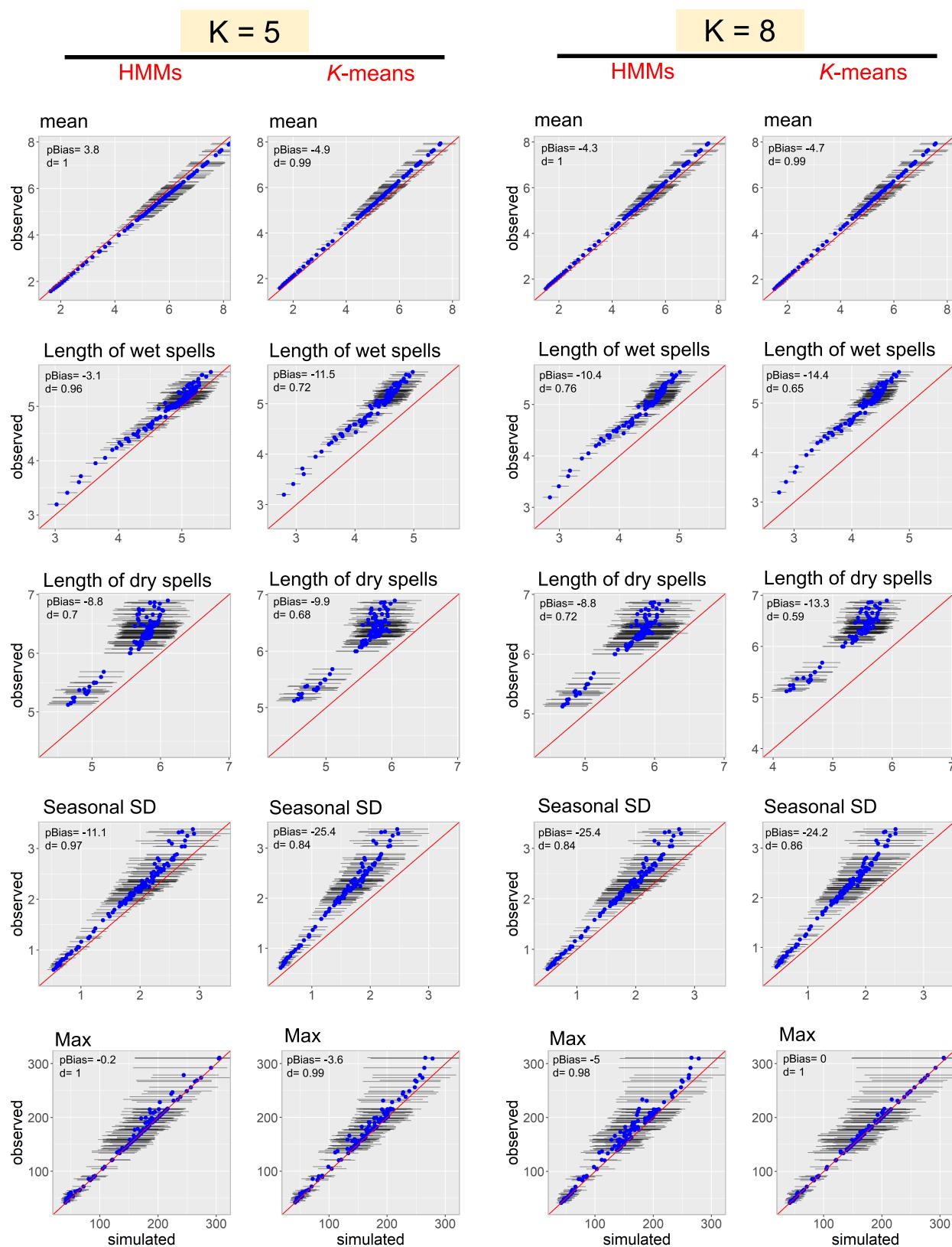
classification of regimes is consistent with improved performance of the WR-SWG.

We follow the methods of Franzke *et al.* (2008) to define MRs based on WRs identified using an HMM (as in Section 2.3). For some number  $K$  of WRs identified by the HMM, MRs can be identified using an eigendecomposition of the t.p.m. The first eigenvalue ( $\lambda_1$ ) will always equal unity and is associated with an eigenvector that represents the stationary distribution of the Markov chain. The other eigenvalues will satisfy  $|\lambda_k| \leq 1$ , and those eigenvalues close to 1 are associated with eigenvectors (i.e., patterns of WR transition probabilities) that decay slowly. That is, the eigenvectors with  $|\lambda_k| \approx 1$  identify sets of WRs that have a high likelihood of transitioning between each other and a low probability of transitioning to WRs outside of that set. Each such set of WRs defines an MR.

The number of MRs can be defined by the number of eigenvalues (after the first) close to 1 (Huisings *et al.*, 2003). However, it is difficult to determine whether an eigenvalue is sufficiently close to 1 to declare a significant MR. Therefore, we follow Franzke *et al.* (2008) and examine the ratio of log-eigenvalues within the eigenvalue spectrum:

$$Fr_k = \log(\lambda_k) / \log(\lambda_{k+1}), k \geq 2 \quad (6)$$

Metastability requires a significant gap in the eigenvalues spectrum, which can be measured effectively when this fraction falls significantly below 1. The statistical significance of  $Fr_k$  values can be determined by comparing them to  $Fr_k$  values based on the eigenvalue spectrum of a reference model, taken here to be an autoregressive model with no metastable behaviour. That is, we fit a multivariate autoregressive model of order 1 (mAR(1)) to the  $J$  PCs of geopotential height on which the HMM is based. New PCs are simulated with this mAR(1), and an HMM is fit to these simulated PCs. Importantly, the linearity of the mAR(1) ensures no regime-like behaviour in the resulting HMM (i.e., no metastable regimes).  $Fr_k$  values are calculated from the t.p.m. of the resulting HMM. This process is repeated 100 times, creating an ensemble of  $Fr_k$  values from the reference model. If  $Fr_k$  from the HMM fit to the original PCs falls below some threshold of the ensemble of  $Fr_k$  values from the reference model (taken as the median in this study), then this indicates significant metastable behaviour. Ultimately, we choose the number of MRs to be the index where  $Fr_k$  both reaches a minimum and is below the reference value. To ensure robustness of the results, the number of MRs should be similar regardless of the number of WRs on which the MRs are based.

**FIGURE 2** Legend on next page.

Once MRs are identified, we can assign membership of the WRs to their corresponding MRs. We employ Generalized Perron Cluster Cluster Analysis (G-PCCA) (Reuter *et al.*, 2018) to conduct this clustering. The G-PCCA is an improved version of robust Perron cluster analysis that has been widely used in molecular studies in chemistry and biology (e.g., Deuflhard *et al.*, 2000; Deuflhard and Weber, 2005; Weber and Kube, 2005; Röblitz and Weber, 2013; Fackeldey *et al.*, 2018; Reuter *et al.*, 2018). It performs a fuzzy clustering conditional on an optimality criterion (Weber and Fackeldey, 2015), and can be applied to any HMM that exhibits periodic transitions between sets of states (e.g., Weber, 2003; Reuter *et al.*, 2018). In our application, we first use the eigenspectrum test above to determine the number of MRs that exist in the data. With this number pre-determined, we then use G-PCCA to estimate the probability that a given WR belongs to a specific MR. We refer the reader to Reuter *et al.* (2018) for additional technical details of the G-PCCA method.

### 3 | RESULTS

#### 3.1 | HMM vs. K-means for weather regime identification

We first examine WR-SWG performance based on WRs identified using HMMs and *K*-means clustering. In these first results, we focus on a specific number of WRs (five and eight) and a specific basin (the Upper Tuolumne River Basin; see Figure 1 and Table 1) to more clearly assess skill across locations within a single basin. Figure 2 shows a set of selected performance metrics for observed and simulated daily precipitation at each grid cell within the Upper Tuolumne across 200 ensemble members. We also indicate the percent bias (Equation 4) and the index of agreement (Equation 5) across all 200 ensemble members. Overall, performance is similar for both numbers of states and methods used to identify the WRs. The mean and maximum precipitation are well simulated, while spell lengths and seasonal variability of total precipitation are underestimated. These results are consistent across all grid cells within the basin, and for

some metrics (length of dry spells), the observed statistics are almost never within the variability of the simulated ensemble. However, some differences between the methods do emerge. Wet spells and seasonal variance are better simulated using HMMs. The downward bias in these metrics is almost eliminated using the HMM with five states. This is consistent with our hypothesis that the HMM accounts better for persistence in WRs, which propagates into more persistence (and therefore seasonal variance) in simulated weather. The use of HMMs also appears to provide some benefits over *K*-means when using eight states, but these improvements are marginal. In addition, for certain metrics (precipitation maximum), HMMs provide better performance for five states, while *K*-means performs better with eight states.

