A Changing Climatology of Precipitation Persistence across the United States Using Information-Based Measures

ALLISON E. GOODWELL

Department of Civil Engineering, University of Colorado Denver, Denver, Colorado

PRAVEEN KUMAR

Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois

(Manuscript received 17 January 2019, in final form 9 May 2019)

ABSTRACT

The sequencing, or persistence, of daily precipitation influences variability in streamflow, soil moisture, and vegetation states. As these factors influence water availability and ecosystem health, it is important to identify spatial and temporal trends in precipitation persistence and predictability. We take an information theoretic perspective to address regional and temporal trends in daily patterns, based on the Climate Prediction Center (CPC) gridded gauge-based dataset of daily precipitation over the continental United States from 1948 to 2018. We apply information measures to binary sequences of precipitation occurrence to quantify uncertainty, predictability in the form of lagged mutual information between the current state and two time-lagged histories, and associated dominant time scales. We find that this information-based predictability is highest in the western United States, but the relative influence of longer lagged histories in comparison to a 1-day history is highest in the east. Information characteristics and time scales vary seasonally and regionally and constitute an information climatology that can be compared with traditional indices of precipitation and climate. Trend analyses over the 70-yr time period also show varying regional characteristics that differ between seasons. In addition to increasing precipitation frequency over most of the country, we detect increasing and decreasing predictability in western and eastern regions, respectively, with average trend magnitudes corresponding to shifts in predictability ranging from -50% to 110% . This new perspective on precipitation persistence has broad potential to link shifts in climate and weather to patterns and predictability of related environmental factors.

1. Introduction

As ecosystems approach thresholds due to increasing stress from climate change and other forcings, the study of variability in ecosystem drivers becomes increasingly relevant (Koster et al. 2017). Precipitation is one such integral driver of ecosystem function, and changes in both the frequency and durations of wet and dry periods have implications for the variability of many downstream ecosystem components (Rodríguez-Iturbe 2000; Knapp et al. 2008; Wu et al. 2011; Roque-Malo and

Kumar 2017). Precipitation intensities, durations, and frequencies have local and downstream impacts on vegetation, soil moisture, heat and carbon fluxes, stream ecosystems, and water scarcity or abundance (Kumar 2013). Persistence relates to the presence of temporal ordering, or dependence between past and future states. For example, in a given area, a rainy day may be likely to be followed by another rainy day, or dry periods tend to persist for several days. We focus on persistence across a range of event magnitudes, which maintains ecosystem functionality, base flow in streams, and land–atmosphere interactions. This feature of temporal ordering has long been recognized in rainfall modeling based on Markov processes (Chin 1977; Gabriel and Neumann 1961; Hay et al. 1991), autoregressive models (Muhammad et al. 2016), mixed distributions (Foufoula-Georgiou and Lettenmaier 1987), or as functions of weather types (Hay et al. 1991). Here, we define measures of persistence from an information

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-19-0013.s1>.

Corresponding author: Allison E. Goodwell, allison.goodwell@ucdenver.edu

DOI: 10.1175/JHM-D-19-0013.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

theory perspective. Regional and seasonal patterns in these measures, in addition to their temporal trends, reveal previously undefined characteristics of precipitation persistence and predictability.

Various aspects of rainfall predictability have been addressed from monthly (Gianotti et al. 2014; Anderson et al. 2015, 2016; Jiang et al. 2016a,b) to daily or variable (Dhanya and Villarini 2018) time scales. For example, potential predictability (Gianotti et al. 2014; Anderson et al. 2015, 2016) defines the degree to which rainfall is predictable at climatic or monthly time scales due to climate drivers, and daily rainfall is viewed as “weather noise.” A different predictability perspective uses the Colwell index, an information theory-based measure, to quantify variability and consistency within and between seasons (Jiang et al. 2016a,b), where predictability is defined as the “recurrent likelihood of patterns described by the metrics of magnitude, variability, and seasonality.” In contrast, inherent predictability (Dhanya and Villarini 2018) uses nonlinear analyses to assess maximum predictability, predictive error, and predictive instability of rainfall on any time scale. While perspectives on predictability differ, there are also many studies that characterize regional climatology within the United States (Marquardt Collow et al. 2016; Moore et al. 2014; Li and Li 2013; Mahoney et al. 2015; Parker and Abatzoglou 2016; Costa-Cabral et al. 2016; Hamlet et al. 2005). Some of these have detected temporal trends in intensity, quantity (Groisman et al. 1999; Chen and Dirmeyer 2017; Hoerling et al. 2016), and meteorological causes (Kunkel et al. 2012; Catto and Pfahl 2013; Catto 2016; Agel et al. 2015; Moore et al. 2014; Feng et al. 2016) of extreme or nonextreme precipitation events. For example, stronger trends in the number of extreme events, rather than their magnitudes, have been detected in the United States over the past several decades (Mallakpour and Villarini 2017).

Many studies on predictability and hydroclimatology consider nonextreme rainfall at a daily time scale to be weather noise relative to extremes or longer-term seasonal rainfall patterns (Gianotti et al. 2014; Anderson et al. 2015, 2016), such that most rainfall occurrences are omitted entirely from analyses. We characterize climatology and predictability using measures anchored in information theory that can shed light on precipitation persistence and climatic trends for a range of event magnitudes. We particularly focus on short-term persistence of daily precipitation occurrence, where the demarcation between wet and dry states is varied from a low threshold detection value of 0.3 mm to 10th and 50th percentiles for a given location. We use daily gridded precipitation data from the Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation over the contiguous United States (CONUS),

which is available at a 0.25° resolution from 1948 to 2018 (Chen et al. 2008; Xie et al. 2007). The application of information theory measures to this dataset enables us to characterize the temporal structure of precipitation in terms of information transfer, that is, the dependency between past and current states. Based on the strengths of these dependencies, we characterize many aspects of persistence, such as uncertainty, predictability in the form of reduced uncertainty given past states, and dominant time scales of predictability. Advantages of information theory over traditional linear correlation measures include the capture of both linear and nonlinear interactions, and different types of influences that may be redundant, joint, or unique from different time-lagged sources (Goodwell and Kumar 2017). Here, we show that precipitation persistence can be captured in ways that are not considered based on traditional analysis methods.

Specifically, we use information theory-based measures to quantify (i) the extent to which lagged precipitation histories reduce the uncertainty of current precipitation, (ii) associated dominant time scales of influence, and (iii) relative influences from shorter versus longer lagged histories. The extent of information transfer from past to current states provides a measure of persistence, or “control” that lagged histories exert on a current wetness state. This persistence can also be broken down into relative influences from particular lagged histories. We assess regional and temporal trends in information metrics, and compare detected trends to observed patterns in climate and meteorological causes of precipitation events. These characterizations of variability also reveal the extent to which the “dry gets drier, wet gets wetter” paradigm of climate change is an oversimplification in many regions (Roque-Malo and Kumar 2017).

This paper is organized as follows. In section 2, we discuss the information theory measures used in this study, provide an illustrative example based on synthetic datasets, and describe the daily precipitation dataset and analysis. In section 3, we present results for seasonal and spatial patterns in precipitation persistence measures for watersheds in the continental United States. In section 4, we assess temporal trends based on 3-yr nonoverlapping windows for seasonal and annual cases. We discuss these patterns and trends in a broader context of joint variability between precipitation and related processes in section 5 and provide concluding remarks in section 6.

2. Methods

a. Information theory measures

We consider X to be a binary variable describing daily precipitation occurrence, where $X_t = 1$ on day t if

precipitation is observed in an amount above a specified threshold, and $X_t = 0$ otherwise. A given temporal sequence of precipitation has attributes $p(X_t) \equiv p(X_t = 1)$ and $p(X_t = 0) = 1 - p(X_t)$, denoting the probability or frequency of wet or dry days, respectively. Shannon entropy, $H(X_t) = -\sum p(x_t) \log_2 p(x_t)$, is a measure of uncertainty of a random variable (Shannon 1948; Cover and Thomas 2006) and is measured in units of bits. In information theory, this uncertainty captures the average surprise associated with the occurrence of an event. When an event has a low probability of occurrence, there is a higher surprise when the event does occur (Pierce 1980). The uncertainty of a binary variable, $H(X_t) = -p \log_2(p) - (1 - p) \log_2(1 - p)$, where $p = p(X_t)$, peaks when $p = 0.5$. In other words, as precipitation occurrence frequency increases or decreases from 0.5, $H(X_t)$ (illustrated as circle areas in Fig. 1) decreases from its upper bound of 1 bit. In this way, the entropy of precipitation $H(X_t)$ is directly related to the proportion of wet versus dry days within a sequence.

