"It's Raining Bits": Patterns in Directional Precipitation Persistence across the United States

ALLISON E. GOODWELL

Civil Engineering, University of Colorado Denver, Denver, Colorado

(Manuscript received 5 June 2020, in final form 29 September 2020)

ABSTRACT: The spatial and temporal ordering of precipitation occurrence impacts ecosystems, streamflow, and water availability. For example, both large-scale climate patterns and local landscapes drive weather events, and the typical speeds and directions of these events moving across a basin dictate the timing of flows at its outlet. We address the predictability of precipitation occurrence at a given location, based on the knowledge of past precipitation at surrounding locations. We identify "dominant directions of precipitation influence" across the continental United States based on a gridded daily dataset. Specifically, we apply information theory–based measures that characterize dominant directions and strengths of spatial and temporal precipitation dependencies. On a national average, this dominant direction agrees with the prevalent direction of weather movement from west to east across the country, but regional differences reflect topographic divides, precipitation gradients, and different climatic drivers of precipitation. Trends in these information relationships and their correlations with climate indices over the past 70 years also show seasonal and spatial divides. This study expands upon a framework of information-based predictability to answer questions about spatial connectivity in addition to temporal persistence. The methods presented here are generally useful to understand many aspects of weather and climate variability.

KEYWORDS: North America; Precipitation; Climate variability; Seasonal variability; Trends

1. Introduction

The focus of this paper is on the following question:

When looking for rainfall
which way is best:
Should I look north, south, east or west?
Or should I, say,
look multiple ways
to make a much better guess?

In prose, we will explore the predictability of precipitation at one location, given information about past precipitation at one or more surrounding locations. At your current location, the knowledge of "Did it rain here yesterday?" is likely somewhat predictive of today's precipitation state due to the temporal persistence of precipitation (Chin 1977; Gabriel and Neumann 1961; Hay et al. 1991; Roque-Malo and Kumar 2017; Goodwell and Kumar 2019). However, the knowledge of past states at neighboring or even distant locations could improve upon this predictability. The temporal persistence of precipitation at a single location relates to typical durations of wet or dry periods, and adding a spatial aspect of persistence relates to how storms move across a landscape. Climate conditions, topography, and regional land–atmosphere feedbacks drive these aspects of temporal persistence and spatial synchronicity of precipitation, which in turn influence soil moisture, flows, and vegetation. For example, the direction, speed, and size of a storm event moving across a basin can impact downstream flows and ecohydrologic processes. Goodwell and Kumar (2019) explored temporal precipitation persistence and predictability, addressing the extent to which the knowledge of past precipitation informs future states, and the associated time delays. This study extends this work to address the predictability of precipitation given not only past states at the location of interest, but at neighboring locations as well.

It is important to understand both spatial and temporal features of precipitation persistence from the perspectives of climate change, weather variability, and the water cycle. For example, Kunkel et al. (2012) distinguishes regions of the continental United States in terms of dominant meteorological causes of precipitation. Changes in these causes would likely lead to shifts in storm lengths and movement. In a study of South American rainfall, Boers et al. (2014) finds connections between the synchronization of extreme precipitation and large-scale climate patterns, which constitute linkages between climate and storm movement. Additionally, precipitation patterns vary based on landscape properties, such as eco-regions that show different trends in precipitation durations and frequencies (Roque-Malo and Kumar 2017).

Influences of global climate on regional precipitation have been studied in the context of climate indices, or patterns of pressure gradients and sea surface temperatures, that are

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-20-0134.s1.

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

DOI: 10.1175/JHM-D-20-0134.1

linked to weather patterns in the United States. For example, different phases of the North Pacific index (NP) are associated with anticyclonic and cyclonic circulation in the U.S. West and Southeast, respectively (NOAA 2019c). In the Great Plains in the central United States, precipitation fluctuations are subject to both tropical and northern Pacific sea surface temperatures at different time scales (Ting and Wang 1997). The Pacific decadal oscillation (PDO) has been found to influence winter air temperatures and precipitation in North America (NOAA 2019d; Mantua et al. 1997), and multiple Pacific climate indices have been linked to long-term U.S. droughts during the warm season (Barlow et al. 2001). Meanwhile, the Atlantic multidecadal oscillation (AMO) is associated with summer climate and Atlantic hurricane patterns (NOAA 2019a), and indices related to ENSO are linked to severe weather patterns throughout the United States (Gershunov and Barnett 1998).

In general, changes in phases or amplitudes of different modes of climate variability under natural and human drivers can lead to altered precipitation patterns with further impacts downstream (Hurrell and Deser 2010). For examples, the PDO has been studied in the context of shifting salmon production in the Pacific Northwest (Mantua et al. 1997), and a strongly positive east Pacific–North Pacific (EPNP) index was associated with extreme 1993 flooding in the U.S. Midwest (NOAA 2019b; Bell and Janowiak 1995). Regardless of associations with climate or landscape drivers, the movement of storm events themselves drive downstream hydrological processes. Rainstorm movement has been studied in terms of its influence on peak streamflows at an outlet in natural, urban, and modeled systems (Singh 1997; Lee and Huang 2007; Seo et al. 2012; Volpi et al. 2013; Seo and Schmidt 2013; Kim and Seo 2013; Seo and Schmidt 2014; Gao and Fang 2018). It has been found that storm timing, size, and duration can be an important factor in flow response, especially when considered in the context of the orientation of the drainage network (Seo et al. 2012). Particularly, it has been found that the cumulative mean areal precipitation is more informative to peak discharge than precipitation at a given point, and the dependency increases with the spatial extent of the rainfall (Gao and Fang 2018). In urban regions, the design of drainage networks could be better informed based on not only typical event sizes, but durations and directions of moving rainstorms (Seo and Schmidt 2013). These findings indicate the importance of understanding how climate patterns influence weather, and how persistence and spatial dependencies impact downstream processes.

Here we study the predictability of precipitation from an information theory–based perspective, in which predictability is a reduction in uncertainty about precipitation occurrence due to the knowledge of past precipitation in one or more surrounding areas. Specifically, we use several information theory–based metrics to quantitatively address questions about precipitation persistence and directionality within the continental United States. Based on a daily gridded precipitation dataset, we answer the following questions:

1) From what direction, or combination of directions, is precipitation best predicted at a given location?
2) How strong is this spatial component of predictability for different regions?
3) How do aspects of directionality and level of precipitation predictability vary with season, climate condition, and over the historic record?

These questions are related to the predictability of precipitation based on the knowledge of both spatial and temporal variability, and are relevant to weather prediction and water resources. We hypothesized that this information-based spatial predictability varies seasonally and regionally across the continental United States, and in certain areas is associated with climate indices. For example, Goodwell and Kumar (2019) found that predictability due to the knowledge of lagged, or past, precipitation states has increased over much of the western United States, and decreased over much of the east. This study further addresses the directions linked to this predictability, along with multivariate measures and associated climate drivers.

The paper is organized as follows. In section 2, we introduce our information theory–based methodology and application to the Climate Prediction Center (CPC) Unified gridded gauge-based precipitation dataset (Chen et al. 2008). We include an illustrative example of how information measures, which quantify predictability as reductions in uncertainty, vary under different categories of spatial variability. In section 3, we present results for nationally averaged and regional dominant directions and strengths associated with predictability as well as trend analyses of these measures and correlations with several climate indices. In sections 4 and 5 we provide a discussion and conclusions.

2. Methods

We consider $X_{t,c}$ to be a binary variable describing daily precipitation occurrence at a "central" location c, where $X_{t,c} = 1$ on day t if precipitation is observed above a specified threshold, and $X_{t,c} = 0$ otherwise. A given time series has attributes $p_1 = p(x_{t,c} = 1)$ and $p_0 = 1 - p_1$, denoting the probability, or relative frequency, of wet or dry days, respectively. Shannon entropy, $H(X_{t,c}) = -\Sigma p(x_{t,c}) \log_2 p(x_{t,c})$, where the summation is over all possible states of $x_{t,c}$, is a measure of uncertainty of a random variable (Shannon 1948; Cover and Thomas 2006) and is measured in units of bits. This formulation of entropy can be interpreted as the average or expected number of (yes or no) questions that need to be asked to determine a value of $X_{t,c}$. Since it is a binary variable, $H(X_{t,c}) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$, and $H(X_{t,c})$ is always less than or equal to 1 (Fig. 1a). Values close to 1 indicate that precipitation is maximally *uncertain*, since it is nearly equally likely to be "wet" or "dry" on a given day. In contrast, a low value of $H(X_{t,c})$ indicates that the precipitation state is nearly always dry (or wet), such that it is less uncertain (Fig. 1a). This uncertainty of precipitation at a given location, based on the probability distribution derived from binary daily occurrence data, provides the amount of information (in bits) needed for an accurate prediction. This information, or reduction in uncertainty, could come from knowledge about historical states of precipitation at that