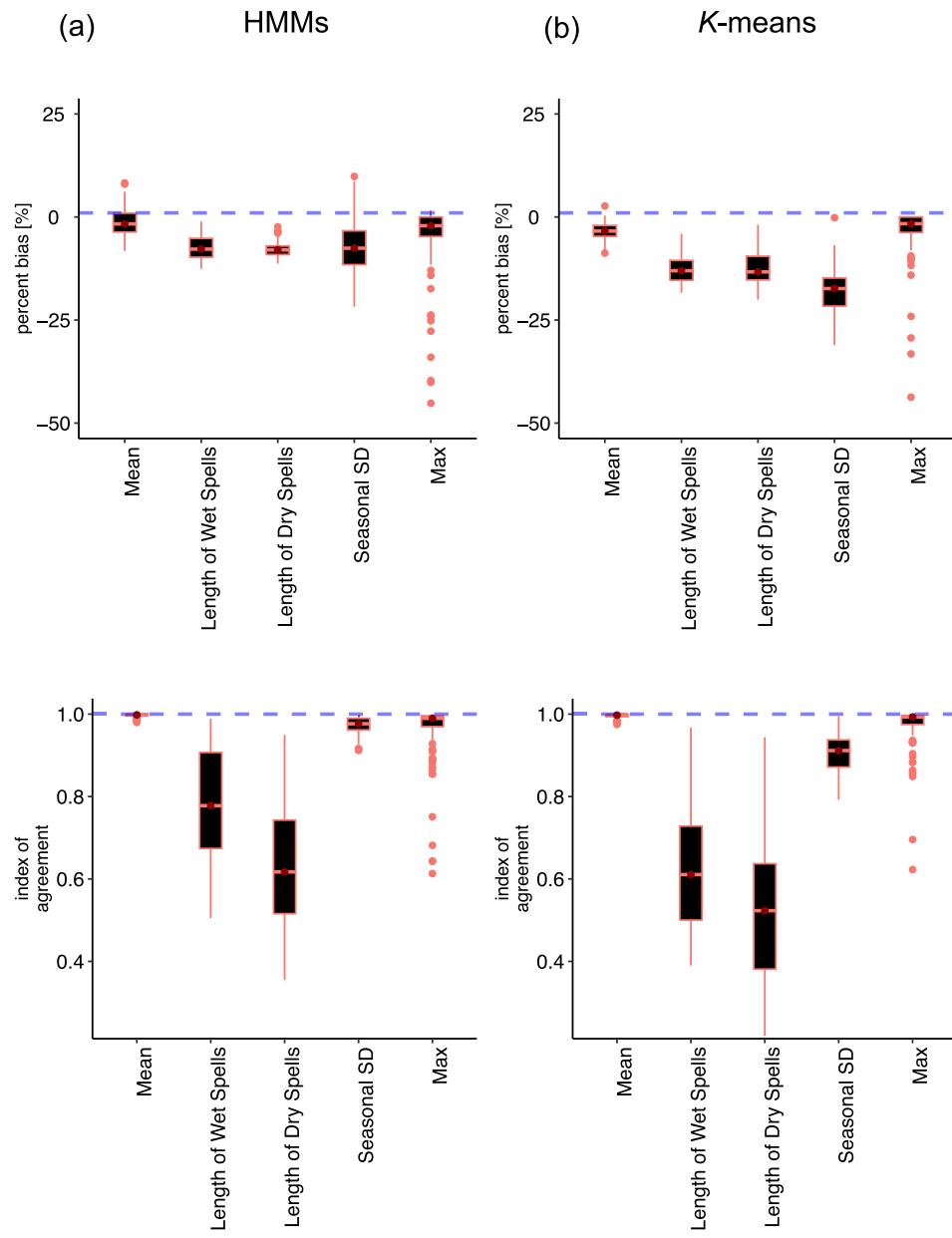
Figure 3 compares the same performance metrics between HMMs and *K*-means, but summarized across all river basins and for all numbers of states. Here, we show the distribution of percent bias and index of agreement pooled across all experiments (13 basins and 2–12 number of states). These comparisons provide a broad overview of performance differences that emerge when using HMM or *K*-means approaches. In general, the HMM approach outperforms *K*-means in all metrics besides maximum precipitation. In particular, HMMs significantly improve the length of wet and dry spells and seasonal standard deviation. These improvements include a reduction in the negative bias, but also a decline in the index of agreement, suggesting reduced variance in these metrics across grid cells within each basin. For mean precipitation, the two methods are similar, but *K*-means does lead to a slightly more prominent negative bias. Given these results, we hereafter focus on the WR-SWG performance base on WRs identified using HMMs, and explore how this performance varies with the number of WRs selected.

#### 3.2 | Number of weather regimes

Figures 4 and 5 present the distribution of percent bias for all 14 performance metrics across all 13 river basins, but separately for *K* = 2 through *K* = 12 numbers of states. Figure 4 shows these results for precipitation

**FIGURE 2** Selected performance metrics for observed and WR-SWG simulated precipitation in the upper Tuolumne River basin based on (*K* = 5 and 8) WRs identified using HMMs and *K*-means clustering. Each simulated point represents the median performance metric across 200 ensemble members at a gridded location within the basin. The thin horizontal lines (whiskers) denote the 95% range for simulated metrics across the ensemble. The *percent bias* [pBias, %] and *index of agreement* [*d*, –] are two evaluation criteria to determine the performance of WR-SWG across the basin (*pBias* = 0 and *d* = 1 indicate a perfect agreement between the observed and simulated precipitation for that metric at the basin scale; note that these metrics are rounded to 1 and 2 significant digits, respectively) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

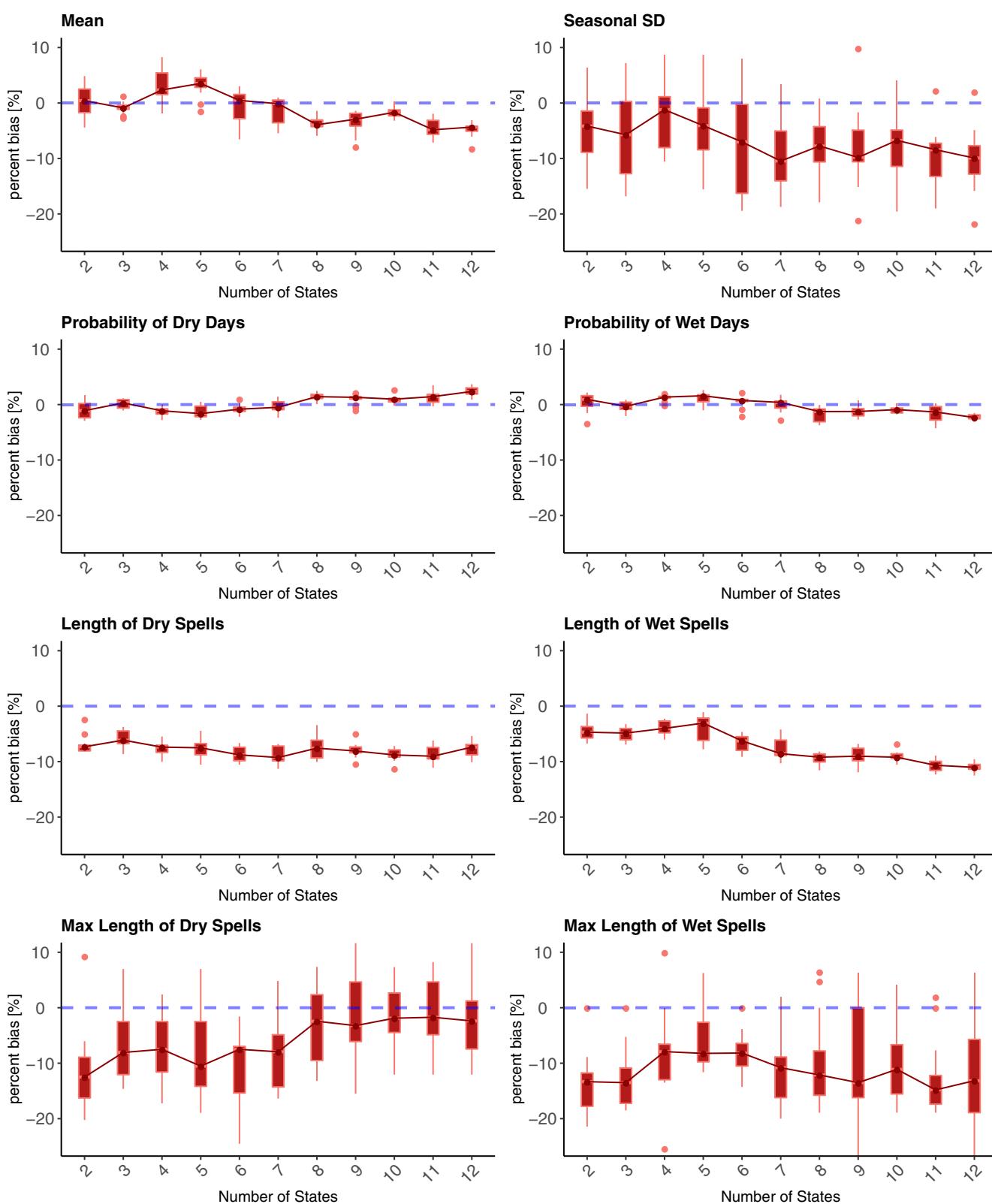
**FIGURE 3** Comparison of hidden Markov models (HMMs) and K-means algorithms used to identify weather regimes (WRs) employed in the WR-based stochastic weather generator (SWG). Percent bias and index of agreement determine the level of inconsistencies between the simulated and observed precipitation statistics from the entire set of 2–12 WRs—identified by HMMs (a) and K-means (b)—and across all 13 river basins in California. Selected diagnostics for precipitation characteristics include the mean, length of wet and dry spells, seasonal standard deviation (SD), and maximum of daily precipitation (max). The box represents the interquartile range (IQR, or middle 50%), with its two whiskers extending to 1.5 times the IQR from the box edges, and red-coloured points (if any) are located beyond such extents. The blue dashed-line denotes where there is no deviation between the simulated and observed statistics [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



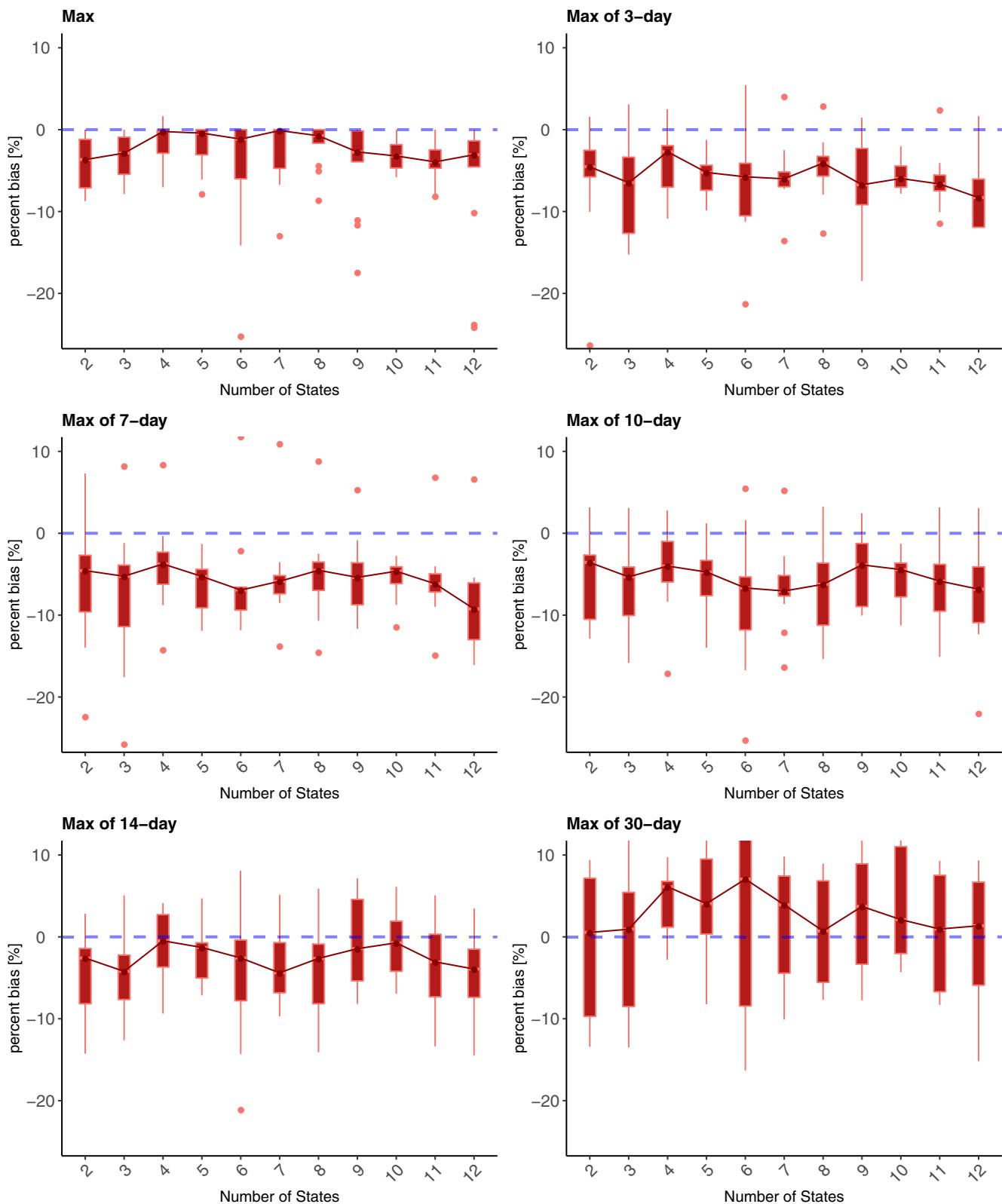
moments and the probability and spell lengths of dry and wet days, while Figure 5 focuses on extreme precipitation over different window lengths. In general, percent bias fluctuates between –15 and 10% for all metrics, with a tendency towards negative biases. However, results differ significantly by metric, and performance varies with the number of states more so for some metrics compared to others. For instance, mean precipitation tends to be unbiased for a small number of states (2–3), exhibits a slight positive bias for 4–5 states, becomes unbiased again at 6–7 states, and then exhibits a negative bias for 8 or more states. For the seasonal standard deviation of precipitation, biases are smallest with four states, but become more negative if fewer or more states are used. The probability of dry and wet days generally exhibits a percent

bias equal to zero (no systematic deviation) regardless of the number of states used.