Due to persistence in precipitation occurrence, we expect that the uncertainty of the current state $H(X_t)$ can be somewhat reduced by the knowledge of past precipitation states. Here we consider two particular past precipitation states of X_{t-1} , or “yesterday,” and a second longer lagged history as $X_{t-\tau}$ or “ τ days ago”. Mutual information $I(X_{t-1}; X_t)$ quantifies this reduction in uncertainty of X_t given X_{t-1} , and is given by

$$I(X_{t-1}; X_t) = \sum p(x_{t-1}, x_t) \log_2 \left[\frac{p(x_{t-1}, x_t)}{p(x_{t-1})p(x_t)} \right]. \quad (1)$$

Similarly, conditional mutual information $I(X_{t-\tau}; X_t | X_{t-1})$ measures a further reduction in uncertainty of X_t given the longer lagged history $X_{t-\tau}$, beyond the information already provided by X_{t-1} . If the lagged history X_{t-1} reduced all the uncertainty in X_t , the mutual information value in Eq. (1) would be equal to the entropy $H(X_t)$. In this case, the conditional mutual information value $I(X_{t-\tau}; X_t | X_{t-1})$ would be negligible, since no other lagged history can provide any information beyond this. In other words, the sequence would be well represented as a first-order Markov process. However, we expect that in many cases, the temporal structure reaches farther back in time, and the knowledge of additional time lagged histories will further reduce our uncertainty of X_t . When we consider the time lag, τ , to be that which maximizes this further reduction in uncertainty $I(X_{t-\tau}; X_t | X_{t-1})$ we can estimate time scales of longer dependencies while maintaining a low dimensional probability distribution estimation problem. The total mutual information between the current state X_t and

FIG. 1. Illustration of information measures that characterize precipitation persistence. (a) A binary sequence, for which the most recent 20 values are illustrated, has a probability distribution $p(X_t)$. Its uncertainty $H(X_t)$ is shown as a circle area at the upper right. Aspects of temporal structure are shown as ovals that indicate dominant time lags, and shading within the $H(X_t)$ circle that indicates total mutual information $I(X_{t-\tau}, X_{t-1}; X_t)$ given the knowledge of those two lagged histories. The proportions of information from the τ -day and the 1-day lagged histories, self-transfer entropy η and $\Gamma = 1 - \eta$, respectively, are indicated as arrows connecting the time-lagged histories to the current state. Trends in each of these metrics are illustrated in (b)–(d). (b) A trend in $p(X_t)$ indicates a change in precipitation frequency and thus uncertainty $H(X_t)$ (change in circle area), while a trend in I_{frac} indicates a change in the predictability of X_t based on the knowledge of past states (change in shaded area within circle). (c) A trend in τ (position of leftmost oval) indicates a shift in dominant time scales, and (d) a trend in η (relative arrow size) indicates a change between longer vs shorter term, or synergistic vs redundant, influences to X_t .

two lagged histories (X_{t-1} and $X_{t-\tau}$, $\tau > 1$), is the sum of the two previously mentioned information terms as follows:

$$\begin{aligned} I_{\text{tot},\tau} &= I(X_{t-1}, X_{t-\tau}; X_t) \\ &= I(X_{t-1}; X_t) + I(X_{t-\tau}; X_t | X_{t-1}). \end{aligned} \quad (2)$$

We denote I_{tot} as the maximum estimated value of $I_{\text{tot},\tau}$ computed over a range of τ values. For probability distributions based on precipitation data, the longer time lag τ is set to 2, 3, 4, ..., 19, 20 days such that 18 three-dimensional probability distributions $p(x_{t-\tau}, x_{t-1}, x_t)$ are estimated for a given binary precipitation sequence. The maximum value of $\tau = 20$ days was chosen based on

data requirements for probability distribution estimations. For a seasonal analysis of daily precipitation data, a season is approximately 90 consecutive days, and increasing lags could reduce the quality of the estimates of the 3D probability distributions due to decreasing effective data size. We also tested cases where both time lags varied from 1 to 20 days, but found that on average the $\tau = 1$ lagged history most reduces the uncertainty of X_t . Based on this, we simplify the analysis to consider specifically how X_{t-1} and $X_{t-\tau}$ individually and jointly reduce precipitation uncertainty $H(X_t)$. This maximum value of I_{tot} quantifies the uncertainty in today's precipitation state $H(X_t)$ that is reduced given the joint knowledge of yesterday's state and a longer τ -day lagged state. We denote normalized total mutual information I_{frac} as the proportion of $H(X_t)$ that is reduced given the information from these two time lagged histories:

$$I_{\text{frac}} = \frac{I_{\text{tot}}}{H(X_t)}. \quad (3)$$

The variable I_{frac} has units of bits per bit and indicates the extent to which the current precipitation is "controlled" by a maximally informative pair of lagged histories. For example, $I_{\text{frac}} = 0$ indicates that X_t is independent of past states X_{t-1} and $X_{t-\tau}$. At the other extreme, $I_{\text{frac}} = 1$ indicates that precipitation is completely predictable given these two past states.

When two sources jointly affect the outcome of a target, their interactions can be partitioned into additive components using information decomposition (Williams and Beer 2010). These components consist of unique, redundant, and synergistic information. Unique information is information that a single time lagged history shares with the current state X_t individually. Meanwhile, redundancy captures overlapping information that both lagged histories share with the current state, and synergy is information that is provided to the current state only when both lagged histories are known together, or jointly. In this context, we consider lagged precipitation histories as sources that provide information to a target, the current state, via these different types of relationships.

The total information from the two lagged histories I_{tot} [Eq. (2)] is composed of two unique components, one from each source, and synergistic and redundant components that sources provide jointly or in overlap. Similarly, the mutual information and conditional mutual information terms that make up I_{tot} can also be described in terms of these components. Specifically, mutual information $I(X_{t-1}; X_t)$ is composed of one unique component that X_{t-1} shares with X_t , and a redundant component that $X_{t-\tau}$ also shares with X_t . In our precipitation occurrence context, when knowing

yesterday's state reduces some amount of uncertainty about today's precipitation, some of this information comes only from knowing yesterday's state, while the rest could have been obtained given the knowledge of either yesterday's state or a previous one. This overlapping information about today's precipitation that could have been obtained given either lagged state is therefore redundant between the two lags. On the other hand, conditional mutual information $I(X_{t-\tau}; X_t | X_{t-1})$ is information that comes uniquely from the longer lagged history $X_{t-\tau}$, along with some information that arises synergistically due to the knowledge of both histories together. In precipitation terms, conditional mutual information is the uncertainty about today's state, that is, only reduced when the longer τ -lagged state is known, in addition to some extra predictability that comes from knowing both lagged histories jointly. As a synthetic example, if two days of rainfall are always followed by a third day of rainfall, knowing that it rained yesterday does not enable us to accurately predict whether it will rain today. Similarly, knowing that it rained two days ago is not enough, but if we know that it rained both yesterday and the day before, this information synergistically helps us predict that it will rain today.

Although various methods have been proposed to estimate these unique, redundant, and synergistic components of information (Goodwell and Kumar 2017; Griffith and Ho 2015; Williams and Beer 2010; Ince 2017), no single method is universally agreed upon since information theory does not provide a method directly. Here, we do not use any proposed method, but interpret existing information theory measures as the previously mentioned combinations of these information types. Specifically, we want to consider the extent to which the total information provided to the current precipitation state can be attributed to the knowledge of the immediate past, redundantly or uniquely, versus a longer time-lagged history, uniquely or synergistically. To this effect we consider these unique (U_1 , U_2), synergistic (S), and redundant (R) components as proportions of total information I_{tot} as follows:

$$\frac{I(X_{t-1}; X_t)}{I_{\text{tot}}} = U_1 + R = \Gamma, \quad (4)$$

$$\frac{I(X_{t-\tau}; X_t | X_{t-1})}{I_{\text{tot}}} = U_2 + S = \eta. \quad (5)$$

When Eqs. (2), (4), and (5) are combined, we see that $\eta + \Gamma = 1$, or correspondingly $U_1 + U_2 + R + S = 1$, such that the information components are proportions of the total information shared between two lagged precipitation states and current precipitation. We note

that η is equivalent to a normalized “self”-transfer entropy (Schreiber 2000), or conditional mutual information, while $\Gamma = 1 - \eta$ is the normalized mutual information between the current state and 1-day lagged history. Due to this, we denote η as the self-transfer entropy. For a precipitation sequence X_t , Γ quantifies the proportion of total information about X_t that can be obtained solely through the knowledge of yesterday’s state, X_{t-1} . This information could be uniquely provided by X_{t-1} , or redundantly provided by both X_{t-1} and $X_{t-\tau}$. The remaining proportion of shared information η quantifies the unique or synergistic contribution from $X_{t-\tau}$, since the information quantity η could be either unique from $X_{t-\tau}$ or synergistically provided when X_{t-1} and $X_{t-\tau}$ are known jointly. In this way, η quantifies the relative importance of knowledge of a longer time lagged history versus the assumption that the precipitation occurrence sequence is a first-order Markov process. For instance, $\eta = 0$ indicates that the maximum reduction in uncertainty about X_t is obtained from the knowledge of X_{t-1} , and any information provided by the τ -lagged history $X_{t-\tau}$ must be redundant. In contrast, $\eta = 1$ indicates that the knowledge of X_{t-1} alone does not reduce any uncertainty in X_t , and the only way it could provide any information would be through the additional knowledge of a longer τ -lagged history which enhances the dependency between X_{t-1} and X_t .