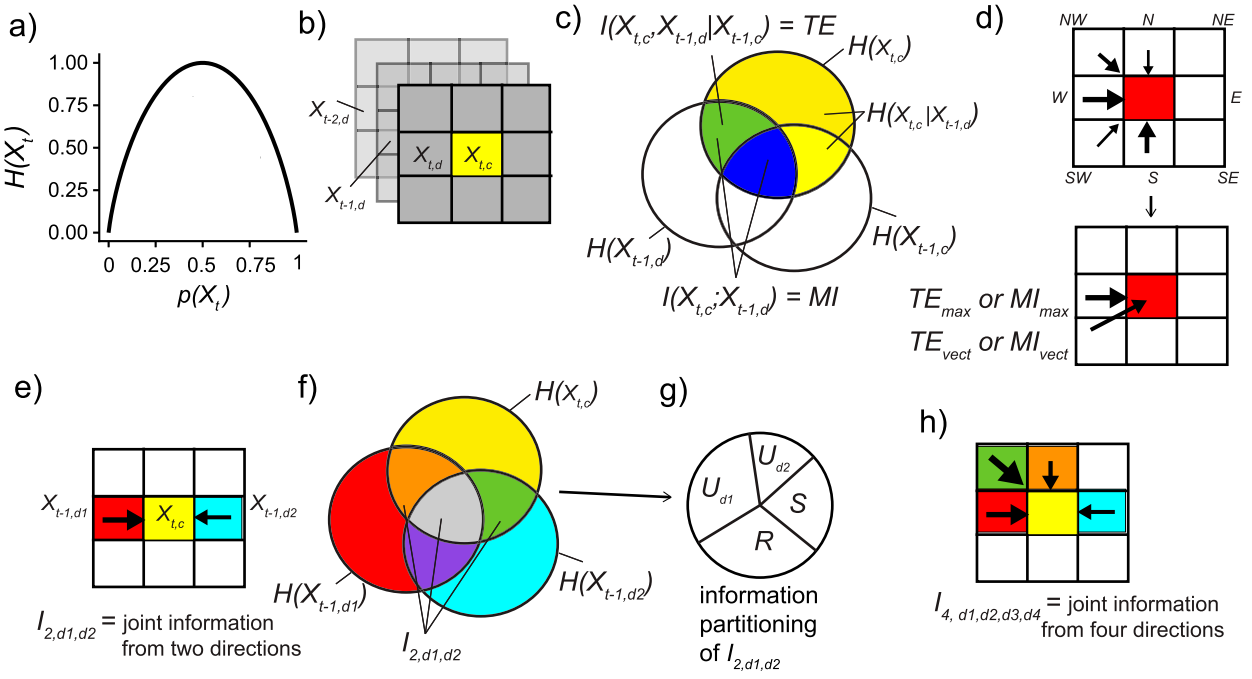

FIG. 1. Illustration of information theory–based methods for spatiotemporal predictability of precipitation. (a) Shannon entropy $H(X_t)$ for a binary variable is a measure of uncertainty that varies from 0 to 1 bit as a function of $p(X_t) \equiv p(X_t = 1)$. (b) We consider a precipitation time series $X_{t,c}$ at a central location, where the uncertainty of $X_{t,c}$ is reduced given knowledge about lagged histories of neighboring areas, e.g., $X_{t-1,d}$. (c) Venn diagram illustrating entropy, mutual information (MI), and transfer entropy (TE). Overlapping blue and green regions indicate reductions in uncertainty, $H(X_{t,c})$, given a lagged state at a neighboring cell. (d) TE and MI are computed for each of eight directions, and the maximum values and the vector resolved values are computed as MI_{max} (or TE_{max}) and MI_{vect} (or TE_{vect}), respectively, as described in the text. (e)–(g) Mutual information between the central target cell and two lagged histories in opposite directions, $I_{2,d1d2}$ (here illustrated as $I_{2,EW}$) is partitioned into unique (U_{d1}, U_{d2}), redundant (R), and synergistic (S) components. (h) Mutual information between the central target cell and four surrounding lagged histories, I_4, is a higher-dimensional measure of precipitation predictability.

location, precipitation at other locations, or other climate-related features such as wind, humidity, or clouds. We focus on the information that can be obtained from knowledge of the past precipitation state at the central location c and neighboring locations. This focus on past precipitation rather than multivariate drivers enables a relatively simple analysis of a single gridded dataset. Moreover, we assume that past precipitation somewhat integrates these other drivers, as it directly captures the duration and movement of events. Depending on a storm's size, shape, speed, and direction of movement, precipitation states in multiple directions can provide information that reduces the uncertainty at a central location (Fig. 1b).

a. Dominant directionality of precipitation

The mutual information between the current precipitation state at location c, $X_{t,c}$, and a lagged state at a neighboring grid cell at location d, $X_{t-1,d}$ is

$$I(X_{t-1,d};X_{t,c}) = \sum p(x_{t-1,d},x_{t,c}) \log_2 \left[\frac{p(x_{t-1,d},x_{t,c})}{p(x_{t-1,d})p(x_{t,c})} \right]$$

$$= H(X_{t,c}) - H(X_{t,c}|X_{t-1,d}), \quad (1)$$

where the summation is over all possible states of $X_{t-1,d}$ and $X_{t,c}$. In the context of daily precipitation, this measure is the reduction in precipitation uncertainty that we obtain from knowing yesterday's state at a neighboring location d. In Eq. (1), the term $H(X_{t,c}|X_{t-1,d})$ is conditional entropy, which is the remaining uncertainty in X_t when the variable $X_{t-1,d}$ is known (Fig. 1c). Hereafter, we refer to this mutual information between one central "target" and one neighboring "source" variable as MI.

While MI quantifies a lagged relationship between precipitation at two locations, we may further inquire about the relevance of this relationship if we also know yesterday's precipitation state at the central location c. For example, if we could have obtained the same or a larger reduction in uncertainty based on knowing $X_{t-1,c}$, the information provided by $X_{t-1,d}$ may be somewhat "redundant" or unnecessary. We address this with transfer entropy (TE), which is a version of conditional mutual information defined here for our specific case as follows:

$$\mathrm{TE}(X_{t-1,d};X_{t,c}|X_{t-1,c}) = \sum p(x_{t-1,d},x_{t,c},x_{t-1,c}) \log_2 \left[\frac{p(x_{t,c}|x_{t-1,d},x_{t,c})}{p(x_{t,c}|x_{t-1,c})} \right]. \quad (2)$$

This TE is a simplification of the initial formulation by Schreiber (2000), which involves multiple time-lagged histories.

Versions of TE used in other studies also detect TE for time lags other than 1 (Ruddell and Kumar 2009; Goodwell and Kumar 2015; Sendrowski and Passalacqua 2017). However, the version defined in Eq. (2) captures information about precipitation at a central location that is provided by a neighboring cell, conditioned on also knowing the lagged state at the central location. In other words, TE can be interpreted as the information that a neighboring location provides beyond that already provided due to the knowledge of past precipitation at the location of interest. For example, if rainfall tends to be synchronous between two neighboring grid cells, rainfall at one grid cell does not predict future rainfall at the next grid cell beyond the predictability that exists due to local knowledge.

We compute MI and TE between precipitation at a central grid cell and its eight nearest neighbors, such that there are two measures of information transfer associated with each of the eight directions [north, northeast, east, southeast, south, southwest, west, and northwest (N, NE, E, SE, S, SW, W, and NW, respectively)]. To obtain a dominant directionality of information transfer at a given location, we compare two methods (Fig. 1d):

1) $\mathrm{MI}_{\max}$ ($\mathrm{TE}_{\max}$) and θ_{MImax} (θ_{TEmax}) are the magnitude and angular direction, respectively, associated with the maximum of the eight values from all neighboring directions.
2) $\mathrm{MI}_{\mathrm{vect}}$ ($\mathrm{TE}_{\mathrm{vect}}$) and θ_{MIvect} (θ_{TEvect}) are the magnitude and direction, respectively, based on vector resolution of the eight neighboring values [see "Information Theory Methods" and Eqs. (1)–(4) in the online supplemental material]. Here, for example, two equal information flows from opposite directions (e.g., east and west) cancel out, while information components from similar directions (e.g., west and southwest) build up.

We employ both of these methods for MI and TE, resulting in four estimates of "dominant direction" and "dominant predictability" for each grid cell. Similarities or differences between measures indicate particular features of directional information flow that relate to storm directions, sizes, and speeds. For example, consider a case where θ_{TEmax} and θ_{TEvect} are very different at some location. This indicates that the single dominant direction based on $\mathrm{TE}_{\max}$ is mitigated by similar information transfer from opposing directions, such that the dominant influence based on $\mathrm{TE}_{\mathrm{vect}}$ is associated with a different direction. Similarly, a large difference between MI-based and TE-based dominant strengths and directions would indicate that knowing the lagged history of the central target location plays a significant role in predictability.

b. *Joint predictability from two directions*

Next we consider higher-dimensional information theory–based measures to determine the total amount of uncertainty that can be reduced when lagged states in multiple directions are known together, rather than individually. This contrasts with the previously described methods, which only assume the knowledge of past precipitation in one neighboring direction at a time. As a first step that limits analyses to 3D probability functions, we consider the total mutual information between the central current state $X_{t,c}$ and two lagged histories from neighboring locations, $X_{t-1,\mathrm{d1}}$ and $X_{t-1,\mathrm{d2}}$, where d1 and d2 indicate two of the eight possible neighboring directions (Figs. 1e,f). We define this bidirectional mutual information as follows:

$$I_{2,\mathrm{d1d2}} = I(X_{t-1,\mathrm{d1}}, X_{t-1,\mathrm{d2}}; X_{t,c}) = \sum p(x_{t-1,\mathrm{d1}}, x_{t-1,\mathrm{d2}}, x_{t,c}) \log_2 \left[\frac{p(x_{t-1,\mathrm{d1}}, x_{t-1,\mathrm{d2}}, x_{t,c})}{p(x_{t-1,\mathrm{d1}}, x_{t-1,\mathrm{d2}}) p(x_{t,c})} \right]. \quad (3)$$