Similarly, the average length of dry spells exhibits a consistent negative bias of approximately 8–9% regardless of the number of states. However, the average length of wet spells exhibits a negative bias of around 5% for 2–5 states that grows towards a –10% bias for 6 or more states. Lastly, the maximum length of dry spells shows a larger negative percent bias for a small number of states (2–7), but it improves when a greater number of states are utilized. In contrast, percent bias for the maximum length of wet spells is at its absolute minimum for 4–5 states and the negative bias increases for smaller or larger number of states, similar to the behaviour of the seasonal standard deviation.



**FIGURE 4** The distribution of percent bias across 13 California river basins for different performance metrics, including mean daily precipitation, seasonal standard deviation (SD) of total precipitation, the probability of dry and wet days, and the average and maximum length of dry and wet spells. Each boxplot represents the distribution of percent bias when using a different number (2–12) of WRs identified using an HMM. The box represents the interquartile range (IQR), with whiskers extending to 1.5 times the IQR from the box edges, and light red points indicating outliers located beyond these extents. The blue dashed-line denotes where percent bias = 0 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



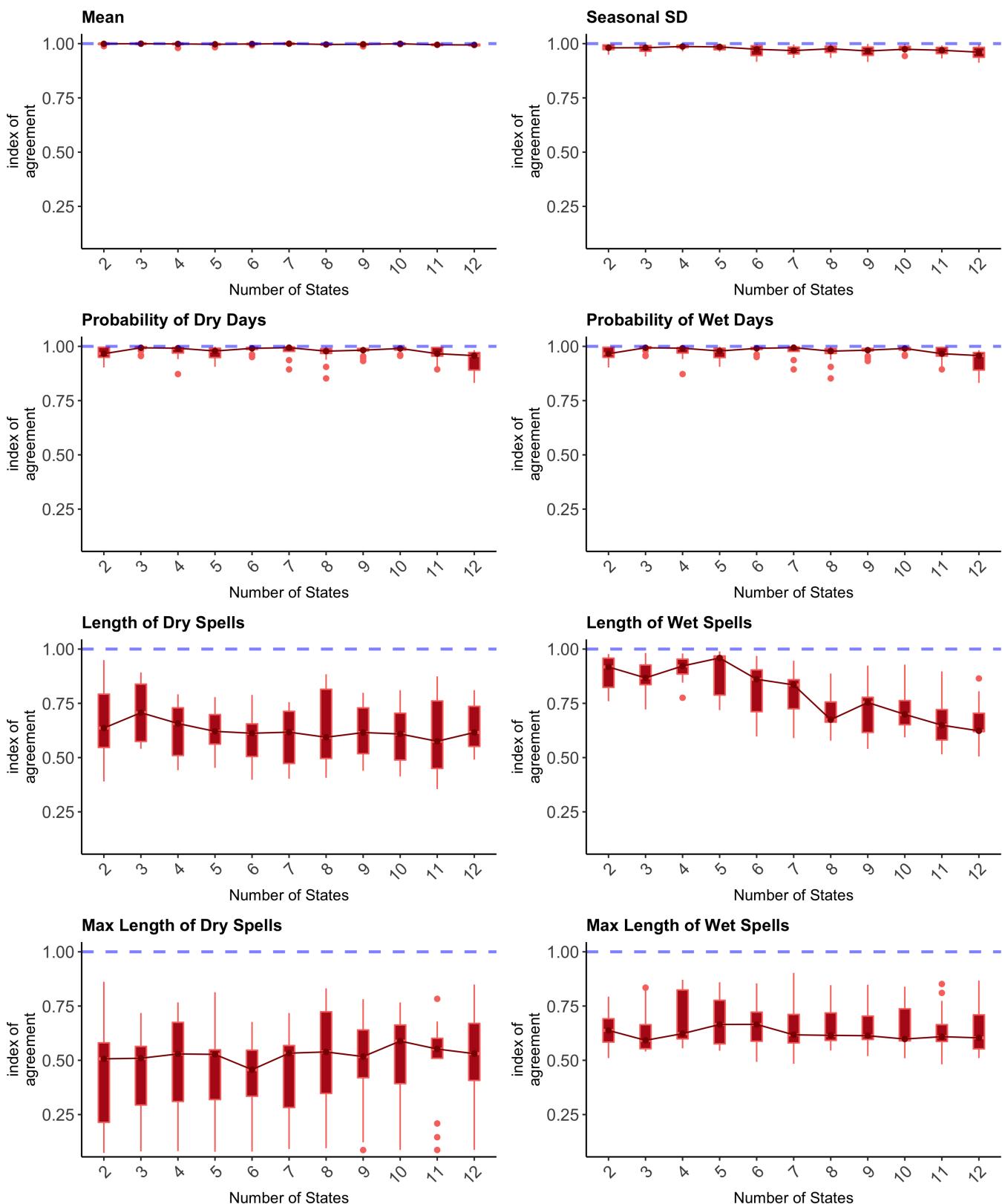
**FIGURE 5** Same as Figure 4 but for maximum precipitation over 1-, 3-, 7-, 10-, 14-, and 30-day moving averages [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

When examining bias in precipitation extremes over different windows (seasonal maxima of 1-, 3-, 7-, 10-, 14-,

and 30-day moving averages), biases tend to be negative for shorter windows but become positive for the longest

30-day window (Figure 5). In addition, the spread of percent bias increases as the window length grows. For the

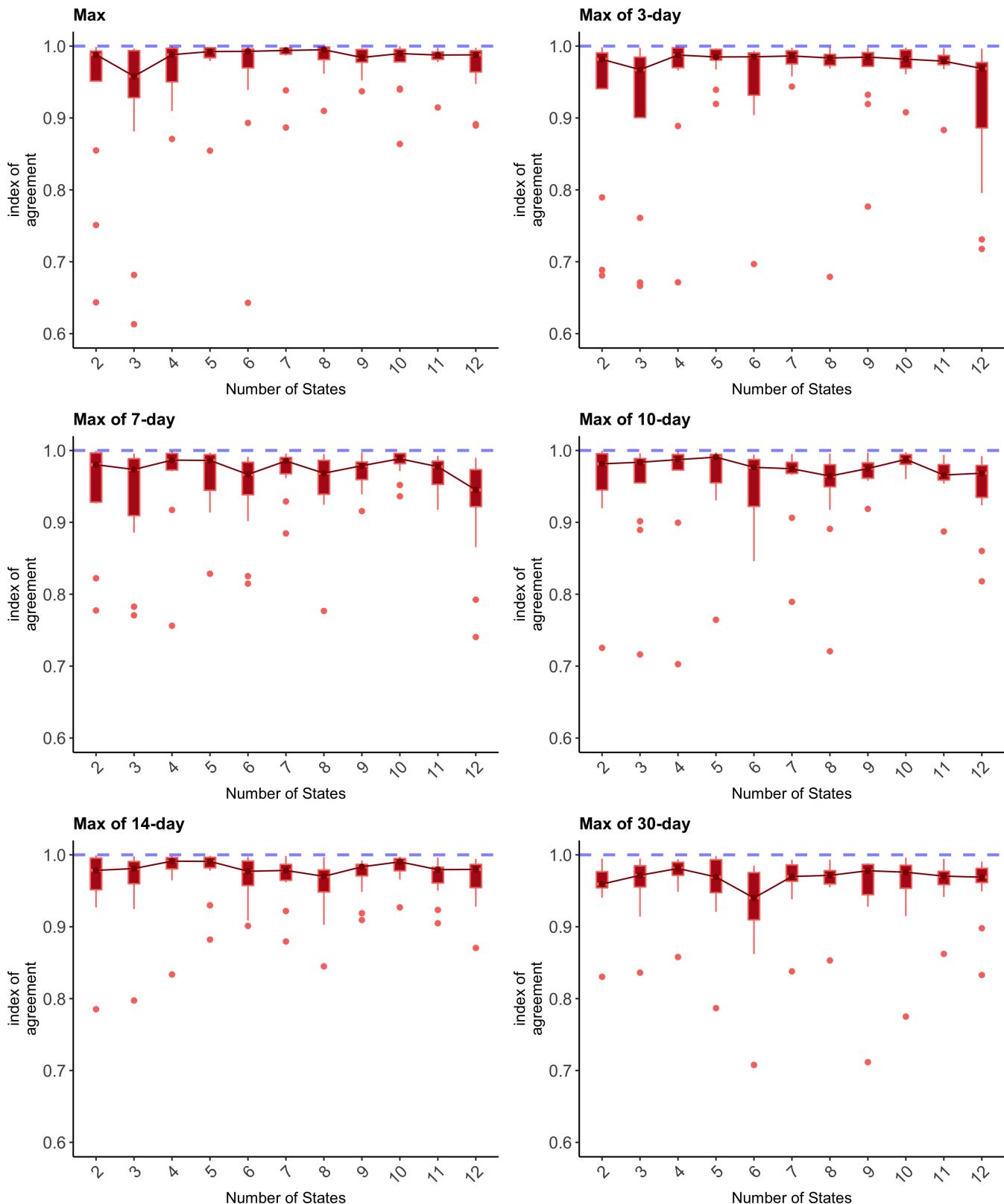
1-day maximum, 4–8 WRs generally lead to small negative biases, although some basins can exhibit significant



**FIGURE 6** Same as Figure 4, but for the index of agreement. The blue dashed-line denotes where index of agreement = 1 (i.e., full agreement) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

negative biases (e.g., for six WRs). For 3-, 7-, and 10-day windows, the bias is relatively constant across the

number of states, although it tends to be smallest (<5%) at  $K = 4$  states for each of them. When the window size



**FIGURE 7** Same as Figure 5, but for the index of agreement. The blue dashed-line denotes where index of agreement = 1 (i.e., full agreement) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

grows to 14 days, biases become small for all numbers of states, and for certain values of  $K$ , positive biases begin to emerge. For an even longer window of 14 days, biases are generally positive (on average around 5%) and become more variable. In general, these results highlight the complexity of simulating both short and more prolonged episodes of heavy precipitation across multiple basins. Biases in these simulated extreme events tend to be minimized when using  $K = 4$  states.