In this study, we analyze spatial patterns and temporal trends in the previously introduced four measures that quantify different aspects of precipitation (X_t) persistence:

- 1) Precipitation frequency $p(X_t)$: Regardless of temporal structure within a sequence, precipitation occurrence uncertainty $H(X_t)$ decreases from a maximum bound of 1 bit as $p(X_t)$ varies away from 0.5 (changing circle sizes in Fig. 1b). In general, very dry [$p(X_t) \rightarrow 0$] and very wet locations [$p(X_t) \rightarrow 1$] have low uncertainties.
- 2) Information transfer as a fraction of $H(X_t)$, I_{frac} [Eq. (3)]: An increase in I_{frac} indicates an increase in temporal ordering of precipitation relative to its uncertainty $H(X_t)$ (changing overlapping regions in circles in Fig. 1b). From a predictability perspective, an increase in I_{frac} corresponds to the potential to better predict X_t given a certain set of two lagged histories, X_{t-1} and $X_{t-\tau}$.
- 3) Time lag associated with I_{frac} , $2 \leq \tau \leq 20$: An increase in τ would indicate a longer time range of dependency, in that a long τ provides more information along with X_{t-1} relative to a short τ (shift along temporal axis illustrated in Fig. 1c). Together, I_{frac}

and τ provide a strength and temporal range of a joint dependency for a given precipitation sequence.

- 4) Self-transfer entropy η [Eq. (5)]: An increase in η indicates that the τ -lagged history provides an increasing proportion of information relative to the 1-day lagged history, and synergistic dependencies are increasing over redundant ones (changing arrow thickness in Fig. 1d). While I_{frac} indicates a level of predictability in the form of mutual information provided by past states, trends in η and/or τ indicate changes in the distributions of the lengths of wet or dry periods, or changes in precipitation persistence over time.

We compute information measures based on empirical probability distributions, and use a shuffled surrogates method to determine statistical significance ($\alpha = 0.01$) of $I_{\text{tot},\tau}$ for each measure (Ruddell and Kumar 2009; Goodwell and Kumar 2015). While this threshold seems strict, we note that information metrics for the majority of binary precipitation sequences are statistically significant at this level, due to largely prevalent temporal persistence. For any statistically significant I_{tot} and its associated τ , we further compute I_{frac} and η . However, if I_{tot} does not pass the significance test, the values of self-transfer entropy η and time lag τ are rendered null.

b. Example using synthetic data

Here we present an example of measures $p(X_t)$, $H(X_t)$, I_{frac} , τ , and η using generated sequences of binary time series data of length $N = 5000$. For any repeating sequence, we note that $I_{\text{tot}} = H(X_t)$, such that $I_{\text{frac}} = 1$, for at least one combination of lags 1 and $2 \leq \tau \leq 20$ (Figs. 2a–c), since the knowledge of some characteristic time lag completely reduces the uncertainty of X_t . For example, for a repeating case of 3-day rainfalls followed by 6-day dry periods [a repeating sequence of 111000000...], where $p(X_t) = 0.333$, we find that $\tau = 9$ days, which is equal to the time delay at which the sequence repeats (Fig. 2c). Similarly, for 1-day rainfalls followed by 3-day dry periods [a repeating sequence of 1000...], where $p(X_t) = 0.25$, we detect a dominant time lag $\tau = 4$ days (Fig. 2a).

For any of these repeating sequences, a shift from the repeating pattern toward an independent binomial distribution with the same $p(X_t)$ would cause I_{frac} to approach zero. To illustrate this, we flip the values of a randomly selected 5% of data points to add noise to each repeating sequence. This introduction of 5% random noise causes the value of I_{tot} to decrease drastically relative to $H(X_t)$ (inner circles in Figs. 2a–c). Meanwhile, $H(X_t)$ remains nearly constant since the overall $p(X_t)$ only changes slightly. As the noise increases

FIG. 2. Information partitioning into self-transfer entropy, $\eta = U_{\tau} + S$, and short-term mutual information, $\Gamma = 1 - \eta = U_1 + R$, for a range of synthetic binary sequences. (a)–(c) Binary repeating sequences with increasing lengths of wet durations from left to right, and decreased wet day frequencies $p(X_t)$ from top to bottom. The outer circle size indicates I_{tot} [equal to $H(X_t)$] provided to X_t from the two maximally informative lagged histories, 1 and $2 \leq \tau \leq 20$. The inner circle sizes indicate the smaller I_{tot} when 5% of the values are randomly flipped. The shaded portions indicate the proportion of short-term mutual information, $\Gamma = 1 - \eta$, relative to information obtained from the longer τ -lagged history. (d) When lengths of rainy periods are uniformly distributed instead of repeating, $I_{\text{tot}} < H(X_t)$ (i.e., $I_{\text{frac}} < 1$), and the proportions of η and Γ also change. The dashed circle indicates $H(X_t)$ for these cases, where some uncertainty remains once X_{t-1} and $X_{t-\tau}$ are known. Higher proportions of Γ here indicate a larger amount of predictability attributed to the most recent state.

further (e.g., if half of the values were flipped), I_{frac} would reach zero when I_{tot} becomes statistically insignificant. In contrast, the addition of noise does not significantly impact η or τ since the proportions of information from 1-day versus τ -lagged histories have not changed.

While the introduction of randomness in the form of noise does not influence τ or η , these attributes do shift as the temporal sequencing is altered. For example, a shift from rainfall lengths of 1 day (Fig. 2a) to 2 days (Fig. 2b), with no change in $p(X_t)$ (e.g., a shift from 101010... to 11001100...) would cause a change in η and τ but not in I_{frac} .

For these synthetic repeating cases, we find that self-transfer entropy η , which quantifies the relative information transfer from the longer time scale $X_{t-\tau}$, is larger than $\Gamma = 1 - \eta$, or the 1-day lagged individual mutual information. This is due to the previously mentioned characteristic time scale at which the sequence repeats, causing X_t to be completely predictable given a certain τ -lagged history. For example, for the sequence 10001000..., the knowledge of X_{t-1} provides some information concerning the value of X_t , but the knowledge

of X_{t-4} completely reduces the uncertainty of X_t due to the time scale of repetition. Specifically, if we know that $X_{t-4} = 1$ or $X_{t-4} = 0$, we also know that $X_t = 1$ or $X_t = 0$, respectively. In contrast, knowing that $X_{t-1} = 1$ informs us that $X_t = 0$ since a rainy day is always followed by a dry day. However, if $X_{t-1} = 0$, this could either be followed by $X_t = 0$ or $X_t = 1$, thus on average, some uncertainty remains even when X_{t-1} is known.

For precipitation, we really expect an opposite behavior in which the most recent history is more informative to the current state, that is, $\Gamma > \eta$. To illustrate this, we shift from generating repeating sequences to cases where lengths of wet periods (days where $X_t = 1$) and dry periods (days where $X_t = 0$) are uniformly distributed (Fig. 2d). For example, consider a case where $p(X_t) = 0.5$ and lengths of wet or dry periods follow $U[1, 3]$, where U here indicates a discrete uniform distribution. Here, an example sequence may be along the lines of 100111010001100..., where the sequence does not repeat but wet and dry periods have typical durations. In this case, I_{frac} is much lower (inner circle in Fig. 2d), since the current state is no longer completely predictable given any τ -lagged history. Moreover,

self-transfer entropy η is also lower, indicating that X_{t-1} provides a larger proportion of information to X_t relative to a longer lagged history. However, the lags τ at which $X_{t-\tau}$ provides the next highest shared information are similar to those of the synthetic repeating sequences (Figs. 2a–c), indicating that significant information is still provided at approximately the average duration of a wet and dry sequence combined.

This comparison between generated sequences of repeating patterns, repeating patterns with noise, and uniformly distributed durations illustrates how our precipitation persistence metrics of frequency $p(X_t)$, information transfer I_{frac} , time lag τ , and self-transfer entropy η vary with particular features. This understanding will help us interpret regional and seasonal characteristics and trends based on a gridded precipitation dataset.

c. Precipitation dataset

The gauge-based gridded daily precipitation data (Chen et al. 2008; Xie et al. 2007) are converted into binary time series based on a threshold of 0.3 mm. This is chosen as an estimate for a lower limit of gauge resolution (Roque-Malo and Kumar 2017), such that lower magnitudes in the gridded dataset likely arise from interpolation of gauge data, and are set to zero. We also perform the entire analysis with thresholds of 10th and 50th percentiles of precipitation (where magnitudes of 0.3 mm or less are already set to zero) at each grid cell to capture the persistence of precipitation sequencing for events larger than moderate (10th percentile) or median for a particular location (see “Results for higher thresholds” in the online supplemental material). The percentiles are determined for each grid cell individually based on the entire data record for that cell, such that they vary between regions but not seasons. We segment data into four seasons per year, where DJF is winter, MAM is spring, JJA is summer, and SON is fall. Bin counting is used to estimate probability distributions for each grid cell over the entire time period and 3-yr nonoverlapping windows, for both seasonal and annual cases (Fig. S1 in the online supplemental material). The 3-yr seasonal windows result in approximately 270 data points per window, a reasonable length to estimate binary probability distributions while retaining enough windows to detect trends over a 70-yr period of 1948–2018.