For this analysis, we choose the pair (d1, d2) to be opposing cardinal directions, i.e., (d1 = N, d2 = S) or (d1 = E, d2 = W). This results in two measures for a given central location c of $I_{2,\mathrm{NS}}$ and $I_{2,\mathrm{EW}}$. These indicate a level of precipitation predictability given knowledge of two lagged states in opposite directions. In an information decomposition (Williams and Beer 2010), $I_{2,\mathrm{d1d2}}$ can be partitioned into unique, synergistic, and redundant components as follows:

$$I_{2,\mathrm{d1d2}} = U_{\mathrm{d1}} + U_{\mathrm{d2}} + S + R. \quad (4)$$

Here, U_{d1} and U_{d2} are information that is shared between each neighboring state and the central state individually, R is "overlapping," or redundant, information that the knowledge of both states provide, and S is "joint" or synergistic information that only arises when both neighboring states are known together (Fig. 1g). Existing information theory measures can be explained as combinations of these four components. For example, TE is composed of one unique component and one synergistic component of information. In a precipitation predictability context, TE is the information that a lagged source provides to a current target individually and/or synergistically when both the lagged source and the lagged target are known together (Fig. 1c). In contrast, MI is composed of the same unique component, but contains a redundant component instead of a synergistic one. In other words, MI is the information that a source provides to a target individually, and/or in overlap when the lagged central state is also known. Since information theory does not provide a method to compute R, U, and S directly, we use a measure proposed in Goodwell and Kumar (2017), which estimates R as a function of the mutual information between the two sources. The terms U_{d1}, U_{d2}, and S are then solved from existing information theoretic equations. In this formulation, independent sources [where $I(X_{t-1,\mathrm{d1}}; X_{t-1,\mathrm{d2}}) = 0$] are minimally redundant, while strongly related sources are maximally redundant ("Information Theory Methods" in the supplemental material).

To illustrate this information partitioning into unique, redundant, and synergistic components, here we present four synthetic cases in which precipitation at a central cell $X_{t,c}$ is a simple function of lagged precipitation from either the west, east, or both (Fig. 2). For all cases, precipitation in the westernmost grid cell is a first-order Markov Chain, with a transition probability (probability of a switch from wet to dry or dry to wet) of 0.1 ("Example Case" in

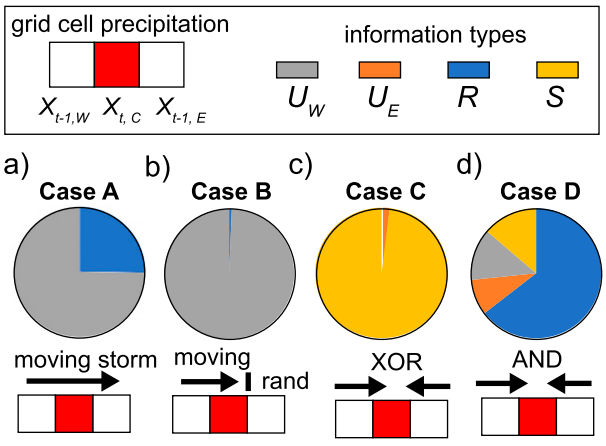

FIG. 2. Four synthetic cases of precipitation patterns with different implications for predictability, as measured by synergistic S, unique U, and redundant R information types. As described in the text, (a),(b) cases A and B illustrate a moving storm from west to east, while (c),(d) cases C and D are cases in which precipitation at a central cell is a function of lagged precipitation in two directions. Pie charts indicate the proportions of shared information, $I_{2,\mathrm{d1d2}}$, attributed to S, U, and R components.

the supplemental material). This introduces persistence in synthetic precipitation, in that dry and wet periods tend to clump together, but overall the relative frequencies of wet and dry days are equal ($p_0 = p_1 = 0.5$), such that entropy $H(X_{t,c}) =$ 1 bit. In all these cases, the uncertainty of the central precipitation state is completely reduced given the knowledge of lagged precipitation in the eastern and western directions.

In case A, precipitation moves from west to east at a 1-day lag. In other words, if it rains on Monday toward the west, it rains on Tuesday and Wednesday in the central and eastern grid cells, respectively. For this case, we find that most information about $X_{t,c}$ is uniquely provided by the knowledge of $X_{t-1,\mathrm{W}}$, and some information is provided redundantly by both $X_{t-1,\mathrm{W}}$ and $X_{t-1,\mathrm{E}}$ due to the persistence in precipitation (Fig. 2a).

In case B, precipitation moves from the west to the central cell, but precipitation at the grid cell to the east is disconnected. Here, all information provided about $X_{t,c}$ comes uniquely from the knowledge of $X_{t-1,\mathrm{W}}$, since $X_{t-1,\mathrm{E}}$ is not informative. This could illustrate a region where a topographical divide dictates storm movement, such that a given grid cell is only influenced by precipitation in one direction (Fig. 2b).

In case C, precipitation at the central grid cell obeys an exclusive "or" (XOR) rule, where precipitation occurs only if the neighboring grid cells were in different states the day before. In this case, precipitation to the east and west are independent of each other, and provide information to the central cell only synergistically. For example, if we knew it rained yesterday toward the west, this does not reduce our uncertainty about $X_{t,c}$ unless we also know yesterday's precipitation state toward the east (Fig. 2c).

In case D, precipitation in the east and west are correlated, and precipitation at the central cell follows an "and" rule where precipitation only occurs if yesterday's state was wet in both directions. This would represent a case where precipitation is somewhat synchronized across the region. Here, we see that all four types of information contribute to the predictability of precipitation, but redundant information dominates since the source regions are correlated (Fig. 2d).

These examples illustrate how different patterns in precipitation movement relate to information measures, but also shows how $I_{2,\mathrm{d1d2}}$ and subsequent partitioning reveals a component of predictability that is missing from MI and TE. For example, in case C, TE and MI would detect that precipitation is completely unpredictable, since the knowledge of a single neighboring lagged state does not reduce any uncertainty in the target cell. Case D provides an example in which two neighboring states are nearly equally informative to $X_{t,c}$, such that they would cancel out in a vector resolution and different dominant directions would be detected for θ_{MIvect} and θ_{MImax}. We refer to the online supplemental material ("Example Case" and Tables S1 and S2) for more details on this example. In general, information partitioning measures of synergy, uniqueness, and redundancy capture the reduction in precipitation uncertainty that is obtained by "looking both ways" simultaneously.

c. *Multivariate predictability*

The previously discussed measures consider precipitation predictability as a function of either one or two neighboring lagged sources. As such, they all fail to detect any additional predictability that could be obtained from knowing more than two sources simultaneously. For this, we introduce a multivariate lagged mutual information measure, I_4, which is the total amount of information that several surrounding locations provide about the current precipitation state. Specifically, I_4 is the information shared between the current state at a central grid cell and the strongest four individual neighboring lagged sources, defined as follows (Fig. 1h):

$$I_4 = I(X_{t-1,\mathrm{d1}}, X_{t-1,\mathrm{d2}}, X_{t-1,\mathrm{d3}}, X_{t-1,\mathrm{d4}}; X_{t,c}), \tag{5}$$

where d1–d4 represent the four neighboring directions that individually provide the maximum amount of information about precipitation at the center grid cell, c, based on TE [Eq. (2)]. For example, d1 corresponds to θ_{TEmax}, and the other three directions (d2, d3, d4) are the neighboring cells with the next highest TE values. Theoretically, we could also compute I_8, which would be the total information transfer from all eight neighboring directions. While this would be a more comprehensive measure of precipitation predictability based on the knowledge of all neighboring states, the dimensionality of the associated probability distribution becomes high relative to the size of the data. For example, for a single season with 90 days, MI involves a 2D binary pdf with $2^2 = 4$ bins. Similarly, TE and I_2 measures involve 3D binary pdfs with $2^3 = 8$ bins, and I_4 involves a 5D binary pdf with $2^5 = 32$ bins. A higher-order measure such as I_8 would have a pdf of size $2^8 = 256$ bins, which would require many years of daily data to obtain a robust result.

For any given measure, statistical significance is estimated using the shuffled surrogates method (Ruddell and Kumar 2009), in which source variables (i.e., lagged histories) are shuffled randomly in time to destroy temporal correlations. This is done 100 times, to obtain a set of

TABLE 1. Summary of information theory–based measures.