Figures 6 and 7 show similar results to Figures 4 and 5, but for the index of agreement. Recall that this criterion quantifies both bias and variance in WR-SWG performance across grid cells within a basin (Equation 5). Values of index of agreement are equal or close to 1 for the mean, seasonal standard deviation, and the probability of dry and wet days for all numbers of WRs tested (Figure 6). Together with Figures 4–5, this suggests that the SWG exhibits very little variance in these metrics across grid cells, and that most of the deviation is due to bias. In contrast, values of index of agreement for the average and maximum length of dry and wet spells are significantly below 1. For the average length of dry spells, these values range between 0.5 and 0.75, and do not vary drastically with the number of states. For the average length of wet spells, values of index of agreement are close to 1 for 4 to 5 states, but fall considerably for 6 or more number of states.

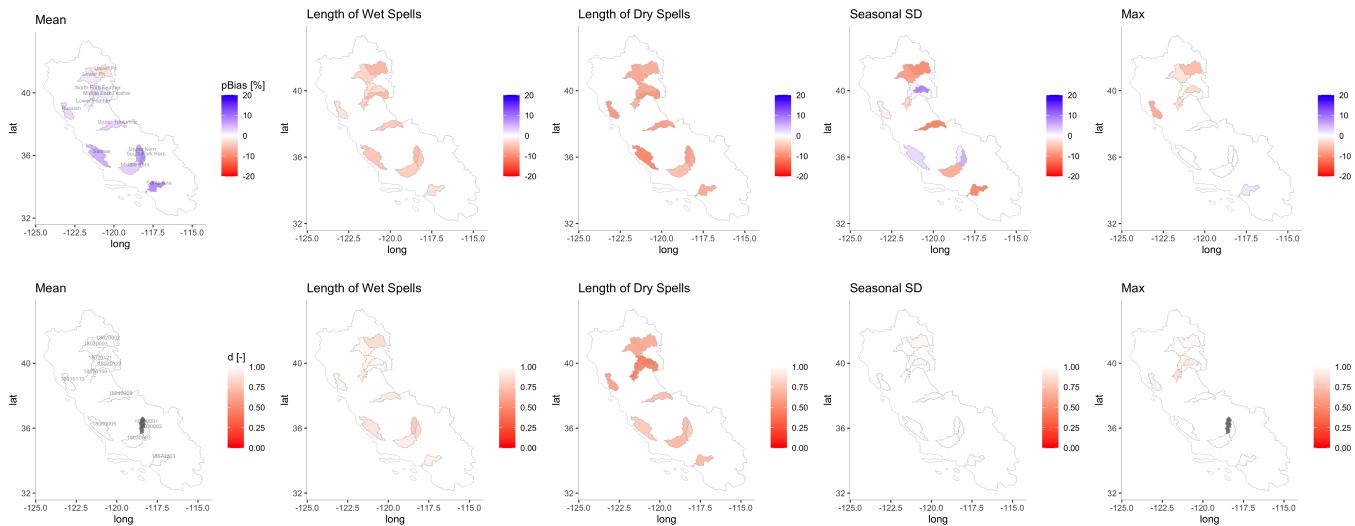
In general, the weather generator produces better simulations of average wet spells if a smaller number of WRs is selected. However, the maximum length of wet spells in Figure 6 shows little dependence on the number of states. Interestingly, values of index of agreement fall significantly below 1 (between 0.25 and 0.5) for the maximum length of dry spells. This contrasts with the bias, which is close to 0 for this metric when using higher numbers of states (Figure 4). These results indicate that for maximum dry spell statistics at a high number of states, there is significant variability in skill across grid cells within each basin (i.e., some overestimations, some underestimations) even if these biases are not systematic across the basins.

For extreme precipitation in Figure 7, the SWG generally scores index of agreement values between 0.9 and 1 for all window lengths, suggesting relatively high agreement between simulated and observed extremes across gridded locations in all basins. However, some outliers fall below 0.7, indicating only fair agreement in certain basins. The index of agreement is relatively similar regardless of the window size, and in general, at its maximum when using 4–5 states. However, the difference in index of agreement values across the different number of states is relatively small, similar to the results for percent bias in Figure 5.

Overall, the results in Figures 4–7 suggest that the use of four WRs in an HMM leads to the best WR-SWG simulations across a comprehensive set of performance metrics for precipitation, with absolute biases that are mostly below 10% and often below 5%, and with minimum variance across grid cells (i.e., high values of index of agreement) for most performance metrics. This is especially the case for the seasonal standard deviation of total precipitation and the length of wet spells. When using more than five WRs, significant bias and variance are introduced into these metrics, as well as for mean and 1-day maximum precipitation events. We note that these conclusions are corroborated if results are examined separately by basin, as is shown in the Supplementary Materials (Figures S1–S4).

In Figure 8, we present the spatial variation of SWG performance across the 13 basins in California when using four WRs. We show the values for percent bias and index of agreement for the five precipitation diagnostics featured in Figure 3 (i.e., mean, length of wet and dry spells, seasonal standard deviation, and daily maximum), and adjust the colour scale to highlight spatial patterns. Overall, spatial patterns in performance are limited, albeit with a few exceptions. For instance, mean precipitation is more positively biased in the south, while daily maximums exhibit a slight negative bias in the north. However, most other metrics generally exhibit consistent percent bias and index of agreement values across all basins, or exhibit seemingly random spatial variations across the state (e.g., seasonal standard deviation).

We also explored the cold-season regional climate conditioned on these four WRs across the Western United States. Figure 9 shows composites of 500-hPa geopotential height and IVT anomalies for these four WRs, together with locally standardized precipitation and temperature anomalies. The average frequency of each WR across the cold season is also provided in Figure S5. WR1 exhibits a ridge over Alaska and a southwest-to-northeast oriented trough over the Central Pacific and U.S. Pacific Northwest, directing substantial IVT towards the Western United States (Figure 9a). Accordingly, much of the region is anomalously wet and northern portions of the domain are colder than average (Figure 9b,c). WR2 exhibits a ridge over the Central/Eastern Pacific Ocean and a trough over Western North America that together direct moisture into British Columbia and the eastern part of the U.S. Southwest, imposing below average precipitation in California and very cold anomalies across much of the West. In WR3, a deep Aleutian low and ridge to the east yields southwesterly flow, above average precipitation over the U.S. Pacific Northwest, and positive temperature anomalies across most of the region. We note that WR3 occurs less frequently than other WRs,



**FIGURE 8** Percent bias [%] and index of agreement [-] for selected diagnostics of precipitation characteristics including the mean, length of wet and dry spells, seasonal standard deviation (SD), and daily maximum of precipitation across 13 river basins in California. These results are based on four WRs identified using an HMM. There is a complete agreement between the observed and simulated precipitation characteristics when percent bias = 0, and index of agreement = 1 (i.e., white shading) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

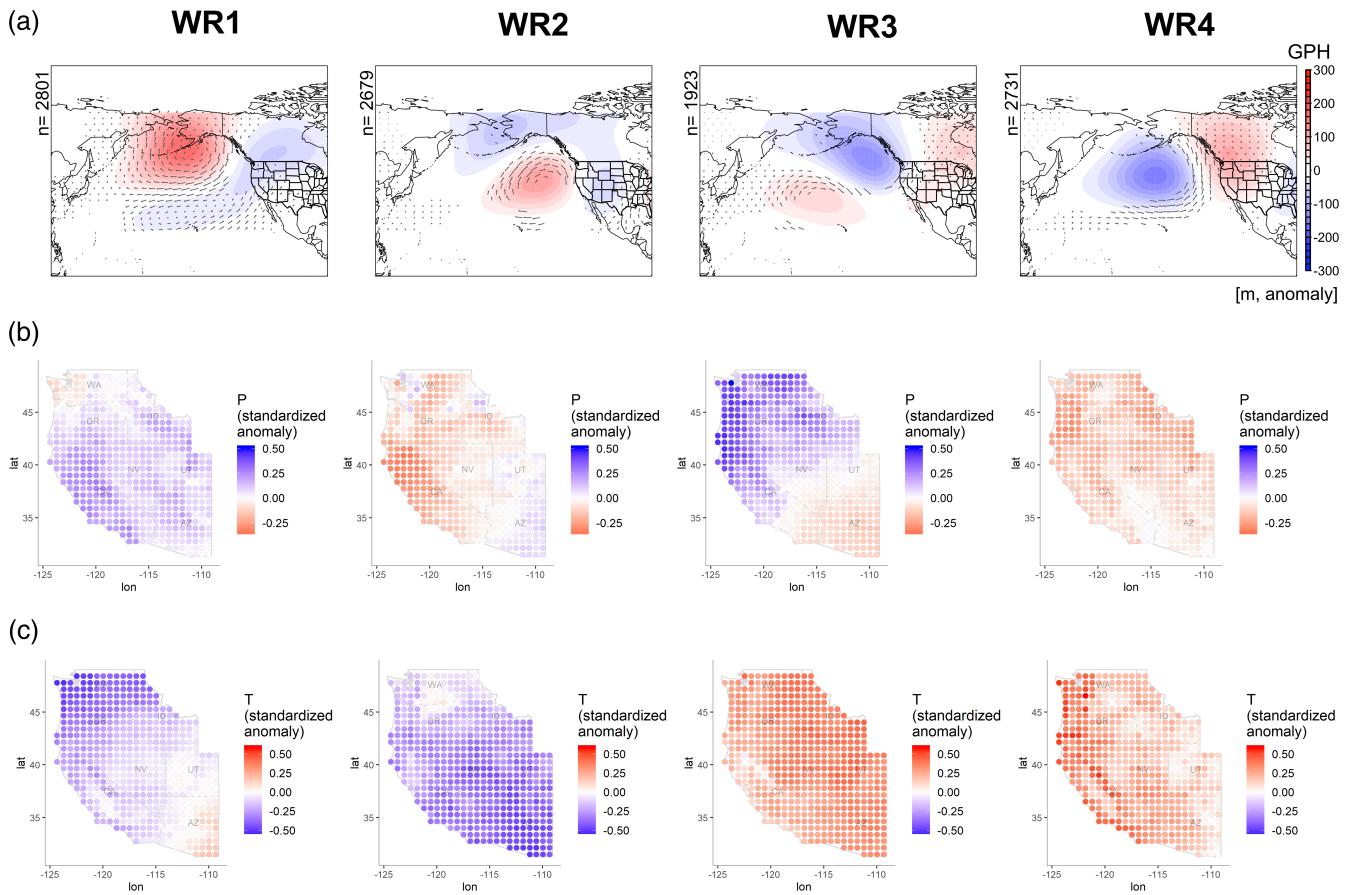
particularly between December through February, while the frequency of the other WRs is relatively consistent across the cold season (see Figure S5). Finally, WR4 exhibits a trough over the Central Pacific and a ridge over western North America, leading much of the Western U.S. to be drier and warmer than average. The regional precipitation patterns by WR seen in Figure 9b also emerge when examining the distribution of non-zero daily precipitation and the number of wet days within each of the 13 river basins in California examined above (see Figure S6).