Values of $p(X_t)$, I_{frac} , τ , and η for each grid cell are averaged at both Hydrologic Unit Code 2 (HUC2) and HUC4 watershed scales (Seaber et al. 1987). There are 18 watersheds at the HUC2 scale, and 201 at the smaller HUC4 scale in the continental United States. We detect trends in watershed averaged η , I_{frac} , τ , and $p(X_t)$ over

the 1948–2018 period based on 3-yr nonoverlapping windows using the Sen slope method, a trend detection technique that is relatively robust to outliers (Sen 1968) (two-sided test with significance level $\alpha = 0.1$). A watershed-averaged trend is defined as the average trend value of all grid cells in a watershed, if a majority of grid cells exhibit a common trend direction (increasing, decreasing, or no trend). Otherwise, the watershed is classified as having “no trend.” Trends are then compared to existing studies of climatic and precipitation trends. While measures of time scale, information transfer, and precipitation frequency are interdependent, we note that a statistically significant trend in one does not necessarily imply a similar trend in another. As illustrated from the synthetic repeating binary sequences, regional and temporal trends in η , I_{frac} , τ , and $p(X_t)$ capture different characteristics of the variability of precipitation persistence in space and time.

3. An information-based climatology of precipitation occurrence

Seasonal and spatial characteristics of precipitation frequency $p(X_t)$, information transfer I_{frac} , and associated attributes of time scale τ , and self-transfer entropy η , reveal ways in which climatologies govern precipitation persistence for different regions. When these attributes are computed seasonally and annually for the entire time range for each grid location and averaged at a HUC2 (Seaber et al. 1987) watershed scale, we find that typically, precipitation frequency $p(X_t) < 0.5$ (Fig. 3). Based on this, high precipitation frequencies can generally be associated with higher uncertainties. In eastern U.S. watersheds, $H(X_t)$ is high and seasonal variability is low, while western U.S. watersheds have lower $H(X_t)$ but more variability between seasons. For example, most West Coast watersheds have very low $p(X_t)$ and $H(X_t)$ in summer and high $p(X_t)$ and $H(X_t)$ in winter (Fig. 3).

For all seasons, western HUC2 watersheds have higher proportions of information transfer from two lagged histories, I_{frac} , compared to eastern watersheds (relative positions of lower and upper dots in Fig. 3). For most western U.S. watersheds, this high I_{frac} can be attributed to both a high I_{tot} and a low $H(X_t)$, due to low precipitation frequency $p(X_t)$. This finding, that the knowledge of past states reduces more uncertainty of X_t in the West, indicates that precipitation occurrences there exhibit greater temporal structure, or more regular patterns of persistence. There is also a general pattern of decreasing self-transfer entropy η from east to west (Fig. 3 insets). This indicates that while precipitation occurrence is less predictable in the East, a relatively large proportion of the information

FIG. 3. Entropy $H(X_t)$ and total mutual information I_{tot} for precipitation occurrence, averaged for grid cells within each HUC2 watershed in the CONUS over the period 1948–2018. Black curves indicate $H(X_t)$ vs X_t for a theoretical binary random variable X_t . Markers along the lines indicate $H(X_t)$ computed from precipitation data, averaged for each HUC2 watershed for each season. Colors indicate average longitude of watersheds, with darker blues representing eastern regions and lighter greens representing western regions. Circles below the $H(X_t)$ curve indicate I_{tot} , the information that knowledge of X_{t-1} and $X_{t-\tau}$ provide to the current precipitation state, and the proportion of I_{tot} to $H(X_t)$ indicates I_{frac} . The insets indicate corresponding self-transfer entropy η for each season.

transfer is gained from knowledge of longer time-lagged histories. Seasonally, η is highest for eastern watersheds in the winter (DJF), indicating that for this region and season, this knowledge of a longer lagged history is the most relevant. While other seasonal differences seem relatively small at the HUC2 watershed scale, they become more apparent at the more detailed HUC4 scale which is used in the following subsections.

a. Patterns of persistence across the United States

Watershed-averaged persistence metrics at the scale of smaller HUC4 basins are similar to the HUC2 trends during all seasons, except in the Pacific Northwest where $p(X_t) > 0.5$ in the winter (Fig. 4a) in certain regions. This leads to $H(X_t) < 1$, or lower than maximum possible uncertainty due to a high probability of precipitation on any given day. Most eastern U.S. watersheds exhibit $0.2 < p(X_t) < 0.4$ over the entire year, while some watersheds in California and the Pacific Northwest vary from low [$p(X_t) < 0.2$] to high [$p(X_t) > 0.4$] precipitation frequencies from summer to winter (Figs. 4a,d).

Patterns in predictability as measured by I_{frac} are spatially coherent, in that neighboring watersheds tend to exhibit similar behaviors. However, spatial and seasonal patterns in I_{frac} differ from those of precipitation frequency $p(X_t)$. Particularly, we find that I_{frac} is consistently much higher in the West and Southwest relative to the rest of the country (Fig. 4). In most of the Central Plains (region labeled CP in Figs. 4 and 5) and eastern United States, I_{frac} peaks in fall but remains an order of magnitude lower than the I_{frac} detected along the West Coast (Fig. 4c). In more localized regions, I_{frac} is high in Florida in the spring, and is very low in the north-central United States in the summer. Differences in spatial and seasonal patterns between information transfer and precipitation frequency indicate that factors other than the number of events in a season dictate the persistence of precipitation. For example, the higher information transfer I_{frac} in the West indicates that regardless of its frequency, precipitation in this region is less “noisy,” or occurs in more predictable patterns, compared to eastern U.S. watersheds.

Further analyses of self-transfer entropy η and dominant time scale τ reveal additional aspects of persistence in terms of relative proportions of influences between shorter and longer time scales (Fig. 5). However, these must be considered within the context of information transfer I_{frac} , since η is a proportion of I_{frac} , much as I_{frac} is a proportion of precipitation uncertainty $H(X_t)$ [Eqs. (3)–(5)]. For example, if I_{frac} is very low, both η and the associated τ are less influential than if I_{frac} were higher. Similarly, if η is very small, then the associated τ is less influential than if η were larger, because a small η indicates that most information is obtained solely from the most recent history, X_{t-1} . To illustrate this, let us consider two cases where $\eta = 0.3$, but $I_{\text{frac}} = 0.3$ in one case and $I_{\text{frac}} = 0.03$ in the other. For the case where $I_{\text{frac}} = 0.3$ and $\eta = 0.3$, we could say that the τ -day time lag reduces 30% of 30%, or about 9%, of the total uncertainty of precipitation occurrence, while the 1-day time lag reduces the remaining 21%. For the other case where $I_{\text{frac}} = 0.03$ and $\eta = 0.3$, we would say that the τ -day time lag reduces 30% of 3%, or about 0.9%, of total uncertainty, and the 1-day lag reduces 2.1%. The latter case illustrates magnitudes similar to those detected over much of the eastern United States (Figs. 4, 5). Even so, seasonal patterns in these relatively small proportions reveal how time dependencies change with prevalent weather patterns from one season to the next.

In a seasonal analysis of η and τ in HUC4 watersheds, we find that η tends to be higher in the eastern United States, where I_{frac} is relatively low, as discussed previously (Figs. 4, 5). This indicates that while information

FIG. 4. Precipitation frequency $p(X_t)$ and information transfer I_{frac} [Eq. (3)] for (a) winter (DJF), (b) spring (MAM), (c) fall (SON), and (d) summer (JJA), during the period 1948–2018, averaged for grid cells within each HUC4 watershed. The color scale indicates I_{frac} , triangle sizes indicate $p(X_t)$, and black lines outline larger HUC2 watersheds.

transfer I_{frac} is low, synergistic time dependencies are more common in the eastern United States, in that the knowledge of multiple past states together is much more informative than only knowing the most recent precipitation state. Seasonally, η is especially high along the East Coast in winter (Fig. 5a), and in the Southwest in summer (Fig. 5d). Dominant lag times, τ , associated with I_{frac} and η also show distinct regional and seasonal patterns. In California, τ is longer than 7 days, except in the spring (Fig. 5b), indicating that long-range time dependencies are stronger than short-range dependencies for most of the year. In contrast, τ is always very short in the Northeast. In this region, high η , low I_{frac} , and short τ indicate that precipitation occurrence is not well predicted given any combination of lagged histories, but a large proportion of the information that is provided is attributable to multiple short time-lagged histories rather than to distant histories or a single state. In the western United States, precipitation uncertainty is more significantly reduced given the knowledge of lagged histories, but a larger proportion of this information is obtained from the 1-day lagged history. This could relate to a tendency in this region toward very long wet and dry periods, such that X_{t-1} is a strong predictor of X_t , and the sequence of wet and dry days typically repeats at a much longer and variable time scale, which corresponds to a longer detected τ .

In this study, we use watershed-averaged metrics to provide a regional to continental scale analysis of persistence. However, we note that some regions exhibit large variability at a much smaller spatial scale. For example, in the Rocky Mountains and surrounding areas of Colorado, the high degree of precipitation variability is reflected in the high variability in the metrics of $p(X_t)$, I_{frac} , τ , and η (Fig. S4) between individual grid cells within watersheds. This can be linked to the known seasonal shifts in extreme events in the region, where extremes occur predominantly in the fall west of the mountains, winter in the mountains, spring in the Front Range, and summer in the eastern plains (Mahoney et al. 2015). In heterogeneous regions such as this, factors such as land use, land cover, and topography are important determining factors in regional and local climates, but are not considered here in detail.