Measure(s)	Description	Equation
MI_{max}, TE_{max}	Maximum mutual information or transfer entropy between current precipitation state at location c and lagged precipitation in a neighboring direction (unit of bits)	(1), (2)
θ_{MImax}, θ_{TEmax}	Dominant neighboring direction associated with MI_{max} or TE_{max} (units of degrees)	
MI_{vect}, TE_{vect}	Weighted average of mutual information or transfer entropy between current precipitation state at location c and lagged precipitation in neighboring directions (unit of bits)	(1), (2)
θ_{MIvect}, θ_{TEvect}	Average direction associated with MI_{vect} or TE_{vect}, based on vector resolution of eight values of mutual information (units of degrees)	
$I_{2,E,W}$, $I_{2,N,S}$	Information provided to current precipitation at location c from two opposite directions (units of bits)	(3)
U_N, U_S, U_E, U_W	Unique information transfers from neighboring directions to current precipitation at location c, based on information decompositions of $I_{2,E,W}$ and $I_{2,N,S}$ (fractions of I_2, units of bits per bit)	(3)
S_{EW}, S_{NS}	Synergistic information transfers from neighboring directions to current precipitation at location c, based on information decompositions of $I_{2,E,W}$ and $I_{2,N,S}$ (fractions of I_2, units of bits per bit)	(3)
R_{EW}, R_{NS}	Redundant information transfers from neighboring directions to current precipitation at location c, based on information decompositions of $I_{2,E,W}$ and $I_{2,N,S}$ (fractions of I_2, units of bits per bit)	(3)
I_4	Multivariate mutual information provided to current precipitation at location c from strongest four of the eight individual lagged information sources, determined from TE values (units of bits)	(5)

shuffled information values, and the detected measure is tested for significance at a 99% confidence level. Table 1 provides a summary list of measures with descriptions of their different characteristics.

d. Precipitation dataset

We apply our information theory–based measures to the CPC Unified gauge-based analysis of daily precipitation over the contiguous United States (CONUS). This dataset is a gridded, gauge-based product available at a 0.25° resolution from 1948 to 2018 (Chen et al. 2008; Xie et al. 2007). We convert the data into binary form, using a magnitude ranging from 0.3 to 3 mm as the threshold between wet ($X_{t,c} = 1$) and dry ($X_{t,c} = 0$) states. This threshold value is chosen at each grid cell individually based on that which maximizes the value of $I(P_{t,c}; P_{t-1,c})$. This corresponds to the "maximally informative" quantization, in that it is the threshold associated with the highest predictability based on the knowledge of yesterday's state. While this threshold varies across grid cells ("Thresholds" and Fig. S1 in the supplemental material), we note that it has previously been found from this dataset that information measures vary little whether a 0.3 mm, a 1 mm, or a 10th percentile threshold is used at a given location (Goodwell and Kumar 2019). In other words, small differences between thresholds for wet and dry states do not have a large influence on results. It has generally been found for gridded datasets that interpolation increases the number of small events and decreases the number of large events (Ensor and Robeson 2008). In this way, we expect our threshold values to omit some of these spurious small events. We also expect that interpolation from gauge data to the gridded dataset causes some redundancies in sparsely monitored regions (Carvalho 2020), in that precipitation may appear more synchronized over large land areas without gauges. However, as later results show, redundancies between grid cells do not appear to dominate regional characteristics, and we assume that spatial dependencies largely reflect the actual timing of daily precipitation occurrences and synchronization. Finally, the data are segmented into seasonal windows (DJF, MAM, JJA, SON) to capture differences between spatiotemporal measures of predictability.

e. Temporal trends and correlations with climate indices

Information measures (Table 1) are first computed seasonally for the entire 1948–2018 period to capture general regional behaviors. Trends are computed based on the same measures computed on individual seasonal time windows. We detect trends in each measure at a 95% confidence level using least squares linear regression, and detected similar results for the nonparametric Sen slope method (Sen 1968; Fernandes and Leblanc 2005). A trend in MI, TE, I_2, or I_4 indicates a change in the level of predictability given knowledge of past precipitation, and a trend in θ indicates a shift in dominant direction. Angular trends were detected with circular correlation statistics (Berens 2009). As an illustration, a $1° yr^{-1}$ positive trend is equivalent to a 70° shift clockwise over 70 years. For a grid cell where information flow as measured by TE_{vect} was predominately from the west, this would indicate a shift clockwise toward the north. Finally, a trend in S, U, or R would indicate a trend in information type, which is related to both the synchronicity and direction of precipitation across the region. In addition to trends over the past 70 years, we detect correlations between information measures and seasonally averaged climate indices to reveal climatological influences on precipitation persistence and direction. We initially choose 10 indices ("Climate Indices" in the supplemental material), and focus on four that are particularly correlated to information measures in certain regions during either winter or summer. Monthly values for each climate index are averaged to seasonal measures to match the time scale of the information-based measures of precipitation predictability.

3. Results: Directional precipitation predictability

Any given information measure (listed in Table 1) can be normalized by precipitation entropy, $H(X_{t,c})$ to obtain a fraction of uncertainty reduced. For example, a case where MI = $H(X_{t,c})$ indicates that knowledge of a neighboring lagged state

fully reduces precipitation uncertainty. Since the proposed measures represent different levels of analysis in that they involve different numbers of lagged neighboring grid cells, we first compare their relative magnitudes as percent reductions in uncertainty. On average for all seasons and grid cells, TE_{vect} is the lowest, at a 0%–5% reduction in uncertainty of precipitation at a central grid cell (Fig. 3). Meanwhile, MI_{vect} is slightly larger with uncertainty reductions up to about 10%. This indicates that conditioning mutual information on the lagged history at the central cell decreases the amount of uncertainty reduced. As expected, the knowledge of the lagged state at a neighboring cell provides some redundant information with the knowledge of the lagged central state due to synchronization of wet and dry days across regions. This feature could also relate to the interpolation of gauges to gridded data, which could cause neighboring cells to be similar. If instead the knowledge of the neighboring cell provided mostly additional information beyond that already provided by the history of the central cell itself, uniquely or synergistically, then TE would be larger than MI. However, this only occurs for a few cases with no particular spatial or seasonal patterns.

Meanwhile, the higher-order measure of $I_{2,EW}$, the joint mutual information provided by the knowledge of lagged neighbors to the east and west, is much higher and more variable than MI or TE, on average reducing about 11% of precipitation uncertainty (Fig. 3). This is expected because $I_{2,d1d2}$ does not condition on the lagged central state, but quantifies the total uncertainty reduced given two neighbors. $I_{2,NS}$ shows very similar behavior (not shown). Finally, I_4, as the highest-dimensional information measure, reduces an average of 12% of uncertainty, and over 20% for some grid cells. This means that the knowledge of past precipitation states at several surrounding locations provides significantly more predictability than only knowing one neighboring state. However, we see that I_4 is only slightly larger than I_2, indicating that knowledge of additional neighboring states provides "diminishing returns" in information gain.

Next we discuss these measures in terms of directionality and seasonal differences across the United States. We first note that directional distributions of MI_{max} and MI_{vect} measures differ because θ_{MImax} can only take on one of the eight neighboring directions, so that its distribution is discrete, while θ_{MIvect} can take any angular direction, since it is the angle associated with the vector resolution of all eight neighboring MI values ("Information Theory Methods" in the supplemental material). On a national average, information flows as measured by lagged MI tend to originate from the west (Figs. 4a–d), which reflects the dominant wind patterns across the United States due to the jet stream. In some cases, we see that information from opposite directions appears to cancel out in the θ_{MIvect} calculation. This occurs for example in the summer, where θ_{MImax} is predominantly from the west, but the dominant θ_{MIvect} is from the southwest (Fig. 4c). This occurs because MI is relatively large in multiple directions, such that east and west cancel out. These results indicate that on a national average, the knowledge of yesterday's precipitation state toward the west or southwest best predicts the current state.

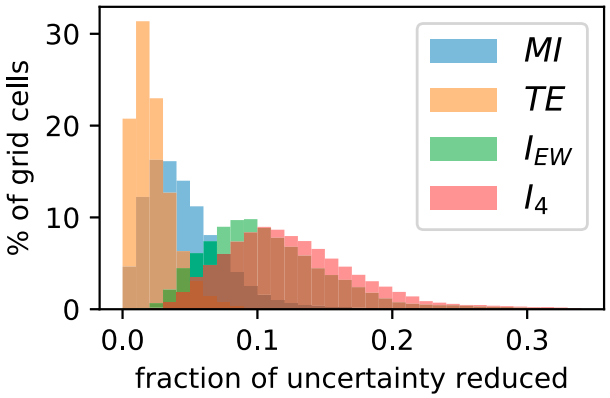

FIG. 3. Distribution of measures MI_{vect}, TE_{vect}, $I_{2,EW}$, and I_4 for all grid cells and seasons, which represent different levels of analysis of precipitation predictability. Measures are normalized by precipitation entropy $H(X_{t,c})$, such that they indicate fractions of reduced uncertainty.

Transfer entropies, TE, tend to be smaller in magnitude as discussed previously, and tend more toward the south relative to the west. This difference between MI_{vect} and TE_{vect} is prevalent in all seasons except summer (Figs. 4c,g) when the directions are very similar. We also observe weaker TE and MI in summer relative to other seasons, indicating lower predictability of precipitation in general given past states. Differences between MI_{max} and TE_{max} are even more apparent (black lines in Fig. 4). Mainly, we see much higher frequencies of TE_{max} in the noncardinal directions compared to MI_{max}, particularly from NW and SW. This is likely due to the conditioning of the mutual information on lagged precipitation at the central grid cell, $X_{t-1,c}$, for TE. Since the centers of the grid cells in N, S, E, and W directions are closer to the center grid cell, the information that they provide appears to be more redundant with the information provided by the history of the center cell itself. This could also relate to interpolation in the gridded dataset, in that neighboring cells are similar. A factor accounting for the distances between neighboring grid cell centers would be expected to even out this redundancy. In general, the tendency for dominant TE to come from the south rather than the west indicates that when lagged precipitation at the central location is known, it is more useful to additionally know the lagged state toward the south than the west. In contrast, a simpler analysis of MI between a lagged neighbor and a central location detects a stronger dependency from the west. This indicates that while storms generally move from west to east, they are somewhat angled northward or southward, such that looking in those directions enhances predictability. In other words, a storm may be largely moving westward so that correlations are high between east to west neighbors, but the shape and angle of the storm are such that north to south neighbors provide more "extra" information. This illustrates the different conclusions that could be drawn from analyses with different dimensions and conditioning.