### 3.3 | Metastability of weather regimes

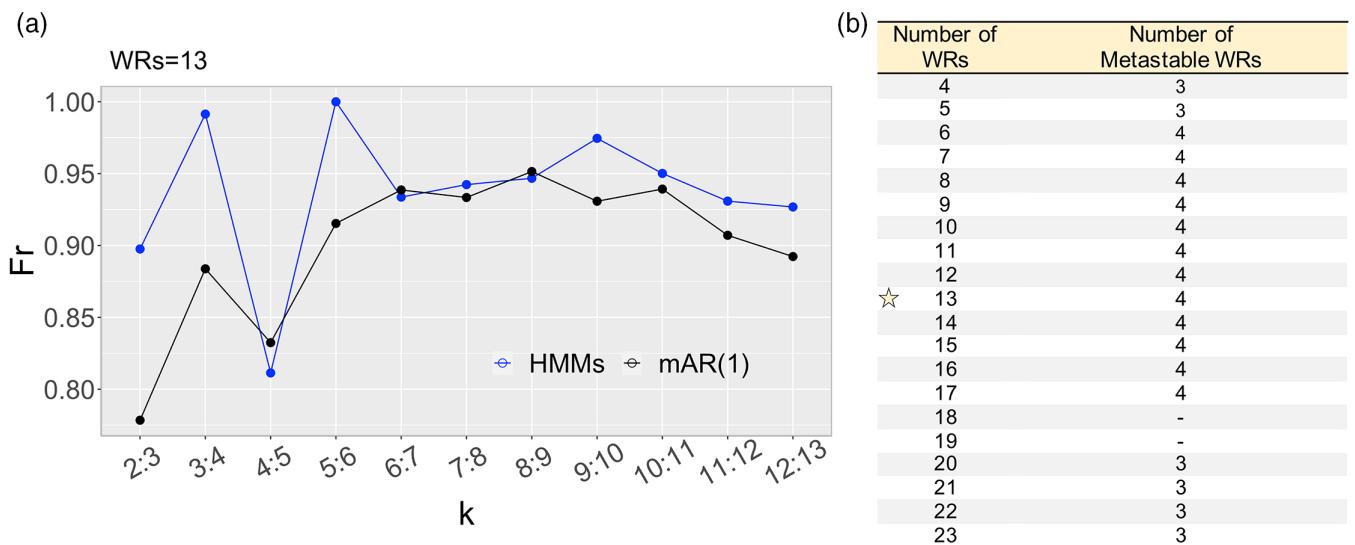
We now assess the number of metastable atmospheric regimes based on the steps in Section 2.5 to determine if this number is related to the number of states that maximize WR-SWG performance. As detailed in Section 2.5, the number of metastable regimes can be identified from the t.p.m. of an HMM defined with any  $K$  number of states. We conducted metastability tests based on the ratio of log-eigenvalues (Fr, Equation 6) for  $K = 4$  through  $K = 23$  states to ensure the results are consistent regardless of the value of  $K$ . In Figure 10a, we show the values of Fr for  $K = 13$  states, which is in the middle of the 20 different numbers of states tested. Also shown in Figure 10a are the median values of Fr taken from a 100-member ensemble simulated from the mAR(1) reference model. The values of Fr for the HMM defined using  $K = 13$  clearly dip below that of the reference model for

the ratio of the fourth and fifth eigenvalue, suggesting that there are four metastable regimes in the atmospheric data. This is equal to the optimal number of regimes for WR-SWG performance identified in Section 3.2. Similar plots to Figure 10a are shown in Figure S7 (Supplementary Materials) when fitting the HMM with  $K = 4$  through  $K = 23$  states, and a summary of those tests is shown in Figure 10b. Regardless of the number of WRs used to define the HMM, we consistently identify 3–4 MRs. We note that three MRs emerge when the number of WRs in the HMM is large ( $K \geq 20$ ), or there are insufficient Fr values ( $K \leq 5$ ), suggesting that the sample size available to estimate the t.p.m might be influencing the number of MRs. Also, in some instances, no MRs were easily identified ( $K = 18, 19$ ). However, the results generally find a relatively consistent, small number of MRs, and we proceed assuming four MRs given that this is the most frequent selection.

Hereafter, we focus on the HMM fit with 13 states as an illustrative example, and based on the analysis in Figure 10. Using G-PCCA and presuming four MRs, we calculate the probability that each of the 13 WRs is a member of the four different MRs (labelled MR1, MR2, MR3, and MR4). Figure 11 shows the results of the G-PCCA membership aggregation. The colour scale is designed to highlight probabilities that are greater than 25%, as this threshold represents the probability of membership across the four MRs by random chance. Figure 11 shows clear clusters of WRs associated with different MRs. For instance, MR1 is most associated with WR4, WR7, and to a lesser extent, WR5, WR1, WR6, and

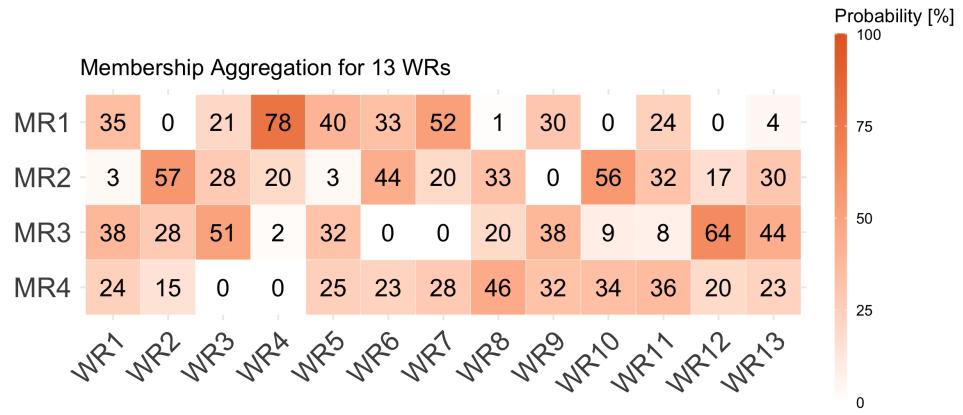


**FIGURE 9** (a) Composites of 500-hPa geopotential height (GPH) anomalies (shaded, [m]) and integrated water vapour transport anomalies (IVT) (arrows) over days categorized under each WR for a 4-state HMM. IVT is only shown when it exceeds its mean seasonal value. The number of days ( $n$ ) for WR occurrences across the entire record is also shown. Composites of locally standardized (b) precipitation ( $P$ ) and (c) temperature ( $T$ ) anomalies over the corresponding days under each WR across the Western United States [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 10** (a) The ratio of log-eigenvalues (Fr) for an HMM fit with  $K = 13$  states (blue), as well as for the median of 100 ensemble members from the mAR(1) reference model (black). (b) Number of metastable regimes (MRS) for an HMM fit with  $K = 4$  through  $K = 23$  states (WRs), concluded from similar plots as in (a) (shown in Figure S7) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**FIGURE 11** Probability that each of the 13 WRs is a member of the four metastable regimes (MRs), as determined by G-PCCA. Probabilities (in %) are rounded up to the nearest integer [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

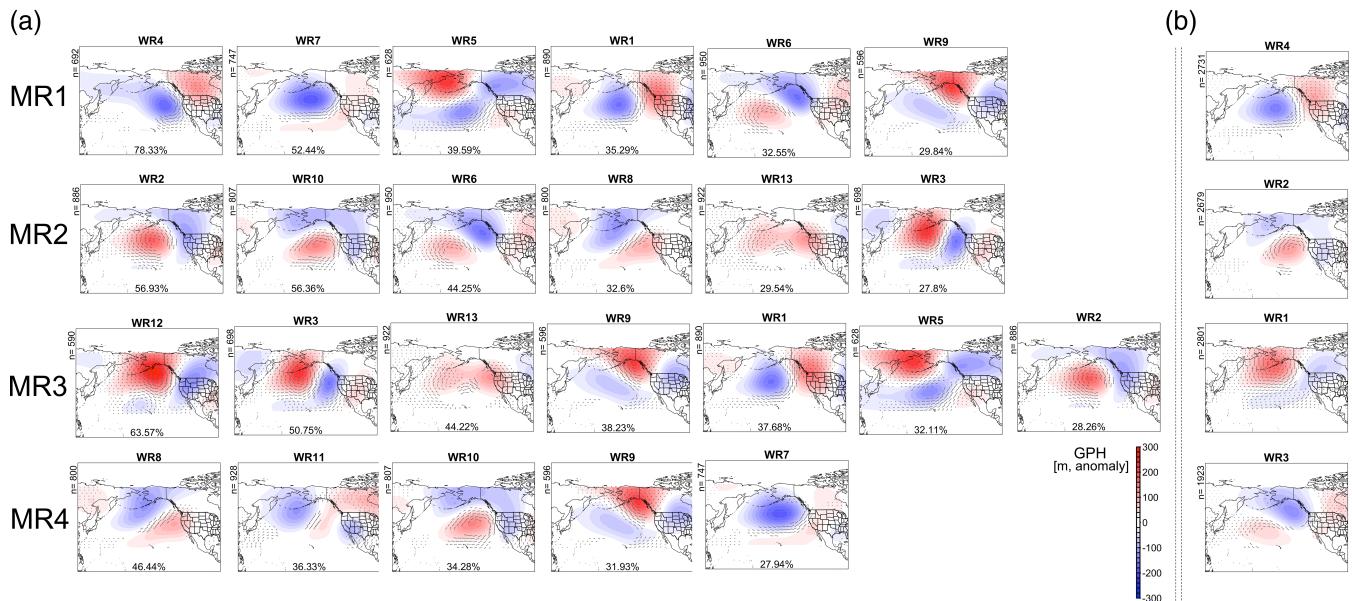


WR9. Similarly, MR3 is composed primarily of WR12, WR3, WR13, WR9, WR1, WR5, and WR2.