While the focus of our analysis is on a large range of precipitation magnitudes due to the 0.3-mm threshold between “wet” and “dry” states, different aspects of precipitation persistence can be observed by studying $p(X_t)$, I_{frac} , τ , and η metrics for 10th and 50th percentile thresholds. The 10th percentile case omits the “light” precipitation events that are included in the 0.3-mm analysis, and similarly, the 50th percentile considers only the range of “median to extreme” events as days where

FIG. 5. Time lag τ and self-transfer entropy η [Eq. (5)] associated with maximum I_{tot} , for (a) winter (DJF), (b) spring (MAM), (c) fall (SON), and (d) summer (JJA), during the period 1948–2018, averaged for stations within each HUC4 watershed. Color indicates η , the proportion of information provided uniquely or synergistically from the τ -lagged source $X_{t-\tau}$, relative to the information provided uniquely or redundantly from the knowledge of X_{t-1} . Circle sizes indicate the τ (days) associated with peak information transfer. Black lines outline HUC2 watersheds.

$X_t = 1$. Differences between these threshold levels indicate how characteristics vary with magnitude (see “Results for higher rainfall thresholds” in the online supplemental material; Figs. S2, S3). We find that the 10th percentile case is very similar to the 0.3-mm case, but when the lightest 50% of precipitation events are set to zero for the 50th percentile case, we find that information transfer I_{frac} is generally lower (Fig. S2). Since $H(X_t)$ is lower due to lower precipitation frequency, this indicates that the total information transfer I_{tot} is also relatively lower, such that the proportion of information to total uncertainty decreases as the precipitation threshold is increased. Since lower frequency events tend to be embedded within sequences of more frequently occurring magnitudes, changes in information transfer as thresholds are varied indicate that the predictability of precipitation occurrence depends on the range of magnitudes considered.

b. Comparisons with regional climatologies

The information-based climatologies discussed above can be compared with existing studies on precipitation predictability and regional climatologies, to determine how our metrics relate to or differ from existing

knowledge. For example, the predominant feature along the West Coast is high information transfer I_{frac} for all seasons. This can be compared to previous findings of high seasonality, a measure of variability between seasons in a year, and high constancy, a measure of similarity between years for a given season, which are prevalent in the western United States (Jiang et al. 2016a,b). This indicates that seasons vary, but within a given season, there are low levels of variability from year to year. This constancy, or similarity between years, may cause our seasonal probability distributions in this region to have a greater structure and less noise relative to regions that have high within-season variability. However, high constancy (Jiang et al. 2016b) has also been detected for the eastern United States, but we find that this does not correspond to high I_{frac} in our analysis. This could indicate that in the eastern United States, although seasons are similar over years in terms of precipitation amounts, daily temporal patterns have more variability such that the information transfer remains low. The high I_{frac} along the West Coast detected in this study can also be compared to high detected potential predictability (Anderson et al. 2015), a measure of climate influence and long-term forcing mechanism, in the

region. In essence, regular seasonal climate-related drivers dominate precipitation occurrences on the West Coast more heavily than the rest of the country.

In the Northwest (region labeled PNW in Figs. 4a and 5a), it has been found that isolated extremes occur in the northern Rocky Mountains and Cascades during the spring, and atmospheric rivers cause widespread events west of the Cascades in the fall and winter (Parker and Abatzoglou 2016). This corresponds to our finding of precipitation frequency peaking in fall and winter (Figs. 4a,c), while information transfer I_{frac} is relatively stable through the year. Meanwhile, the dominant time lag τ is longest in summer (Fig. 5d), indicating that longer gaps between rain events during this drier season likely lead to a longer time range of persistence. In general, we detect trends of higher $p(X_t)$, higher I_{frac} , and longer τ in the westernmost region of the Pacific Northwest relative to regions east of the Cascades.

In the Central Plains region of the United States (region labeled CP in Figs. 4a and 5a), more frequent and longer storms tend to occur in the spring and can be attributed to surface warming in the Rocky Mountains, which increases the pressure gradient across the central part of the country (Feng et al. 2016). In this large region, we find that rainfall frequency is highest in spring and summer, information transfer I_{frac} peaks in fall, self-transfer entropy η peaks in winter, and time lag τ is longest in summer and fall. This indicates that while longer storms occur in the spring, their lengths are variable such that our detected τ tends to be short. In other words, the uncertainty in current precipitation $H(X_t)$ is most greatly reduced by the knowledge of very recent states, rather than a typical wet or dry duration. In contrast, the longer τ in summer and fall for some of the Central Plains region indicates a longer term persistence of rainfall.

In the Southeast (region labeled SE in Figs. 4a and 5a), rainfall totals are highest in the summer and regionally important for agriculture (Li and Li 2013). This summertime high rainfall magnitude corresponds to the season of highest rainfall frequency $p(X_t)$ (Fig. 4d). In this region, the predictive skill of daily rainfall has previously been found to depend on the intensity (Li and Li 2013). Specifically, while light and heavy rainfall are related to developing La Niña and North Atlantic sea surface temperatures, respectively, moderate rainfall events are more difficult to predict from a climatic perspective (Moore et al. 2014). We detect few differences between rainfall threshold cases for this region, but we also do not particularly isolate these rainfall magnitudes. We find that self-transfer entropy η peaks in winter in the Southeast, when I_{frac} is

relatively low (Figs. 4a, 5a). The dominant time scale τ is much higher in the spring and summer compared to winter, indicating more long-range dependencies during the wet growing season.

4. Temporal trends in precipitation persistence

As weather patterns shift due to drivers such as climate change and human activities, it is important to address how aspects of precipitation persistence are also changing over time. A trend in precipitation frequency $p(X_t)$ simply indicates a tendency toward more or fewer wet days on average in a season or year (Roque-Malo and Kumar 2017). An increase in information transfer I_{frac} indicates more regularly occurring precipitation patterns that lead to an increase in its overall predictability in terms of information shared between lagged and current states. Similarly, a decrease in I_{frac} indicates more randomly occurring precipitation, or a greater variety of patterns within precipitation sequences such that the current state is less predictable. Associated with a given information transfer, a trend in self-transfer entropy η indicates an altered proportion of information derived from a longer time scale. An increasing η indicates an increased proportion of synergistic information, in that lagged states are becoming stronger joint predictors of current precipitation state, while decreasing η would indicate increased redundancy between multiple lagged states. Finally, an increasing time lag τ would indicate a shift from shorter to longer term persistence. While a short τ is likely associated with highly variable rainy and dry period lengths such that precipitation is not predictable based on long time histories, a longer τ may be associated with a typical combined wet and dry sequence length, as illustrated from the synthetic binary sequences.

a. Annual and seasonal information-based trends

Since any one of I_{frac} , $p(X_t)$, η , and τ could have increasing, decreasing, or nonstatistically significant trends, there are $3^4 = 81$ possible trend combinations. When annual trends are detected for each of the 13 626 individual grid cells in the CPC gauge-based gridded dataset based on 3-yr windows, we find that 64 of these combinations are realized for the 0.3-mm precipitation threshold (Fig. 6a) but some are very rare, representing less than 0.1% of all cases. The most common trend combination for every precipitation threshold is “no trend” in any category, which 19%–27.1% of grid cells exhibit at the different “wet” versus “dry” thresholds (Fig. 6). However, this indicates that a trend in at least one characteristic does occur in about 80% of grid cells.

FIG. 6. Gridded precipitation data categorized by trends in $p(X_t)$, η , τ , and I_{frac} . Trends vary between analyses based on (a) a 0.3-mm threshold for the “wet” vs “dry” demarcation, (b) a 10th percentile threshold, or (c) a 50th percentile threshold for a given grid cell. Numbers indicate the percentage of 13 626 individual grid cells that exhibit any of the 81 possible trends (increasing, decreasing, or no trend for any of the four indicators). Darker colors correspond to higher percentages.

For all thresholds, increasing precipitation frequency $p(X_t)$ is much more common than decreasing $p(X_t)$, and those cells with increasing $p(X_t)$ are more likely to exhibit increasing information transfer I_{frac} than decreasing I_{frac} (Fig. 6). Conversely, cells with decreasing $p(X_t)$ are more likely to be associated with decreases in I_{frac} for lower precipitation thresholds. We detect very few cases of decreasing self-transfer entropy η , and the trend combination of decreasing $p(X_t)$ and decreasing time lag τ is extremely rare for all thresholds. The least common trend is decreasing precipitation frequency $p(X_t)$ paired with decreasing dominant time scale τ and decreasing self-transfer entropy η (Fig. 6). We find that there are no grid cells exhibiting this particular combination of trends for any threshold.

On a national average for the 0.3-mm threshold case, the number of wet days per year has increased by 15.6 days between 1948 and 2018, while information transfer, associated time scale, and self-transfer entropy exhibit weaker trends (Table 1). The ranges of trend magnitudes vary greatly, and for τ and η metrics, many grid cells do exhibit statistically significant trends. The average increasing and decreasing trends in time scale τ are about 4–6 days over the 70-yr period for all precipitation thresholds. This does not necessarily correspond to average wet or dry durations that have altered by this length, but indicates that the typical lengths of combined wet and dry sequences have shifted. We find that trends in precipitation frequency and predictability, $p(X_t)$ and I_{frac} , maintain direction but decrease in

TABLE 1. Average trend magnitudes over CONUS and land areas exhibiting trend directions, based on analysis of 13 626 grid cells of daily precipitation using thresholds of 0.3-mm, 10th percentile, and 50th percentile precipitation.