While the nationally averaged distributions of MI and TE strengths and directions show broad seasonal trends, the same distribution of measures for regional HUC2 watersheds

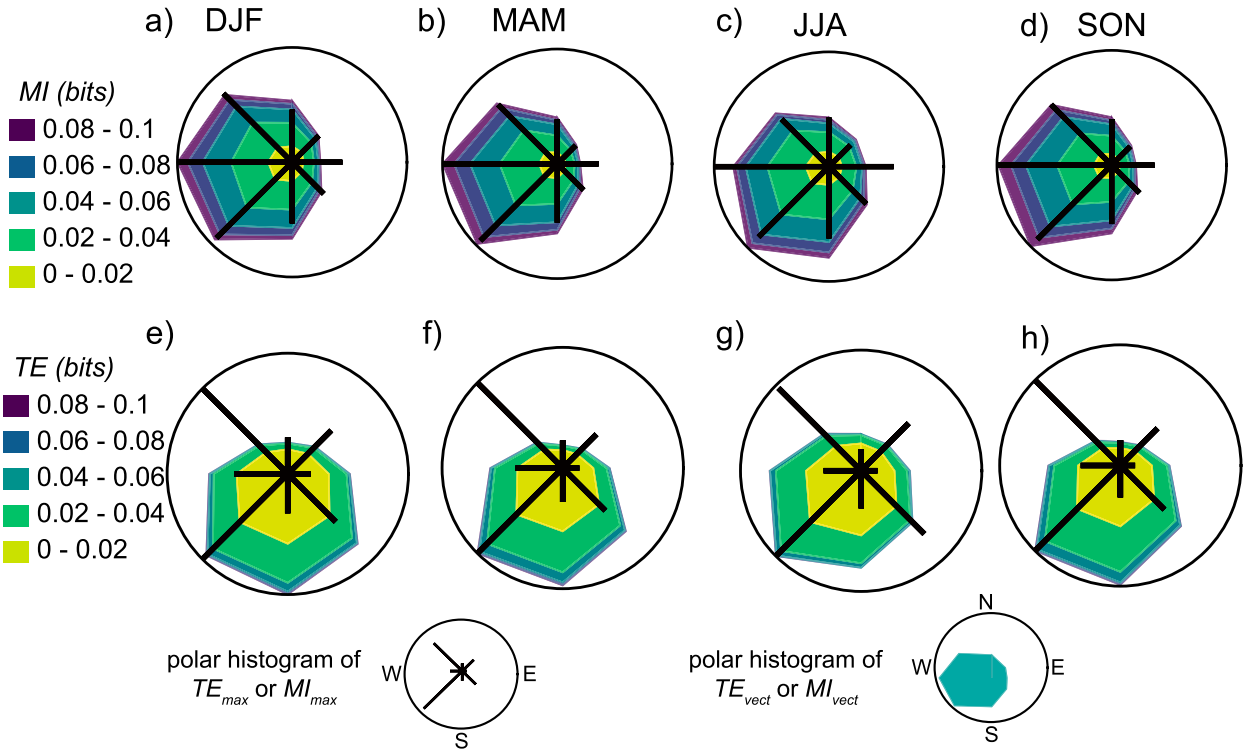

FIG. 4. Seasonal polar histograms of (a)–(d) MI and(e)–(h) TE on a national average, for 13 626 grid cells. The black lines indicate the directions θ_{MImax} and θ_{TEmax} that take on one of the eight surrounding directions, while the filled areas indicate frequencies for dominant directions and strengths of $\mathrm{MI}_{\mathrm{vect}}$ and $\mathrm{TE}_{\mathrm{vect}}$. Colors correspond to the strengths (bits) of information contents. For example, in the winter [DJF, shown in (e)], the most common dominant direction of TE is from the southwest and northwest. However, when information from all directions is resolved into a single vector, the dominant direction is from the south.

(Seaber et al. 1987) shows features of different regions of CONUS (Fig. 5). Differences in these directions should relate to topographic divides and landscape properties, and different seasonal drivers of precipitation. We group these basins into five regions with common features as described below.

- Northeast (Mid-Atlantic, New England, Ohio, Great Lakes, Upper Mississippi): In summer (red in Fig. 5), TE is dominantly from the southwest and MI from the west, following national averages. The majority of extreme precipitation events occur in summer in this region, and most of these are attributable to frontal systems (Kunkel et al. 2012). In winter (blue in Fig. 5), θ_{TEvect} has a broader distribution, or more variability in dominant direction. This could indicate winter influence of both continental and ocean dynamics, relative to summer influences mainly from the west.
- Southeast: (South Atlantic–Gulf, Lower Mississippi, Tennessee, Arkansas–White–Red, Texas–Gulf): In winter, TE is detected from the south, angled toward the Gulf Coast. Distributions are slightly more narrow in winter relative to summer, indicating more of a dominant directionality in winter storms. During winter in Texas, wet periods are often associated with southeasterlies that bring moist air from the Gulf of Mexico (Lyons 1990). In summer, TE is dominantly from north or west, with very broad distributions in some regions, reflecting the large variety of precipitation causes in this season (Kunkel et al. 2012).
- Southwest (California, Lower Colorado, Rio Grande): Here, TE is dominantly from south or southwest in winter. This reflects precipitation events due to atmospheric rivers that convey moisture from the Pacific Ocean during this season (Gershunov et al. 2017). In the summer when conditions are relatively drier in most areas, the distributions are broader and mainly from the north.
- Northwest (Pacific Northwest, Upper Colorado, Great Basin): TE is mainly from south in winter, similar to the Southwest basins, but is derived from the east in summer. This is opposite of the summertime θ_{TEvect} in the eastern part of the country, indicating the prevalence of weather that originates from the mountainous regions in the country. In these basins and to some extent in the Southwest, the patterns in TE reflect the reversal from westerly winds in the winter to more easterly winds in the summer (Dominguez et al. 2008).
- North-central (Missouri, Souris–Red–Rainy): Here, distributions are relatively uniform between summer and winter, indicating the least influence from ocean patterns. Very few of the most extreme storms occur in the winter in this region relative to other seasons (Kunkel et al. 2012), but this analysis shows that the origins of precipitation are likely similar throughout the year.

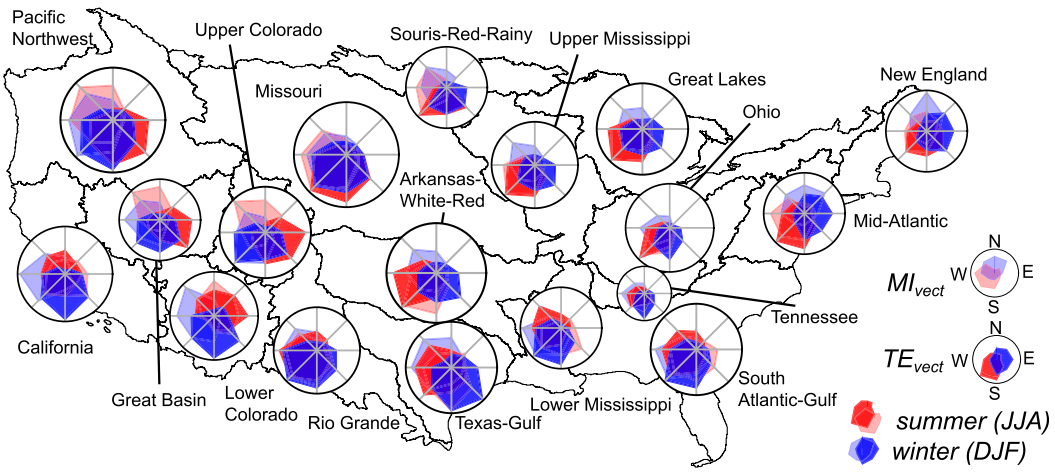

FIG. 5. Spatial variability of directions θ_{TEvect} (solid colors) and θ_{MIvect} (transparent colors in background) during the summer (JJA, red) and winter (DJF, blue). Each polar histogram represents the dominant direction of 1-day lagged precipitation dependencies for a HUC2 watershed. Circle sizes are approximately scaled to the relative basin sizes.

a. Uniqueness and redundancy dominant over synergy

Next, we consider $I_{2,\mathrm{EW}}$, the information provided to a central grid cell due to the knowledge of 1-day lagged states in both east and west directions. When this measure is further partitioned into unique, redundant, and synergistic information types ("Information Theory Methods" in the supplemental material), the main result is a balance between redundancy R and uniqueness U (Figs. 6a–d), while synergistic information S is much lower (Figs. 6e,f). The term $I_{2,\mathrm{NS}}$, the joint information from north and south directions, was similar to this (not shown). While high R indicates synchronization between precipitation in east and west directions, high U indicates that the information contributing to predictability tends to come from one direction or the other, but not both.