The four MRs highlight clusters of WRs that tend to transition between each other frequently, but rarely transition to other WRs in another MR. This can be seen by examining the t.p.m. of the 13-state HMM (Figure S8). Additional insight is possible by visualizing the atmospheric states of each WR that have the highest probability of membership in any given MR. Figure 12a shows a synoptic composite of 500-hPa geopotential height and IVT anomalies for WRs from the 13-state HMM with at least 25% probability of membership in each of the 4 MRs. For comparison, in Figure 12b, we plot similar

composites but for WRs defined using a 4-state HMM (as shown in Figure 9), which was deemed the most skillful in Section 3.2. The purpose of this comparison is to juxtapose the atmospheric patterns in a low order HMM with good WR-SWG performance to atmospheric patterns in a higher-order HMM based on their metastable clustering.

Characteristics of cold season circulation patterns can be identified using the behaviour of MRs. MR1 (see Figure 12a) is defined primarily by a trough over the Central/Eastern Pacific Ocean and a weaker ridge to the east over Western North America. MR1 is composed of different permutations of this pattern, as can be seen by the



**FIGURE 12** (a) Atmospheric flow composited over days categorized into each weather regime (WR) from a 13-state HMM with >25% of being classified into one of the four different metastable weather regimes (MRs). WRs have been ranked from the highest to lowest membership probability (in %). Composites show anomalies of 500-hPa geopotential height (GPH) (shaded, [m]) and integrated water vapour transport (IVT) (arrows). IVT is only shown when it exceeds its mean seasonal value. The number of days (*n*) for WR occurrences is also shown. (b) Atmospheric flow composited over days categorized under each WR for a 4-state HMM [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

WRs with the highest membership probability for MR1. WRs 4, 7, 1, and 6 all exhibit slightly different locations and intensities of the trough/ridge pair, while in WR5 and WR9, the low-pressure region is broader, more developed in the Central Pacific, and extends into central North America. In the 4-state HMM, a characteristic trough/ridge pattern is identified that broadly resembles many of the 13-state HMM patterns clustered into MR1 (Figure 12b). Such a pattern seems to be similar to the positive phase of the Pacific/North American (PNA+) pattern (e.g., Barnston and Livezey, 1987; Kimoto and Ghil, 1993; Risbey *et al.*, 2015).

Conversely, in MR2, a ridge lay over the Central/Eastern Pacific with an adjacent trough over Western North America (comparable with the negative phase of PNA (PNA-)). This pattern strongly resembles WR2 from a 4-state HMM. Similar to MR1, the WRs that compose MR4 from the 13-state model are mostly variants of this basic spatial pattern (e.g., WR2, WR10, WR6, WR8). Importantly though, not all WRs with a high probability of membership in a MR need to exhibit a similar composite to the others, as some WRs may act as transition WRs that have a high probability of membership because they serve more frequently as the entrance or exit WRs from a particular MR.

The WRs that compose MR3 (Figure 12a) correspond to WR1 from a 4-state HMM (Figure 12b), as both show a southwest-to-northeast oriented trough that extends from the Central Pacific to Northwest North America. There is also a strong ridge over the North Pacific Ocean and the Gulf of Alaska, clearly seen in WR12, WR3, WR13, WR9, and WR5 (Figure 12a). This pattern is associated with strong moisture transport over Northern California and has been linked to Atmospheric Rivers and episodes of extreme precipitation across the region (e.g., Ralph *et al.*, 2006; Guan and Waliser, 2015 and Zhang and Villarini, 2018).

MR4 contains a mixture of WR8, WR11, WR10, WR9, and WR7, respectively (Figure 12a), although none of the WRs have a probability of membership greater than 50%. This MR contains a larger diversity of patterns, including a ridge over the Eastern Pacific (WR8, WR10), an Omega block (WR11, WR9), and a deep Aleutian low (WR7). The patterns that compose MR4 exhibit the least resemblance to the remaining WR in the 4-state HMM (WR3; Figure 12b), although a trough near Alaska is common to almost all of them.

## 4 | SUMMARY AND CONCLUSIONS

This study proposes a framework to identify the best set of WRs to support WR-SWG at the regional scale. We

tested two approaches to classify large-scale atmospheric flow across the Pacific/North American sector into WRs, namely *K*-means clustering and HMMs, as well as the number of states that should be selected. These components of WR-SWG design were assessed based on bias and variance in a variety of precipitation metrics (moments, spells, extremes) in a case study across 13 river basins in California. This experimental design is similar to that of recent studies looking to optimize the selection of WRs (e.g., Huth *et al.*, 2016 and citations within), but is unique in that it is tailored towards SWG performance across a region. We then analysed the metastability of WRs in the Pacific/North American sector to determine how the number of WRs that optimize WR-SWG performance compares against a parsimonious, low-order identification of atmospheric regimes.

The results of this work support the primary conclusions below:

1. WR-SWG performance improved when WRs were identified using HMMs rather than *K*-means. Seasonal variability of total precipitation, wet and dry spell durations, and maximum seasonal precipitation were better simulated using WRs defined using HMMs, indicating that HMMs are better able to capture the persistence of WRs with their transitioning dynamics compared to the *K*-means algorithm. This persistence propagates into key statistics of simulated precipitation.
2. Overall, using HMMs with four states led to the best performance of simulated precipitation across multiple metrics. In particular, four states led to minimum bias and variance in seasonal standard deviation, average and maximum length of wet spells, and precipitation maxima over multiple rolling windows. There were limited spatial gradients of performance reported across the river basins tested in this study.
3. The metastability analysis indicated that there are four metastable regimes in the Pacific/North American sector, independent of the initial number of WRs and consistent with the number of WRs that most improved WR-SWG performance.
4. Synoptic composites of four metastable regimes with their prominent WRs from a 13-state HMM indicated that they reasonably match a low order HMM with four states, particularly for metastable regimes that contain individual WRs with a high probability of membership (>50%).

The conclusions above and scope of this work do depend on some assumptions that underscore our analysis and require additional discussion. First, we limited our analysis to WR-based SWGs, which are a subset of

SWGs that focus on linking local weather to global-scale atmospheric circulation. Many other, non-WR-based SWGs are available that perform very well from a statistical perspective (e.g., Peleg *et al.*, 2017; Evin *et al.*, 2018; Singer *et al.*, 2018). The performance of our model is comparable to others, with some underperformance in certain metrics (spell lengths) and over-performance in other metrics (inter-site correlation). In addition, we applied our framework using one WR-SWG and thus some of the results could be specific to the structure of that model. It is not clear if the findings would generalize if we used a different weather generator; this should be the subject of addition tests. However, we do not believe it is a coincidence that the best WR-SWG performance was based on a number of states equal to the number of metastable regimes. We posit that this result stems from the fact that a WR-SWG based on metastable regimes is better able to simulate spells, because it does not have to simulate frequent transitions between WRs that together form the atmospheric setup that underscores extended wet and dry weather. That is, MRs by definition tend to be independent of each other, and so a WR-SWG with WRs equal to the number of MRs does not need to accurately simulate frequent WR transitions to correctly capture persistent weather systems.

Another limitation is that we applied our analysis to a (small) set of 13 basins across California. More case studies are required across the United States and globally to assess the utility of the proposed framework and general performance of the WR-SWG used in this work. Part of the reason we limited the number of case study basins was because of the computational expense of the approach. The experimental design (including long records of simulation, 13 moderate-to-large river basins with data at high spatial resolution, two clustering methods and many different numbers of states, 200 ensemble members, and the calculation of performance metrics by grid cell and ensemble member) required 5–7 days of computing time using a Desktop System Type x64-based PC with Intel(R) Core(TM) i7-9700 CPU 3.00 GHz 8 Cores 8 Logical Processors and an installed physical memory (RAM) of 64.0GB. Therefore, additional applications of this framework for larger regions will likely require high performance computing to support the approach.

Finally, we only tested two design components of WR identification (the classification algorithm and the number of states), but other aspects of the design are likely relevant and should be tested. For instance, the spatial domain of the atmospheric data, the atmospheric variables included in WR clustering, and the temporal scale (e.g., seasonal vs. annual) over which WRs are identified. Such underlying design components have been tested in

other climatological studies (Huth *et al.*, 2008; Beck *et al.*, 2016), but not in the context of WR-SWG performance. We leave these explorations for future work.

Ultimately, the framework and results presented in this paper aim to help bridge the gap between climate vulnerability analyses that utilize SWGs and emerging insights in hydroclimatology. In particular, a better understanding of WRs and their dynamics can help craft process-oriented scenarios of dynamic climate change in vulnerability assessments. For example, future work could examine whether similar WRs to those in the historic record emerge in projections of future climate (e.g., Brigode *et al.*, 2018; Richardson *et al.*, 2019); whether those WRs shift in frequency, persistence, or intra-seasonal variability under warming (Rojas *et al.*, 2013; Francis *et al.*, 2018; Díaz-Estebar *et al.*, 2020; Fabiano *et al.*, 2020); or if completely new WRs are identified that need to be modelled. The metastability of WRs in future projections could also help to assess whether regime-like structures are being captured in the current generation of climate models, and to organize scenarios of future change for impact assessments, for instance by altering the likelihood of all WRs within a MR. This is one of our main research tracks to develop in the near future.

## ACKNOWLEDGEMENTS

The data used in this analysis are publicly available through the NOAA online repositories, including the gridded and station-based precipitation and NCEP/NCAR reanalysis datasets (<https://www.psl.noaa.gov/>). The authors acknowledge support from NSF Grant AGS1702273 and USDA Grant 2019-67019-30122.

## ORCID

Nasser Najibi  <https://orcid.org/0000-0002-0515-7192>

## REFERENCES

- Agel, L., Barlow, M., Feldstein, S.B. and Gutowski, W.J. (2018) Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast. *Climate Dynamics*, 50, 1819–1839. <https://doi.org/10.1007/s00382-017-3724-8>.
- Ailliot, P., Denis, A., Monbet, V. and Naveau, P. (2015) Stochastic weather generators: an overview of weather type models. *J. la Société Française Stat. (Journal French Stat. Soc. J-SFdS)*, 156(1), 101–113.
- Bardossy, A. and Plate, E.J. (1991) Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. *Journal of Hydrology*, 122, 33–47. [https://doi.org/10.1016/0022-1694\(91\)90170-M](https://doi.org/10.1016/0022-1694(91)90170-M).
- Bardossy, A. and Plate, E.J. (1992) Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, 28, 1247–1259. <https://doi.org/10.1029/91WR02589>.
- Barnston, A.G. and Livezey, R.E. (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation

- patterns. *Monthly Weather Review*, 115, 1083–1126. [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2).
- Baum, L.E. and Petrie, T. (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37, 1554–1563. <https://doi.org/10.1214/aoms/1177699147>.
- Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41, 164–171. <https://doi.org/10.1214/aoms/1177697196>.
- Beck, C., Philipp, A. and Streicher, F. (2016) The effect of domain size on the relationship between circulation type classifications and surface climate. *International Journal of Climatology*, 36, 2692–2709. <https://doi.org/10.1002/joc.3688>.
- Betrò, B., Bodini, A. and Cossu, Q.A. (2008) Using a hidden Markov model to analyse extreme rainfall events in central-East Sardinia. *Environmetrics*, 19, 702–713. <https://doi.org/10.1002/env.944>.
- Birt, A.G., et al. (2010) A simple stochastic weather generator for ecological modeling. *Environmental Modelling & Software*, 25, 1252–1255. <https://doi.org/10.1016/j.envsoft.2010.03.006>.
- Breinl, K., Di Baldassarre, G., Girons Lopez, M., Hagenlocher, M., Vico, G. and Rutgersson, A. (2017) Can weather generation capture precipitation patterns across different climates, spatial scales and under data scarcity? *Scientific Reports*, 7, 1–12. <https://doi.org/10.1038/s41598-017-05822-y>.
- Brigode, P., Gérardin, M., Bernardara, P., Gailhard, J. and Ribstein, P. (2018) Changes in French weather pattern seasonal frequencies projected by a CMIP5 ensemble. *International Journal of Climatology*, 38, 3991–4006. <https://doi.org/10.1002/joc.5549>.
- Cahynová, M. and Huth, R. (2016) Atmospheric circulation influence on climatic trends in Europe: an analysis of circulation type classifications from the COST733 catalogue. *International Journal of Climatology*, 36, 2743–2760. <https://doi.org/10.1002/joc.4003>.
- Chen, J. and Brissette, F.P. (2014) Comparison of five stochastic weather generators in simulating daily precipitation and temperature for the loess plateau of China. *International Journal of Climatology*, 34, 3089–3105. <https://doi.org/10.1002/joc.3896>.
- Cioffi, F., Conticello, F., Lall, U., Marotta, L. and Telesca, V. (2017) Large scale climate and rainfall seasonality in a Mediterranean area: insights from a non-homogeneous Markov model applied to the agro-Pontino plain. *Hydrological Processes*, 31, 668–686. <https://doi.org/10.1002/hyp.11061>.
- Corte-Real, J., Xu, H. and Qian, B. (1999) A weather generator for obtaining daily precipitation scenarios based on circulation patterns. *Climate Research*, 13, 61–75. <https://doi.org/10.3354/cr013061>.
- Daly, C., Neilson, R.P. and Phillips, D.L. (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, 33, 140–158. [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2).
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Deuflhard, P. and Weber, M. (2005) Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and Its Applications*, 398, 161–184. <https://doi.org/10.1016/j.laa.2004.10.026>.
- Deuflhard, P., Huiszinga, W., Fischer, A. and Schütte, C. (2000) Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra and its Applications*, 315, 39–59. [https://doi.org/10.1016/S0024-3795\(00\)00095-1](https://doi.org/10.1016/S0024-3795(00)00095-1).
- Díaz-Esteban, Y., Raga, G.B. and Díaz Rodríguez, O.O. (2020) A weather-pattern-based evaluation of the performance of CMIP5 models over Mexico. *Climate*, 8, 5. <https://doi.org/10.3390/cli8010005>.
- Diday, E. and Simon, J.C. (1976) *Clustering Analysis*. (Digital Pattern Recognition). Berlin, Heidelberg: Springer, pp. 47–94. [https://doi.org/10.1007/978-3-642-96303-2\\_3](https://doi.org/10.1007/978-3-642-96303-2_3).
- Evin, G., Favre, A.-C. and Hingray, B. (2018) Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences*, 22, 655–672. <https://doi.org/10.5194/hess-22-655-2018>.
- Fabiano, F., Meccia, V., Davini, P., Ghinassi, P. and Corti, S. (2020) A regime view of future atmospheric circulation changes in Northern mid-latitudes. *Weather Climate Dynamics Discussion*, 1–24. <https://doi.org/10.5194/wcd-2020-37>.
- Fackeldey, K., Sikorski, A. and Weber, M. (2018) Spectral clustering for non-reversible Markov chains. *Computational and Applied Mathematics*, 37, 6376–6391. <https://doi.org/10.1007/s40314-018-0697-0>.
- Farnham, D.J., Doss-Gollin, J. and Lall, U. (2018) Regional extreme precipitation events: robust inference from credibly simulated GCM variables. *Water Resources Research*, 54, 3809–3824. <https://doi.org/10.1002/2017WR021318>.
- Forsythe, N., Fowler, H.J., Blenkinsop, S., Burton, A., Kilsby, C.G., Archer, D.R., Harpham, C. and Hashmi, M.Z. (2014) Application of a stochastic weather generator to assess climate change impacts in a semi-arid climate: the upper Indus Basin. *Journal of Hydrology*, 517, 1019–1034. <https://doi.org/10.1016/j.jhydrol.2014.06.031>.
- Francis, J.A., Skifc, N. and Vavrus, S.J. (2018) North American weather regimes are becoming more persistent: Is arctic amplification a factor? *Geophysical Research Letters*, 45. <https://doi.org/10.1029/2018GL080252>.
- Franzke, C., Crommelin, D., Fischer, A. and Majda, A.J. (2008) A hidden Markov model perspective on regimes and metastability in atmospheric flows. *Journal of Climate*, 21, 1740–1757. <https://doi.org/10.1175/2007JCLI1751.1>.
- Ghahramani, Z. (2001) An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15, 9–42. <https://doi.org/10.1142/S0218001401000836>.
- Ghil, M. and Robertson, A.W. (2002) “Waves” vs. “particles” in the atmosphere’s phase space: a pathway to long-range forecasting? *Proceedings of the National Academy of Sciences of the United States of America*, 99, 2493–2500. <https://doi.org/10.1073/pnas.012580899>.
- Guan, B. and Waliser, D.E. (2015) Detection of atmospheric rivers: evaluation and application of an algorithm for global studies. *Journal of Geophysical Research*, 120, 12514–12535. <https://doi.org/10.1002/2015JD024257>.
- Guo, L., Jiang, Z. and Chen, W. (2018) Using a hidden Markov model to analyze the flood-season rainfall pattern and its

- temporal variation over East China. *J. Meteorol. Res.*, 32, 410–420. <https://doi.org/10.1007/s13351-018-7107-9>.
- Gupta, H.V., Sorooshian, S. and Yapo, P.O. (1998) Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research*, 34, 751–763. <https://doi.org/10.1029/97WR03495>.
- Haines, K. and Hannachi, A. (1995) Weather regimes in the Pacific from a GCM. *Journal of the Atmospheric Sciences*, 52, 2444–2462. [https://doi.org/10.1175/1520-0469\(1995\)052<2444:WRITPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<2444:WRITPF>2.0.CO;2).
- Hannachi, A., Straus, D.M., Franzke, C.L.E., Corti, S. and Woollings, T. (2017) Low-frequency nonlinearity and regime behavior in the northern hemisphere extratropical atmosphere. *Reviews of Geophysics*, 55, 199–234. <https://doi.org/10.1002/2015RG000509>.
- Hay, L.E., McCabe, G.J., Wolock, D.M. and Ayers, M.A. (1991) Simulation of precipitation by weather type analysis. *Water Resources Research*, 27, 493–501. <https://doi.org/10.1029/90WR02650>.
- Hewitson, B. and Crane, R. (2002) Self-organizing maps: applications to synoptic climatology. *Climate Research*, 22, 13–26. <https://doi.org/10.3354/cr022013>.
- Holsclaw, T., Greene, A.M., Robertson, A.W. and Smyth, P. (2016) A Bayesian hidden Markov model of daily precipitation over south and East Asia. *Journal of Hydrometeorology*, 17, 3–25. <https://doi.org/10.1175/JHM-D-14-0142.1>.
- Hughes, J.P. and Guttorp, P. (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, 30, 1535–1546. <https://doi.org/10.1029/93WR02983>.
- Hughes, J.P., Guttorp, P. and Charles, S.P. (1999) A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 48, 15–30. <https://doi.org/10.1111/1467-9876.00136>.
- Huiszinga, W., Schütte, C. and Stuart, A.M. (2003) Extracting macroscopic stochastic dynamics: model problems. *Communications on Pure and Applied Mathematics*, 56, 234–269. <https://doi.org/10.1002/cpa.10057>.
- Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kysely, J. and Tveito, O.E. (2008) Classifications of atmospheric circulation patterns: recent advances and applications. *Annals of the New York Academy of Sciences*, 1146, 105–152.
- Huth, R., Beck, C. and Kučerová, M. (2016) Synoptic-climatological evaluation of the classifications of atmospheric circulation patterns over Europe. *International Journal of Climatology*, 36, 2710–2726. <https://doi.org/10.1002/joc.4546>.
- Ivanov, V.Y., Bras, R.L. and Curtis, D.C. (2007) A weather generator for hydrological, ecological, and agricultural applications. *Water Resources Research*, 43(10), 1–21. <https://doi.org/10.1029/2006WR005364>.
- Johnson, F. and Sharma, A. (2009) Measurement of GCM skill in predicting variables relevant for Hydroclimatological assessments. *Journal of Climate*, 22, 4373–4382. <https://doi.org/10.1175/2009JCLI2681.1>.
- Kalnay, E., et al. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Katz, R.W. and Parlange, M.B. (1995) Generalizations of chain-dependent processes: application to hourly precipitation. *Water Resources Research*, 31, 1331–1341. <https://doi.org/10.1029/94WR03152>.
- Kavvas, M.L., Puri, P.S. and Saquib, M.N. (1988) Analytical extensions on a stochastic description of cyclonic precipitation fields in time and space. *Stochastic Hydrology and Hydraulics*, 2, 263–279. <https://doi.org/10.1007/BF01544040>.
- Khalil, A.F., Kwon, H.-H., Lall, U. and Kaheil, Y.H. (2010) Predictive downscaling based on non-homogeneous hidden Markov models. *Hydrological Sciences Journal*, 55, 333–350. <https://doi.org/10.1080/02626661003780342>.
- Khazaei, M.R., Zahabiyoun, B. and Saghafian, B. (2012) Assessment of climate change impact on floods using weather generator and continuous rainfall-runoff model. *International Journal of Climatology*, 32, 1997–2006. <https://doi.org/10.1002/joc.2416>.
- Kim, Y. and Lee, G.W. (2017) Stochastic precipitation generator with hidden state covariates. *Asia-Pacific Journal of Atmospheric Sciences*, 53, 353–359. <https://doi.org/10.1007/s13143-017-0037-0>.
- Kimoto, M. and Ghil, M. (1993) Multiple flow regimes in the northern hemisphere winter. Part I: methodology and hemispheric regimes. *Journal of the Atmospheric Sciences*, 50, 2625–2644. [https://doi.org/10.1175/1520-0469\(1993\)050<2625:MFRITN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<2625:MFRITN>2.0.CO;2).
- Kwasniok, F. (2008) Regime prediction and predictability in nonlinear dynamical systems. *The European Physical Journal Special Topics*, 165, 35–44. <https://doi.org/10.1140/epjst/e2008-00847-y>.
- Kwasniok, F. (2014) Enhanced regime predictability in atmospheric low-order models due to stochastic forcing. *Philosophical Transactions of the Royal Society A—Mathematical Physical and Engineering Sciences*, 372, 20130286. <https://doi.org/10.1098/rsta.2013.0286>.
- Legras, B. and Ghil, M. (1985) Persistent anomalies, blocking and variations in atmospheric predictability. *Journal of the Atmospheric Sciences*, 42(5), 433–471. [https://doi.org/10.1175/1520-0469\(1985\)042<0433:pabavi>2.0.co;2](https://doi.org/10.1175/1520-0469(1985)042<0433:pabavi>2.0.co;2).
- Livneh, B., Bohn, T.J., Pierce, D.W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D.R. and Brekke, L. (2015) A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and southern Canada 1950–2013. *Scientific Data*, 2, 2. <https://doi.org/10.1038/sdata.2015.42>.
- MacDonald, I.L. and Zucchini, W. (1997) Hidden Markov and Other Models for Discrete-valued Time Series. Vol. 110, CRC Press, Boca Raton, Florida, pp. 1–256.
- Majda, A.J., Franzke, C.L., Fischer, A. and Crommelin, D.T. (2006) Distinct metastable atmospheric regimes despite nearly Gaussian statistics: a paradigm model. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 8309–8314. <https://doi.org/10.1073/pnas.0602641103>.
- Mavromatis, T. and Jones, P.D. (1998) Comparison of climate change scenario construction methodologies for impact assessment studies. *Agricultural and Forest Meteorology*, 91, 51–67. [https://doi.org/10.1016/S0168-1923\(98\)00063-X](https://doi.org/10.1016/S0168-1923(98)00063-X).
- Michelangeli, P.A., Vautard, R. and Legras, B. (1995) Weather regimes: recurrence and quasi stationarity. *Journal of the Atmospheric Sciences*, 52, 1237–1256. [https://doi.org/10.1175/1520-0469\(1995\)052<1237:WRRAQS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2).