National average trends	Precipitation threshold of 0.3 mm				Precipitation threshold of 10th percentile				Precipitation threshold of 50th percentile			
	$p(X_t)$	I_{frac}	τ	η	$p(X_t)$	I_{frac}	τ	η	$p(X_t)$	I_{frac}	τ	η
Units (per 70 years)	days	bits per bit	days	bits per bit	days	bits per bit	days	bits per bit	days	bits per bit	days	bits per bit
National average trend	15.6	−0.006	0.072	0.009	12.9	−0.004	0.109	0.011	3.9	0.000	0.25	0.033
Percent area increasing	43	14	14	15	40	15	14	17	26	15	15	25
Average increasing trend	31.7	0.046	4.5	0.077	28.2	0.045	4.6	0.081	15.1	0.042	5.28	0.135
Percent area decreasing	4	28	12	3	4	26	11	2	7	21	11	4
Average decreasing trend	−17.4	−0.047	−4.6	−0.075	−16.8	−0.045	−4.7	−0.075	−12.0	−0.041	−5.48	−0.109

magnitude as the threshold is increased (Table 1). This likely relates to the decrease of the average values of both $p(X_t)$ and I_{frac} when lighter events are set to zero. In contrast, the trends in time lag and self-transfer entropy, η and τ , increase as the threshold is increased (Table 1). This indicates that the characteristics of time scale and relative influences from shorter and longer time lags are changing more for larger precipitation magnitudes. For example, 15% and 25% of grid cells exhibit increasing η for 0.3-mm and 50th percentile cases, respectively, and the magnitude of this increasing trend nearly doubles (Table 1). Based on this, we can infer that for larger precipitation magnitudes, it is becoming increasingly important to take into account more distant histories in order to predict future conditions. Although magnitudes of trends in I_{frac} seem small, at the lowest 0.3-mm threshold, these correspond to shifts in predictability ranging from -50% to 110% for different grid cells.

Spatial analyses of these trends reveal regions that exhibit particular behaviors in different seasons, and can be compared with existing analyses of trends in weather. While direct comparisons with studies that focus on extreme events are difficult, our results do include occurrences of these extreme events as they are embedded within sequences of precipitation with varying magnitudes. Shifting weather patterns may cause similar changes in both extreme events and the entire range of precipitation magnitudes, or may have a disproportionate impact on certain event sizes.

When grid cells are averaged at the HUC4 watershed scale, over half of watersheds show a trend in at least one characteristic during at least one season, whether it is precipitation occurrence, temporal structure in the form of I_{frac} or η , or time lags τ . Increasing and decreasing information transfer I_{frac} are prevalent in the western and eastern United States, respectively (Fig. 7a, Figs. S5, S6). Particularly in the Pacific Northwest region, precipitation sequences are increasingly more predictable in terms of information transfer. For example, of the 47% of grid cells that exhibit increasing I_{frac} in this region, the average magnitude of this increase is 0.06 bits per bit over the 70-yr period, corresponding to a 54% increase. In the eastern and central United States, watersheds with dominantly decreasing information transfer are spread over a large geographic area, with some watersheds exhibiting no trend on average and others exhibiting decreasing trends in I_{frac} of greater than 0.03 bits per bit over the 70-yr period.

At the 0.3-mm threshold, the only region showing a statistically significant annual trend in self-transfer entropy η is the Gulf Coast region in the U.S. Southeast (Fig. 7a). At higher thresholds, however, this increasing η trend is more prevalent over a large area of the eastern

United States (Fig. S6b). This matches with larger τ and η trends for larger precipitation magnitudes. Similarly, the only regions showing annual trends in time lag τ for the 0.3-mm threshold are several relatively mountainous watersheds in the western United States. While most watersheds do not exhibit trends in η or τ , we see from the national averages (Table 1) that up to 29% of the continental U.S. area does show a trend. This could indicate that these trends are more local in nature relative to the size of the HUC4 watershed areas.

While annual trends based on 3-yr nonoverlapping windows show overall patterns throughout the country, a seasonal analysis attributes trends to different times within the year. An annual trend could arise from a trend that occurs over all seasons, or from a strong trend during a single season that dominates the year. Similarly, a finding of “no trend” on an annual scale could result from opposing seasonal trends that cancel out.

In the Northwest, we find that the annual trend of increasing information transfer I_{frac} is mainly due to increasing trends during the winter and spring (Figs. 7b,c). In the summer and fall, the same watersheds show no I_{frac} trends on average (Figs. 7d,e). Meanwhile, precipitation frequency $p(X_t)$ is generally increasing throughout the year, except for the summer for most areas. Seasonally, we also observe instances of increasing or decreasing watershed trends in τ and η that are not detected at the annual scale.

In the eastern United States, information transfer I_{frac} is prevalently decreasing particularly during the spring and summer (Figs. 7g,h), while self-transfer entropy η is increasing in the summer and fall over large areas (Figs. 7h,j). Precipitation frequency $p(X_t)$ is increasing in the Great Lakes and Northeast for most of the year. In this region of the United States, watersheds tend to exhibit either increasing $p(X_t)$ or decreasing I_{frac} , but not both. This indicates that in areas where precipitation frequency is not changing, the level of predictability is decreasing, and in areas where precipitation is becoming more frequent, the predictability is not changing. In the context of the synthetic binary sequences discussed in section 2, some areas are receiving more rain (e.g., a typical rainfall sequence changes from 0011001100 to 0011100111) while others are exhibiting less order (e.g., a typical rainfall sequence changes from a repeating pattern of 0011001100 to a more random pattern of 0010110010 with the same rainfall frequency).

b. Comparison of trends with climatologies

Similarly to the general information-based characteristics, we next consider these trends in different metrics to existing knowledge on changing precipitation and climate in the United States. As discussed

FIG. 7. Statistically significant trends detected in $p(X_t)$, I_{frac} , τ , and η between 1948 and 2018 (a) annually and (b)–(i) for each season and HUC4 watershed, based on 3-yr nonoverlapping windows. Black outlines indicate larger HUC2 basin boundaries. Blue colors indicate statistically significant decreasing trends in $p(X_t)$ (cross hatching), η (circles), τ (triangles), or I_{frac} (background color), and red colors indicate increasing trends. In (a) trends are detected based on probability distributions estimated from nonoverlapping annual 3-yr windows. Boxes in (a) indicate insets for regions of the Pacific Northwest in (b)–(e) and the eastern United States in (f)–(i) for which seasonal trends are detailed.

previously, precipitation frequency $p(X_t)$ is increasing on average for most of the country, particularly in North-Central, West, and Northeast regions. Our finding of an average 15.6 additional wet days over a 70-yr period generally coincides with a previous finding that the national average number of wet days per year has increased by 11 days over a 50-yr period (Roque-Malo and Kumar 2017). Another study has found that mean summer heavy precipitation has increased by 5% for the United States on average, with no change in the

average number of rainy days (Groisman et al. 1999). This contrasts with our finding of increased $p(X_t)$ over much of the country, when all precipitation amounts are considered. This indicates that light to moderate events have increased in occurrence, rather than heavy rain events, and also agrees with our finding of lower trend magnitudes at higher precipitation thresholds.

For certain regions, trends are much stronger than the national averages. For example, the average trend in $p(X_t)$ when only regions with increasing trends are

considered (43% of U.S. land area) is 31.7 days over the study period (Table 1). In both the Pacific Northwest and upper Colorado regions, the average increasing trend in precipitation frequency is greater than 50 days. Similar to the national averages, these trends in precipitation frequencies are less extreme for higher thresholds. In the Midwest, however, a statistically significant increase of 5 days annually of heavy rain in 100 years has been previously estimated (Groisman et al. 1999). This corresponds to our finding of increasing precipitation frequency $p(X_t)$ for the central United States for all threshold cases. Previous studies on extreme precipitation have detected increasing precipitation in the eastern United States, while trends in the West vary from decreasing or insignificant (Mallakpour and Villarini 2017; Yu et al. 2016).

Particularly in the Northeast, it has previously been found that more heavy rain events are occurring during the summer (Hoerling et al. 2016). Correspondingly, it has been found that the increasing extreme events are due to extratropical cyclones (Kunkel et al. 2012), while less than 10% of events are caused by tropical cyclones (Marquardt Collow et al. 2016). While this analysis considers both extreme and nonextreme events together, the increasing trend in $p(X_t)$ in the Northeast in all seasons reflects this increased heavy rainfall. Meanwhile, in the Southeast, fewer heavy rain events are occurring in winter (Hoerling et al. 2016), which corresponds to our detection of no trend or decreasing $p(X_t)$ during the winter and spring for several Southeastern watersheds. In the Central Plains region of the United States, large springtime storms are dominated by mesoscale convective systems that have become more frequent and intense in the past several decades (Feng et al. 2016). In the Southeast, it has been found that events related to tropical cyclones increasingly cause rainfall, but these are still more rare than other causes of rainfall and occur only in the warm season (Moore et al. 2014).