Redundancy is the strongest information component for most regions in both summer and winter, indicating a large degree of synchronization in precipitation from the eastern and western directions. This matches with the finding of high MI but lower TE in the west-to-east direction, associated with the dominant direction of storm movement in the country. This may be expected given the temporal resolution considered here, as it is likely to be simultaneously wet or dry over regions that cover multiple grid cells within a given day. A precipitation dataset at a higher time resolution, or over a very large spatial grid, may exhibit less redundancy, in that subdaily rainfall would better resolve the movement of storms and a larger grid size would make it less likely that neighboring cells are synchronized. Uniqueness, as the combination of U_{E} and U_{W}, is the next strongest information component, and in some grid cells $U_{\mathrm{E}} + U_{\mathrm{W}}$ makes up more than half of the total value of $I_{2,\mathrm{EW}}$. For example, we detect high uniqueness in Florida and Southern California in the summer (Fig. 6c), associated with prevalent weather influences from the Atlantic and Pacific Oceans, respectively. Unique information is also high in the Rocky Mountains region in the winter, indicating an orographic effect where precipitation moves in one direction and then dissipates (Fig. 6d). These regions of high uniqueness correspond to regions of low redundancy for the same season (Figs. 6a,b). Over the country, we note that R is generally higher in the winter relative to the summer. This indicates a larger degree of spatial uniformity of precipitation during the winter, in that precipitation is more synchronized across grid cells. This finding matches with a larger spatial scale of frontal systems in the winter (Kunkel et al. 2012) relative to the summer.

Synergistic information S is always less than 10% of $I_{2,\mathrm{EW}}$, indicating that the joint knowledge of lagged precipitation states in east and west directions does not provide a large amount of "extra" information about the current central precipitation state, $X_{t,c}$, beyond their individual and overlapping contributions. However, we do see spatial patterns in S, which is particularly high in California and the Southwest in the summer (Fig. 6e), and in the Sierra Nevada and Cascades ranges in the winter (Fig. 6f). In these regions and seasons, it is relevant to "look both ways" simultaneously when making a prediction about the current precipitation state. To illustrate this, $S = 0.05$ would indicate that considering lagged states in both east and west directions provides an extra 5% of predictability relative to only knowing the lagged state in either direction individually. These regions of relatively high synergistic information could mark topographic or climatic boundaries where weather may arrive from either of the two opposing directions independently. Particularly, high rates of precipitation recycling have been detected in these mountainous regions, which is associated with the North American monsoon during the summer (Dominguez et al. 2008). Here, high S could reflect joint drivers of precipitation, which occurs due to both incoming moisture and recycled moisture from the regional landscape.

b. Temporal trends vary from east to west

Next we consider trends in information measures over the 70-yr study period. We particularly focus on multivariate information from the four most individually informative

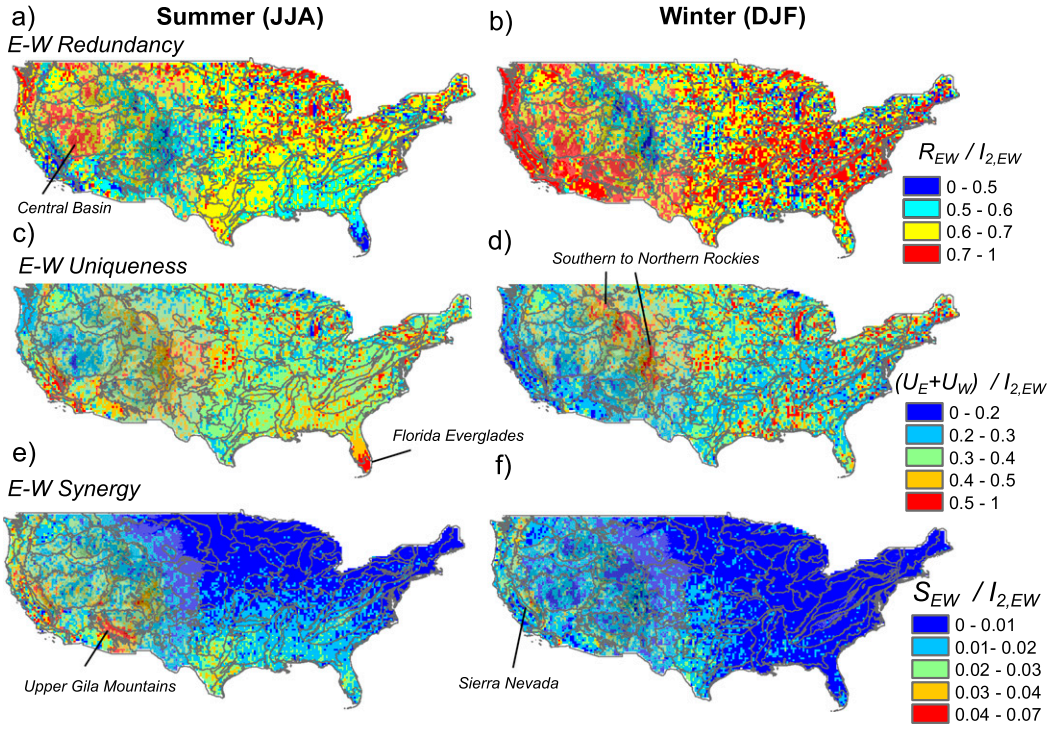

FIG. 6. Proportions of (a),(b) redundant, (c),(d) unique, and (e),(f) synergistic information between east and west directions for summer (JJA) and winter (DJF) across the United States. Values of S_{EW}, $U_E + U_W$, and R_{EW} sum to 1 for a given location and season, such that each value represents a fraction of total information $I_{2,EW}$ from Eq. (4). Unique information combines individual information components from the east and west. Outline borders show level 3 eco-regions (McMahon et al. 2001), several of which are labeled, and shading in the western United States highlights mountainous regions (darker shading indicates higher elevation).

neighbors, I_4. For most seasons, more trends were detected for I_4 than for other information measures such as TE or MI, but spatial and seasonal patterns were similar. S, U, and R information types showed very few statistically significant trends. While the average magnitudes of increasing and decreasing I_4 trends appear small, about 0.0014 bits per year (Table 2), over the 70-yr study period this amounts to percent changes that range from 26% to 27.7% increases and from 21.9% to 25.4% decreases. In other words, the predictability of precipitation given the knowledge of four lagged neighbors has increased or decreased by about 25% for some grid cells. There are also differences in the numbers of grid cells exhibiting statistically significant trends in each season. For winter, spring, and fall, a similar fraction of grid cells show a trend. In summer, 0.26, or about 3500 out of 13 626 grid cells, show a decreasing trend while only 0.05, or 680 grid cells, have increasing trends (Table 2).

Next, we link these national average trends to regional behaviors across seasons. Overall, I_4 tends to be decreasing in the eastern United States, and increasing in the Northwest (Fig. 7). This corresponds to the finding of increased precipitation predictability in general in the Northwest (Goodwell and Kumar 2019), in that the knowledge of lagged precipitation states at a given location provides increasing amounts of information regarding the current state. Here, we find that this trend of increasing predictability also exists when a spatial component of precipitation occurrence is considered. Meanwhile, east of the Continental Divide, I_4 is decreasing in strength. This indicates that in this region, precipitation is becoming less predictable given knowledge of lagged neighboring states. In the winter, decreasing I_4 is particularly clustered along the Texas–Gulf Coast (Fig. 7a). When paired with the dominant θ_{TEvect} from the southeast in this region during the winter (Fig. 5), this indicates a weakening influence, or lower persistence, of precipitation states over the Gulf of Mexico. In summer, very few increasing trends were detected, and the region of decreasing I_4 covers most of the east coast.

TABLE 2. Positive (+) and negative (−) trends in I_4. Fractions are computed as the number of grid cells showing a trend, divided by the total number of grid cells (13 626). The percent (%) indicates the percent increase or decrease in I_4 over the 70-yr window associated with a given trend in bits per year.

I_4 trends	DJF	MAM	JJA	SON
Fraction +	0.14	0.15	0.05	0.13
Avg + trend (bits)	0.0014	0.0014	0.0013	0.0015
Avg + trend (%)	27.7	27.6	26.0	26.0
Fraction −	0.14	0.17	0.26	0.15
Avg − trend (bits)	−0.0013	−0.0013	−0.0014	−0.0014
Avg − trend (%)	−25.4	−22.9	−22.5	−21.9

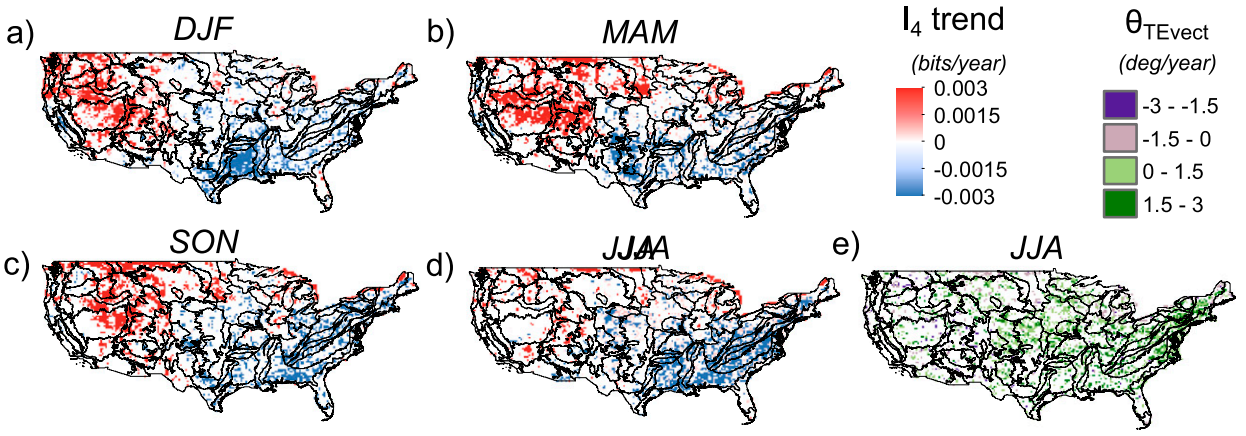

FIG. 7. Trends in information measure I_4 over 1950–2018 study period for (a) winter (DJF), (b) spring (MAM), (c) fall (SON), and (d) summer (JJA) time windows, and (e) θ_{TEvect} for summer. White regions indicate no statistically significant trend. An increase in I_4 indicates an increase in predictability given lagged neighboring histories, while a positive angular trend in θ_{TEvect} indicates a trend in the clockwise direction, e.g., from west toward northwest or from east toward southeast.