- Milrad, S. (2018) Synoptic analysis and forecasting: an introductory kit. *Weather*, 17, 41–43. <https://doi.org/10.1002/j.1477-8696.1962.tb07301.x>.
- Muñoz, Á.G., Yang, X., Vecchi, G.A., Robertson, A.W. and Cooke, W.F. (2017) A weather-type-based cross-time-scale diagnostic framework for coupled circulation models. *Journal of Climate*, 30, 8951–8972. <https://doi.org/10.1175/JCLI-D-17-0115.1>.
- Peleg, N., Fatichi, S., Paschalidis, A., Molnar, P. and Burlando, P. (2017) An advanced stochastic weather generator for simulating 2-D high-resolution climate variables. *Journal of Advances in Modeling Earth Systems*, 9, 1595–1627. <https://doi.org/10.1002/2016MS000854>.
- Philipp, A., Beck, C., Huth, R. and Jacobbeit, J. (2016) Development and comparison of circulation type classifications using the COST 733 dataset and software. *International Journal of Climatology*, 36, 2673–2691. <https://doi.org/10.1002/joc.3920>.
- Pineda, L.E. and Willems, P. (2016) Multisite downscaling of seasonal predictions to daily rainfall characteristics over pacific-andean river basins in Ecuador and Peru using a non-homogeneous hidden markov model. *Journal of Hydrometeorology*, 17, 481–498. <https://doi.org/10.1175/JHM-D-15-0040.1>.
- Rabiner, L.R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 257–286.
- Ralph, F.M., Neiman, P.J., Wick, G.A., Gutman, S.I., Dettinger, M.D., Cayan, D.R. and White, A.B. (2006) Flooding on California's Russian River: role of atmospheric rivers. *Geophysical Research Letters*, 33, L13801. <https://doi.org/10.1029/2006GL026689>.
- Reinhold, B.B. and Pierrehumbert, R.T. (1982) Dynamics of weather regimes: quasi-stationary waves and blocking. *Monthly Weather Review*, 110(9), 1105–1145. [https://doi.org/10.1175/1520-0493\(1982\)110<1105:dowrqs>2.0.co;2](https://doi.org/10.1175/1520-0493(1982)110<1105:dowrqs>2.0.co;2).
- Reuter, B., Weber, M., Fackeldey, K., Röblitz, S. and Garcia, M.E. (2018) Generalized Markov state modeling method for non-equilibrium biomolecular dynamics: exemplified on amyloid  $\beta$  conformational dynamics driven by an oscillating electric field. *Journal of Chemical Theory and Computation*, 14, 3579–3594. <https://doi.org/10.1021/acs.jctc.8b00079>.
- Richardson, C.W. (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17, 182–190. <https://doi.org/10.1029/WR017i001p00182>.
- Richardson, D., Kilsby, C.G., Fowler, H.J. and Bárdossy, A. (2019) Weekly to multi-month persistence in sets of daily weather patterns over Europe and the North Atlantic Ocean. *International Journal of Climatology*, 39, 2041–2056. <https://doi.org/10.1002/joc.5932>.
- Risbey, J.S., O'Kane, T.J., Monselesan, D.P., Franzke, C. and Horenko, I. (2015) Metastability of northern hemisphere teleconnection modes. *Journal of the Atmospheric Sciences*, 72, 35–54. <https://doi.org/10.1175/JAS-D-14-0020.1>.
- Robertson, A.W. and Ghil, M. (1999) Large-scale weather regimes and local climate over the Western United States. *Journal of Climate*, 12, 1796–1813. [https://doi.org/10.1175/1520-0442\(1999\)012<1796:LSWRAL>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1796:LSWRAL>2.0.CO;2).
- Röblitz, S. and Weber, M. (2013) Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification*, 7, 147–179. <https://doi.org/10.1007/s11634-013-0134-6>.
- Rojas, R., Feyen, L. and Watkiss, P. (2013) Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation. *Global Environmental Change*, 23, 1737–1751. <https://doi.org/10.1016/j.gloenvcha.2013.08.006>.
- Robertson, A.W., Kirshner, S. and Smyth, P. (2004) Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model. *Journal of Climate*, 17, 4407–4424. <https://doi.org/10.1175/JCLI-3216.1>.
- Robertson, A.W., Kushnir, Y., Lall, U. and Nakamura, J. (2015) *Weather and Climatic Drivers of Extreme Flooding Events over the Midwest of the United States. Extreme Events: Observations, Modeling, and Economics.* (Geophysical Monograph Series) Hoboken, NJ: Wiley Blackwell, pp. 113–124.
- Sansom, J. and Thomson, P. (2001) Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability*, 38, 142–157. <https://doi.org/10.1239/jap/1085496598>.
- Schlabing, D., Frassl, M.A., Eder, M.M., Rinke, K. and Bárdossy, A. (2014) Use of a weather generator for simulating climate change effects on ecosystems: a case study on Lake Constance. *Environmental Modelling & Software*, 61, 326–338. <https://doi.org/10.1016/j.envsoft.2014.06.028>.
- Shepard, D.S. (1984) Computer mapping: the SYMAP interpolation algorithm. In: *Spatial Statistics and Models*. Netherlands: Springer, pp. 133–145.
- Singer, M.B., Michaelides, K. and Hobley, D.E.J. (2018) STORM 1.0: a simple, flexible, and parsimonious stochastic rainfall generator for simulating climate and climate change. *Geoscientific Model Development*, 11, 3713–3726. <https://doi.org/10.5194/gmd-11-3713-2018>.
- Steinschneider, S., Ray, P., Rahat, S.H. and Kucharski, J. (2019) A weather-regime-based stochastic weather generator for climate vulnerability assessments of water systems in the western United States. *Water Resources Research*, 55, 6923–6945. <https://doi.org/10.1029/2018WR024446>.
- Tveito, O.E. and Huth, R. (2016) Circulation-type classifications in Europe: results of the COST 733 action. *International Journal of Climatology*, 36, 2671–2672. <https://doi.org/10.1002/joc.4768>.
- Verbist, K., Robertson, A.W., Cornelis, W.M. and Gabriels, D. (2010) Seasonal predictability of daily rainfall characteristics in central northern Chile for dry-land management. *Journal of Applied Meteorology and Climatology*, 49, 1938–1955. <https://doi.org/10.1175/2010JAMC2372.1>.
- Visser, I. and Speekenbrink, M. (2010) depmixS4: An R package for hidden markov models. *Journal of Statistical Software*, 36, 1–21. <https://doi.org/10.18637/jss.v036.i07>.
- Weber, M. (2003) Improved Perron cluster analysis. *Zuse Institute Berlin (ZIB) Report 03-0*, 1–9. <http://nbn-resolving.de/urn:nbn:de:0297-zib-7260>.
- Weber, M. and Fackeldey, K. (2015) G-PCCA: Spectral Clustering for Non-reversible Markov Chains. *Zuse Institute Berlin (ZIB) Report, 1-13*, 1–9. <http://nbn-resolving.de/urn:nbn:de:0297-zib-55505>.
- Weber, M. and Kube, S. (2005) Robust Perron cluster analysis for various applications in computational life science. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 3695 LNBI of, Springer, Berlin, Heidelberg, pp. 57–66.

- Wilks, D.S. (1992) Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, 22, 67–84. <https://doi.org/10.1007/BF00143344>.
- Wilks, D.S. (1998) Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210, 178–191. [https://doi.org/10.1016/S0022-1694\(98\)00186-3](https://doi.org/10.1016/S0022-1694(98)00186-3).
- Wilks, D.S. and Wilby, R.L. (1999) The weather generation game: a review of stochastic weather models. *Progress in Physical Geography-Earth and Environment*, 23, 329–357. <https://doi.org/10.1177/030913339902300302>.
- Willmott, C.J. (1981) On the validation of models. *Physical Geography*, 2, 184–194. <https://doi.org/10.1080/02723646.1981.10642213>.
- Wilson, L.I., Lettenmaier, D.P. and Skillingstad, E. (1992) A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *Journal of Geophysical Research*, 97, 2791–2809. <https://doi.org/10.1029/91JD02155>.
- Zambrano-Bigiarini, M. (2019) HydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series [R package hydroGOF version 0.4-0]. Comprehensive R Archive Network (CRAN), R package version 0.4-0.
- Zhang, W. and Villarini, G. (2018) Uncovering the role of the east Asian jet stream and heterogeneities in atmospheric rivers affecting the western United States. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 891–896. <https://doi.org/10.1073/pnas.1717883115>.
- Zhang, W. and Villarini, G. (2019) On the weather types that shape the precipitation patterns across the U.S. Midwest. *Climate Dynamics*, 53, 4217–4232. <https://doi.org/10.1007/s00382-019-04783-4>.
- Zucchini, W., Macdonald, I.L. and Langrock, R. (2017) *Hidden Markov Models for Time Series: An Introduction Using R*, 2nd edition. Boca Raton, FL: CRC Press, pp. 1–370.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Najibi N, Mukhopadhyay S, Steinschneider S. Identifying weather regimes for regional-scale stochastic weather generators. *Int J Climatol*. 2021;41: 2456–2479. <https://doi.org/10.1002/joc.6969>