During both summer and winter, it has been found that most nonextreme precipitation occurs as single day events, while extreme precipitation days are embedded within longer durations of rainfall between 2 and 5 days (Agel et al. 2015). Along the East Coast, as the number of these larger events is increasing, an increase in η for our 50th percentile threshold case (Fig. S6) may indicate that the spacing between these single-day precipitation events is becoming more regular.

Trends in precipitation persistence addressed in this study may also relate to shifting meteorological causes of precipitation, and climate indices that influence different regions at different times. For example, a nationally increasing trend in extreme events due to tropical cyclones and fronts has been detected (Kunkel et al. 2012, 2013),

which could impact the temporal ordering of rainfall. In terms of regional climate influences, in Northern California the North Pacific High index is more strongly correlated to rainfall compared to traditional indices such as El Niño–Southern Oscillation (ENSO; Costa-Cabral et al. 2016), such that a combination of climate drivers, in addition to more local changes in land cover could lead to altered rainfall patterns. precipitation characteristics in the Southeast are also influenced by climatic features such as the Atlantic multidecadal oscillation (AMO) and ENSO (Goly and Teegavarapu 2014), such that trends may be related to these. While certain aspects of precipitation may remain constant as climate changes, others may exhibit significant changes. In many areas, greater than predicted increases are attributed to increases in convective intensity or transitions between convective and frontal storms (Ivancic and Shaw 2016).

5. Discussion

Characteristics of precipitation persistence, such as the information theory-based metrics presented here, may have implications for ecosystem function, agriculture, water availability, and downstream impacts. These characteristics and trends also provide a new perspective on the influence of changing climatological patterns on precipitation. Here we review several results of this study, and place them into the context of these potential implications and associated climate patterns.

a. Result: Increasing precipitation frequency

Precipitation frequency has increased over the past 70 years over much of the country, particularly for lighter magnitudes. The finding that the frequency of wet days per year increased by 15.6 days during the 70-yr period based on a 0.3-mm threshold, and only 3.9 days based on a 50th percentile threshold, indicates that most of the increase in precipitation frequency can be associated with light to median precipitation magnitudes. However, even changes in these precipitation amounts can be important to ecosystems. Particularly in the Southwest and other arid regions, increasing demands on water cause any changes in rainfall variability to be particularly important to water availability (Diaz and Anderson 1995). In the context of this study, a change in precipitation frequency $p(X_t)$ may lead to a subsequent change in streamflow variability, evapotranspiration, and groundwater recharge. Additionally, influences of land cover change can manifest as changes in precipitation frequency rather than magnitude (Chen and Dirmeyer 2017). The middle part of the country has experienced the most land cover change since the 1850s in the form of grassland conversion to crops and

continued intensification of agriculture, while the eastern part of the country experienced similar land conversion in the form of deforestation in earlier decades. Land cover changes such as these lead to altered albedo, surface roughness, and heat fluxes, which in turn result in changes in precipitation (Chen and Dirmeyer 2017).

b. Result: Regional trends in predictability

In the western United States, predictability in the form of information transfer I_{frac} is high and has generally increased over the past 70 years. In the eastern United States, predictability is low, and has decreased over the past 70 years. The higher information transfer in the western United States indicates stronger persistence within seasons compared to the rest of the country. Particularly in the Pacific Northwest region, this information transfer has increased over the past 70 years, indicating more regularly occurring precipitation patterns. Second, trends in predictability could translate to altered patterns in flow rates, reservoir levels, and irrigation needs. Although it has been found that changes in precipitation are not necessarily tightly correlated to similar changes in streamflows when soil moisture is not considered (Ivancic and Shaw 2015), it is still likely that altered precipitation variability and persistence do have some nonlinear impact on both soil moisture conditions and downstream flow. For example, an increase in predictability could lead to increased temporal structure in downstream flows, or more predictable timing of water availability. In the eastern United States, information transfer I_{frac} is an order of magnitude lower relative to some regions to the west, and precipitation occurrence is becoming even more highly variable. This divergence between eastern and western regions of the United States, in terms of information transfer and trends in other metrics, could have implications for hydrologic prediction, agricultural management, and natural ecosystem functions. For example, a change in precipitation variability or persistence could lead to “shocks” to vegetation in regions that have not historically experienced the same level of variability.

c. Patterns vary with event sizes

As the precipitation threshold is increased from 0.3-mm to 50th percentile events, trend magnitudes decrease on a national average for precipitation frequency and information transfer. Meanwhile, trend magnitudes increase for time lag and self-transfer entropy associated with maximum information transfer. While decreases in precipitation frequency and information transfer trends are generally associated with lower $p(X_t)$ and I_{frac} values as a “wet” versus “dry” threshold is increased, increasing trend magnitudes for

self-transfer entropy η or time scale τ indicate other types of changes. These trends could influence downstream processes in more complex ways. For example, if there were no changes in $p(X_t)$ or I_{frac} but an increase in τ for a given region, this could influence the amplitude of streamflow, since it may indicate more consecutive rainy days followed by more consecutive dry days. An increasing self-transfer entropy η , such as a shift from uniformly distributed rainy and dry periods to a more ordered repeating sequence, would indicate that a longer time lagged history is becoming more influential relative to the most recent state. In terms of downstream variability, an increased η would indicate that knowledge about longer time-lagged histories is increasingly important, while short-term predictability may have decreased.

6. Conclusions

Through these and other driving mechanisms of connectivity between environmental variables, the trends in information metrics detected here could have varying implications. Here we limit our focus to precipitation, but extensions to variables such as temperature, flows, or evapotranspiration, or attributes such as topography and land cover could further characterize these effects. Similarly, we limit this study to binary precipitation occurrence rather than magnitudes, and only consider magnitudes as a “wet” versus “dry” threshold is changed. With a higher temporal resolution dataset, similar techniques could be used to consider precipitation magnitudes more specifically. Another source of uncertainty in this analysis is the potential for particular stations used in the gridded gauge-based dataset to come into and out of service over the 70-yr period. This feature, along with averaging of multiple stations in heterogeneous regions for interpolation to a single grid cell, could influence trends over time. However, we assume that these impacts are minor given the quality assured precipitation data source used. Overall, this study presents an alternate perspective of precipitation occurrences and persistence in the context of information flow from past to future states, which can augment other studies of hydroclimatology and predictability at various spatial and temporal scales. Particularly as climate and land cover are changing, and human populations exert increasing pressures on water resources, this type of characterization of weather variability is becoming more important.

Acknowledgments. This study was supported by the IML-CZO project NSF EAR 1331906. CPC U.S. Unified precipitation data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their

website at <https://www.esrl.noaa.gov/psd/>. Python codes for this analysis are available at <https://github.com/allisongoodwell/RainingBits>.