In terms of the direction associated with TE_{vect}, θ_{TEvect}, we see that more grid cells show a positive, or clockwise trend, relative to a negative trend (Fig. 7e). This pattern is similar for all seasons (not shown). To some extent, clockwise trends in θ are most clustered the eastern United States where measures of spatial predictability (I_4, MI, and TE) are decreasing. In the western United States where spatial predictability is increasing, there are more grid cells that show counterclockwise (negative) trends in direction. From this and the prior analysis for dominant directions, we can broadly interpret that in the western United States, information about precipitation is generally strongest from the east in the summer (Fig. 5), and has increased in strength of predictability and shifted from the east toward the north over the 70-yr study period. In contrast, in the eastern United States the dominant direction of summertime predictability is from the west, and the strength of this predictability has decreased and shifted clockwise toward the north over the study period.

c. *Correlations between information measures and climate indices*

Aside from linear trends over the past 70 years, correlations between seasonally averaged climate indices and our information theory–based measures reveal potentially more complex aspects of spatial and temporal precipitation predictability. Climate teleconnections can influence storm tracks and precipitation magnitudes, and can also feedback with the land surface to cause further events or droughts (Trenberth and Guillemot 1995). Here we look at correlations between multivariate information, I_4, and several climate indices. In general for the set of 10 climate indices tested for correlations (''Climate Indices'' in the supplemental material), we find fewer statistically significant correlations relative to linear trends, but several climate indices do exhibit patterns of correlations for certain seasons.

The PDO tends to be positively correlated with I_4, particularly in the western United States in winter (Fig. 8a). There are pockets of positive PDO correlations in other regions in the summer, such as in Missouri, northern Michigan, and the Pacific Northwest. Meanwhile, the AMO is mainly negatively correlated with I_4, and more correlations are spread over the country. These negative correlations are particularly clustered in the eastern United States during the summer. The PDO and the AMO have been linked to climate variability in the Americas on decadal to multidecadal time scales (Carvalho 2020). Particularly, a negative correlation has been detected between maximum streamflows and the PDO in winter in the western United States, and some negative correlations between streamflows and the AMO have been detected during all seasons in the eastern United States (Dickinson et al. 2019). While we focus on nonextreme precipitation events that do not necessarily lead to flood events, these relationships between climate indices and flows are driven by the spatial movement of precipitation events such as those studied here. The widespread correlations between I_4 and the AMO detected here also reflect previous findings that the Atlantic Warm Pool (AWP) impacts precipitation in much of the central and eastern United States (Liu et al. 2015).

We also observe regional correlations that differ between seasons, for example positive correlations between I_4 and the EPNP in the Southeast in summer, and between I_4 and the NAO in winter (Fig. 8c). Specifically, the EPNP is positively correlated with I_4 in a band from North Carolina to Alabama, while the NAO is positively correlated with the same metric slightly farther south, in northern Florida. These correlations dominate during opposite parts of the year, indicating different drivers of precipitation movement during different seasons. The NAO has been associated with high precipitation in the Southeast (Stenseth et al. 2003), and streamflow variability has been tied to Atlantic Ocean sea surface temperatures (Sadeghi et al. 2019). In parts of this region, durations of dry periods have also been found to be increasing (Brown et al. 2020; Sadeghi et al. 2019). Particularly in northern Florida, the NAO is positively correlated with maximum streamflows

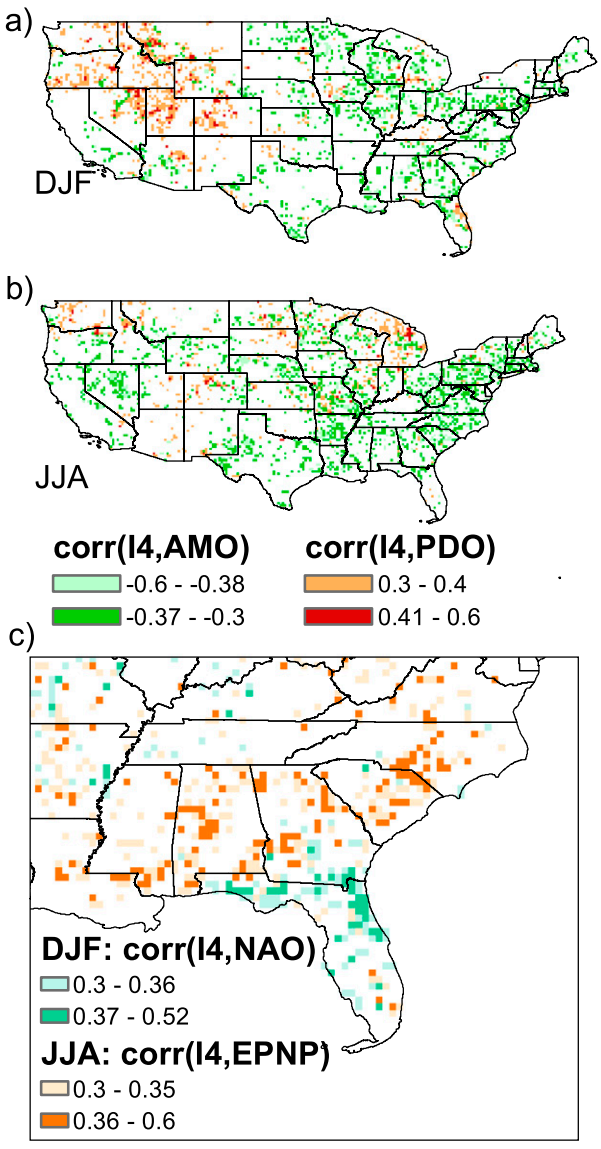

FIG. 8. Correlations between information measures and climate indices. Correlations between I_4 and the PDO and AMO for (a) winter (DJF) and (b) summer (JJA). Red colors indicate positive correlations between I_4 and PDO, while green colors indicate negative correlations between I_4 and AMO. There were very few statistically significant negative correlations for PDO, or positive ones with the AMO. (c) Correlations between I_4 and the EPNP and I_4 and the NAO for different seasons in the U.S. Southeast.

(Dickinson et al. 2019). The positive correlation between information measures and the NAO in this study indicates that the index is also associated with predictability based on the knowledge of neighboring regions. This also matches with the known maximum influence of the NAO on climate variability in the winter (Hurrell and Deser 2010).

Finally, we note that opposite correlations are detected between I_4 and the NAO and AMO, both of which are related to Atlantic Ocean temperatures. This reflects a potential difference in time scales of influences. For example Liu et al. (2015) found that warming Atlantic temperatures have opposite effects on U.S. precipitation on interannual and multidecadal time scales. Specifically, warming Atlantic temperatures on multidecadal time scales results in less precipitation, which in this study correlates to lower predictability, or a negative correlation between I_4 and the AMO. Atlantic warming on interannual time scales is linked to more precipitation, and here correlates to higher predictability, or a positive correlation between I_4 and the NAO. Meanwhile, there are fewer known linkages to precipitation and Pacific indices such as the EPNP, although a high EPNP has previously been associated with severe flooding in the Midwest (Bell and Janowiak 1995). This is an example of a climate teleconnection, or large-scale atmospheric dynamic (Leathers et al. 1991), that can influence weather patterns in distant locations.

4. Discussion

The spatial and temporal persistence of precipitation occurrence characterized here could have implications for hydrologic behaviors, such as the timing of flows due to a moving storm event, or extreme events that lead to flooding or droughts. Here we discuss several findings in the context of potential downstream implications, predictability, and climate connectivity.

Dominant directions of precipitation predictability vary from east to west in the United States, and uncertainty reduced from knowing lagged precipitation at neighboring locations is partially redundant with the knowledge of the local state. This also relates to the partitioning of information measures, which shows that unique and redundant information types dominate over synergistic information. While some of this redundancy likely results from the nature of the spatially interpolated gridded dataset, it still reflects the persistence of precipitation in both time and space. The characteristic directions of dominant information influences vary by region, and the influence of oceanic currents that cause weather events is apparent in the coastal regions. For example, precipitation predictability in California is typically derived from the knowledge of the lagged precipitation state toward the west, indicating the prevalence of events coming from the Pacific Ocean. On a national average, patterns in spatial predictability follow the prevailing wind pattern in the United States from west to east.