REFERENCES

- Agel, L., M. Barlow, J.-H. Qian, F. Colby, E. Douglas, and T. Eichler, 2015: Climatology of daily precipitation and extreme precipitation events in the northeast United States. *J. Hydrometeorol.*, **16**, 2537–2557, <https://doi.org/10.1175/JHM-D-14-0147.1>.
- Anderson, B. T., D. Gianotti, and G. Salvucci, 2015: Characterizing the potential predictability of seasonal, station-based heavy precipitation accumulations and extreme dry spell durations. *J. Hydrometeorol.*, **16**, 843–856, <https://doi.org/10.1175/JHM-D-14-0111.1>.
- , D. J. Gianotti, and G. Salvucci, 2016: Dominant time scales of potentially predictable precipitation variations across the continental United States. *J. Climate*, **29**, 8881–8897, <https://doi.org/10.1175/JCLI-D-15-0635.1>.
- Catto, J. L., 2016: Extratropical cyclone classification and its use. *Rev. Geophys.*, **54**, 486–520, <https://doi.org/10.1002/2016RG000519>.
- , and S. Pfahl, 2013: The importance of fronts for extreme precipitation. *J. Geophys. Res. Atmos.*, **118**, 10 791–10 801, <https://doi.org/10.1002/jgrd.50852>.
- Chen, L., and P. A. Dirmeyer, 2017: Impacts of land-use/land-cover change on afternoon precipitation over North America. *J. Climate*, **30**, 2121–2140, <https://doi.org/10.1175/JCLI-D-16-0589.1>.
- Chen, M., W. Shi, P. Xie, V. B. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, <https://doi.org/10.1029/2007JD009132>.
- Chin, E. H., 1977: Modeling daily precipitation occurrence process with Markov chain. *Water Resour. Res.*, **13**, 949–956, <https://doi.org/10.1029/WR013i006p00949>.
- Costa-Cabral, M., J. S. Rath, W. B. Mills, S. B. Roy, P. D. Bromirski, and C. Milesi, 2016: Projecting and forecasting winter precipitation extremes and meteorological drought in California using the North Pacific high sea level pressure anomaly. *J. Climate*, **29**, 5009–5026, <https://doi.org/10.1175/JCLI-D-15-0525.1>.
- Cover, T. M., and J. A. Thomas, 2006: *Elements of Information Theory*. 2nd ed. Wiley, 776 pp.
- Dhanya, C. T., and G. Villarini, 2018: On the inherent predictability of precipitation across the United States. *Theor. Appl. Climatol.*, **133**, 1035–1050, <https://doi.org/10.1007/s00704-017-2231-5>.
- Diaz, H. F., and C. A. Anderson, 1995: Precipitation trends and water consumption related to population in the southwestern United States: A reassessment. *Water Resour. Res.*, **31**, 713–720, <https://doi.org/10.1029/94WR02755>.
- Feng, Z., L. R. Leung, S. Hagos, R. A. Houze, C. D. Burleyson, and K. Balaguru, 2016: More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. *Nat. Commun.*, **7**, 13429, <https://doi.org/10.1038/ncomms13429>.
- Foufoula-Georgiou, E., and D. P. Lettenmaier, 1987: A Markov renewal model for rainfall occurrences. *Water Resour. Res.*, **23**, 875–884, <https://doi.org/10.1029/WR023i005p00875>.
- Gabriel, K., and J. Neumann, 1961: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90–95, <https://doi.org/10.1002/qj.49708837511>.
- Gianotti, D. J. S., B. T. Anderson, and G. D. Salvucci, 2014: The potential predictability of precipitation occurrence, intensity, and seasonal totals over the continental United States. *J. Climate*, **27**, 6904–6918, <https://doi.org/10.1175/JCLI-D-13-00695.1>.
- Goly, A., and R. S. V. Teegavarapu, 2014: Individual and coupled influences of AMO and ENSO on regional precipitation characteristics and extremes. *Water Resour. Res.*, **50**, 4686–4709, <https://doi.org/10.1002/2013WR014540>.
- Goodwell, A. E., and P. Kumar, 2015: Information theoretic measures to infer feedback dynamics in coupled logistic networks. *Entropy*, **17**, 7468–7492, <https://doi.org/10.3390/e17117468>.
- , and —, 2017: Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resour. Res.*, **53**, 5920–5942, <https://doi.org/10.1002/2016WR020216>.
- Griffith, V., and T. Ho, 2015: Quantifying redundant information in predicting a target random variable. *Entropy*, **17**, 4644–4653, <https://doi.org/10.3390/e17074644>.
- Groisman, P. Ya., and Coauthors, 1999: Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climatic Change*, **42**, 243–283, <https://doi.org/10.1023/A:1005432803188>.
- Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier, 2005: Effects of temperature and precipitation variability on snowpack trends in the western United States. *J. Climate*, **18**, 4545–4561, <https://doi.org/10.1175/JCLI3538.1>.
- Hay, L. E., G. J. McCabe, D. M. Wolock, and M. A. Ayers, 1991: Simulation of precipitation by weather type analysis. *Water Resour. Res.*, **27**, 493–501, <https://doi.org/10.1029/90WR02650>.
- Hoerling, M., J. Eischeid, J. Perlwitz, X. W. Quan, K. Wolter, and L. Cheng, 2016: Characterizing recent trends in U.S. heavy precipitation. *J. Climate*, **29**, 2313–2332, <https://doi.org/10.1175/JCLI-D-15-0441.1>.
- Ince, R. A. A., 2017: Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, **19**, 318, <https://doi.org/10.3390/e19070318>.
- Ivancic, T. J., and S. B. Shaw, 2015: Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change*, **133**, 681–693, <https://doi.org/10.1007/s10584-015-1476-1>.
- , and —, 2016: A U.S.-based analysis of the ability of the Clausius-Clapeyron relationship to explain changes in extreme rainfall with changing temperature. *J. Geophys. Res. Atmos.*, **121**, 3066–3078, <https://doi.org/10.1002/2015JD024288>.
- Jiang, M., B. S. Felzer, and D. Sahagian, 2016a: Characterizing predictability of precipitation means and extremes over the conterminous United States, 1949–2010. *J. Climate*, **29**, 2621–2632, <https://doi.org/10.1175/JCLI-D-15-0560.1>.
- , —, and —, 2016b: Predictability of precipitation over the conterminous U.S. based on the CMIP5 multi-model ensemble. *Nat. Sci. Rep.*, 29962, <https://doi.org/10.1038/srep29962>.
- Knapp, A. K., and Coauthors, 2008: Consequences of more extreme precipitation regimes for terrestrial ecosystems. *BioScience*, **58**, 811–821, <https://doi.org/10.1641/B580908>.
- Koster, R. D., and Coauthors, 2017: Hydroclimatic variability and predictability: A survey of recent research. *Hydrol. Earth Syst. Sci.*, **21**, 3777–3798, <https://doi.org/10.5194/hess-21-3777-2017>.
- Kumar, P., 2013: Seasonal rain changes. *Nat. Climate Change*, **3**, 783–784, <https://doi.org/10.1038/nclimate1996>.
- Kunkel, K. E., D. R. Easterling, D. A. R. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological causes of

- the secular variations in observed extreme precipitation events for the conterminous United States. *J. Hydrometeor.*, **13**, 1131–1141, <https://doi.org/10.1175/JHM-D-11-0108.1>.
- , and Coauthors, 2013: Monitoring and understanding trends in extreme storms. *Bull. Amer. Meteor. Soc.*, **94**, 499–514, <https://doi.org/10.1175/BAMS-D-11-00262.1>.
- Li, L., and W. Li, 2013: Southeastern United States summer rainfall framework and its implication for seasonal prediction. *Environ. Res. Lett.*, **8**, 044017, <https://doi.org/10.1088/1748-9326/8/4/044017>.
- Mahoney, K., F. M. Ralph, K. Wolter, N. Noeskin, M. Dettinger, D. Gottas, T. Coleman, and A. White, 2015: Climatology of extreme daily precipitation in Colorado and its diverse spatial and seasonal variability. *J. Hydrometeor.*, **16**, 781–792, <https://doi.org/10.1175/JHM-D-14-0112.1>.
- Mallakpour, I., and G. Villarini, 2017: Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. *Theor. Appl. Climatol.*, **130**, 345–363, <https://doi.org/10.1007/s00704-016-1881-z>.
- Marquardt Collow, A. B. M., M. G. Bosilovich, and R. D. Koster, 2016: Large-scale influences on summertime extreme precipitation in the northeastern United States. *J. Hydrometeor.*, **17**, 3045–3061, <https://doi.org/10.1175/JHM-D-16-0091.1>.
- Moore, B. J., K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. Hamill, 2014: Climatology and environmental characteristics of extreme precipitation events in the Southeastern United States. *Mon. Wea. Rev.*, **143**, 718–741, <https://doi.org/10.1175/MWR-D-14-00065.1>.
- Muhammad, N. S., P. Y. Julien, and J. D. Salas, 2016: Probability structure and return period of multiday monsoon rainfall. *J. Hydrol. Eng.*, **21**, 04015048, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001253](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001253).
- Parker, L. E., and J. T. Abatzoglou, 2016: Spatial coherence of extreme precipitation events in the Northwestern United States. *Int. J. Climatol.*, **36**, 2451–2460, <https://doi.org/10.1002/joc.4504>.
- Pierce, J., 1980: *An Introduction to Information Theory: Symbols, Signals, and Noise*. Dover Publications, 336 pp.
- Rodriguez-Iturbe, I., 2000: Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamics. *Water Resour. Res.*, **36**, 3–9, <https://doi.org/10.1029/1999WR900210>.
- Roque-Malo, S., and P. Kumar, 2017: Patterns of change in high frequency precipitation variability over North America. *Nat. Sci. Rep.*, **7**, 10853, <https://doi.org/10.1038/s41598-017-10827-8>.
- Ruddell, B. L., and P. Kumar, 2009: Ecohydrologic process networks: 1. Identification. *Water Resour. Res.*, **45**, W03419, <https://doi.org/10.1029/2008WR007279>.
- Schreiber, T., 2000: Measuring information transfer. *Phys. Rev. Lett.*, **85**, 461, <https://doi.org/10.1103/PhysRevLett.85.461>.
- Seaber, P. R., F. P. Kapinos, and G. L. Knapp, 1987: Hydrologic unit maps. USGS Water-Supply Paper 2294, 66 pp., http://pubs.usgs.gov/wsp/wsp2294/pdf/wsp_2294.pdf.
- Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's Tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, <https://doi.org/10.1080/01621459.1968.10480934>.
- Shannon, C., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Williams, P. L., and R. D. Beer, 2010: Nonnegative decomposition of multivariate information. arXiv, 14 pp., <https://arxiv.org/abs/1004.2515>.
- Wu, Z., P. Dijkstra, G. W. Koch, J. Peñuelas, and B. A. Hungate, 2011: Responses of terrestrial ecosystems to temperature and precipitation change: A meta-analysis of experimental manipulation. *Global Change Biol.*, **17**, 927–942, <https://doi.org/10.1111/j.1365-2486.2010.02302.x>.
- Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeor.*, **8**, 607–626, <https://doi.org/10.1175/JHM583.1>.
- Yu, L., S. Zhong, L. Pei, X. Bian, and W. E. Heilman, 2016: Contribution of large-scale circulation anomalies to changes in extreme precipitation frequency in the United States. *Environ. Res. Lett.*, **11**, 044003, <https://doi.org/10.1088/1748-9326/11/4/044003>.