Precipitation predictability has increased in the western United States and decreased in the eastern United States, with corresponding shifts in the dominant direction of influence. Trends in spatial information transfers match with the previous finding of increased or decreased persistence in similar areas of the country (Goodwell and Kumar 2019). For example, in the Pacific Northwest, Goodwell and Kumar (2019) found that precipitation at a given location has become more predictable given the knowledge of its own time-lagged history. In contrast, in the eastern part of the country, precipitation has become less predictable, given both knowledge of lagged states at a single location and lagged neighboring states. This indicates that not only is precipitation occurring more

sporadically, but that storms are not coherent enough spatially that precipitation would become more predictable given knowledge of neighboring states.

Several climate indices are correlated to our measure of multivariate precipitation predictability for different regions and seasons. Climate indices represent large-scale processes of the Earth system, and have been found to link to global temperatures and patterns of carbon fluxes (Zhu et al. 2017). Climate indices have also been found to influence cycles of extreme events. For example, ENSO modulates the annual tornado cycle (Allen et al. 2018), climate cycles relate to extreme precipitation over Texas (Bhatia et al. 2019), and the NAO has been found to be a dominant driver of events in the North Atlantic region (Stenseth et al. 2003). Some ecological studies have found that large-scale climate indices are better predictors of populations and ecosystem health than local indices because they combine multiple stressors rather than a single metric (Hallet et al. 2004). However, there are various challenges in relating local precipitation patterns to climate indices. First, climate indices simplify very large scale behaviors that may or may not be relevant at local scales (Stenseth et al. 2003). On top of this, nonstationarity and nonlinear relationships may result in a climate index only explaining a small part of local climate variability, in the form of relatively weak correlations (Stenseth et al. 2003). This nonstationarity may also exist within a climate index itself. For example, a shift in the PDO has been detected since the late 1980s which may change any interpretation of its relationship to regional climate variability (Litzow et al. 2020). Finally, it has been found that some climate indices are more strongly correlated to local environmental features on a lagged time scale, where a climate index may predict fluxes one or two seasons in advance (Zhu et al. 2017). These challenges are apparent in this study in that most climate indices were not found to correlate with information-based precipitation measures. A lagged correlation analysis or a more targeted focus on particular climate features may reveal more dependencies, or it could be that climate indices correlate better in general with other precipitation properties, such as durations or magnitudes, rather than persistence and spatial directions of predictability. However, several regional clusters of correlations and seasonal patterns indicate that there are relevant connections between storm movement and precipitation predictability and large-scale climate features. Some of these, such as the correlation between the EPNP and I_4 measure in the Southeast, potentially indicate very long range connections. Others, such as the widespread correlation of I_4 with the AMO, indicate common drivers of precipitation variability over large regions of the United States (Kim et al. 2020).

5. Conclusions

Although we focus only on predictability based on the knowledge of past precipitation occurrences, the results presented here could further link to changes in the water balance and ecosystem properties for various locations. For example, the regional clustering of flood events (Dickinson et al. 2019) and storm tracks and intensities (Hurrell and Deser 2010) both relate to spatial and temporal patterns in precipitation, and have been found to relate to climate indices. To further understand implications of shifts in precipitation predictability, a more focused analysis of individual regions could involve basin characteristics such as the stream network structure, soil types, evapotranspiration, and human management. In determining how spatial persistence is related to flow rates in a given stream, aspects such as evapotranspiration, water uses, and groundwater may be confounding factors. For example, it has been found that up to 40% of evaporation dynamics can be explained by climate indices (Martens et al. 2018). In this way, large-scale climate forcing and local conditions may exert different controls on different aspects of the water balance. From a climate perspective, there may be more correlations between precipitation and climate indices detected if climate indices were lagged in time, or if precipitation magnitudes were considered more explicitly.

The presence of trends over time and correlations of information measures with climate indices indicates that the movement of precipitation in space, regardless of event size, has changed in certain regions. In general, this type of study can be useful to both understand patterns in precipitation across large regions, or more regional linkages between precipitation, streamflow, and watershed characteristics. In a broader context, information-based methods such as those applied here can be used to study persistence, predictability, and causal interactions for a variety of Earth system processes.

Acknowledgments. This research is supported by NSF Grant EAR 2012850, for the project Network Cluster CINet: Critical Interface Network in Intensively Managed Landscapes, a multidisciplinary and multi-institution collaborative effort. The support by NSF (grant EAR-1928724) and NASA (grant 80NSSC19K0726) to organize the 12th International Precipitation Conference (IPC12), Irvine California, June 2019, and produce the IPC12 special collection of papers is gratefully acknowledged. CPC U.S. Unified Precipitation data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their website at https://www.esrl.noaa.gov/psd/. Python codes for this analysis are available at https://github.com/allisongoodwell/RainingBitsSpatial.

REFERENCES

Allen, J. T., M. J. Molina, and V. A. Gensini, 2018: Modulation of annual cycle of tornadoes by El Niño–Southern Oscillation. *Geophys. Res. Lett.*, **45**, 5708–5717, https://doi.org/10.1029/2018GL077482.

Barlow, M., S. Nigam, and E. H. Berbery, 2001: ENSO, Pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow. *J. Climate*, **14**, 2105–2128, https://doi.org/10.1175/1520-0442(2001)014<2105:EPDVAU>2.0.CO;2.

Bell, G. D., and J. E. Janowiak, 1995: Atmospheric circulation associated with the midwest floods of 1993. *Bull. Amer. Meteor. Soc.*, **76**, 681–696, https://doi.org/10.1175/1520-0477(1995)076<0681:ACAWTM>2.0.CO;2.

Berens, P., 2009: CircStat: A MATLAB toolbox for circular statistics. *J. Stat. Software*, **31** (10), 1–21, https://doi.org/10.18637/jss.v031.i10.

Bhatia, N., V. P. Singh, and K. Lee, 2019: Variability of extreme precipitation over Texas and its relation with climatic cycles. *Theor. Appl. Climatol.*, **138**, 449–467, https://doi.org/10.1007/s00704-019-02840-w.

Boers, N., A. Rheinwalt, B. Bookhagen, H. M. J. Barbosa, N. Marwan, J. Marengo, and J. Kurths, 2014: The South American rainfall dipole: A complex network analysis of extreme events. *Geophys. Res. Lett.*, **41**, 7397–7405, https://doi.org/10.1002/2014GL061829.

Brown, V. M., B. D. Keim, and A. W. Black, 2020: Trend analysis of multiple extreme hourly precipitation time series in the southeastern United States. *J. Appl. Meteor. Climatol.*, **59**, 427–442, https://doi.org/10.1175/JAMC-D-19-0119.1.

Carvalho, L. M., 2020: Assessing precipitation trends in the Americas with historical data: A review. *Wiley Interdiscip. Rev.: Climate Change*, **11**, e627, https://doi.org/10.1002/wcc.627.

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.

Cover, T. M., and J. A. Thomas, 2006: *Elements of Information Theory*. 2nd ed. Wiley, 792 pp.

Dickinson, J. E., T. M. Harden, and G. J. McCabe, 2019: Seasonality of climatic drivers of flood variability in the conterminous United States. *Sci. Rep.*, **9**, 15321, https://doi.org/10.1038/s41598-019-51722-8.

Dominguez, F., P. Kumar, and E. R. Vivoni, 2008: Precipitation recycling variability and ecoclimatological stability—A study using NARR data. Part II: North American Monsoon region. *J. Climate*, **21**, 5187–5203, https://doi.org/10.1175/2008JCLI1760.1.

Ensor, L. A., and S. M. Robeson, 2008: Statistical characteristics of daily precipitation: Comparisons of gridded and point datasets. *J. Appl. Meteor. Climatol.*, **47**, 2468–2476, https://doi.org/10.1175/2008JAMC1757.1.

Fernandes, R., and S. G. Leblanc, 2005: Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. *Remote Sens. Environ.*, **95**, 303–316, https://doi.org/10.1016/j.rse.2005.01.005.

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.

Gao, S., and Z. Fang, 2018: Using storm transposition to investigate the relationships between hydrologic responses and spatial moments of catchment rainfall. *Nat. Hazards Rev.*, **19**, 04018015, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000304.

Gershunov, A., and T. P. Barnett, 1998: ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *J. Climate*, **11**, 1575–1586, https://doi.org/10.1175/1520-0442(1998)011<1575:EIOIER>2.0.CO;2.

——, T. Shulgina, F. M. Ralph, D. A. Lavers, and J. J. Rutz, 2017: Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophys. Res. Lett.*, **44**, 7900–7908, https://doi.org/10.1002/2017GL074175.

Goodwell, A. E., and P. Kumar, 2015: Information theoretic measures to infer feedback dynamics in coupled logistic networks. *Entropy (Basel)*, **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.

——, and ——, 2019: A changing climatology of precipitation persistence across the United States using information-based measures. *J. Hydrometeor.*, **20**, 1649–1666, https://doi.org/10.1175/JHM-D-19-0013.1.

Hallet, T. B., T. Coulson, J. G. Pilkington, T. H. Clulton-Brock, J. M. Pemberion, and B. T. Grenfell, 2004: Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature*, **430**, 71–75, https://doi.org/10.1038/NATURE02708.

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.

Hurrell, J. W., and C. Deser, 2010: North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, **79**, 231–244, https://doi.org/10.1016/j.jmarsys.2009.11.002.

Kim, D., S.-K. Lee, H. Lopez, G. R. Foltz, V. Misra, and A. Kumar, 2020: On the role of Pacific-Atlantic SST contrast and associated Caribbean Sea convection in August-October US regional rainfall variability. *Geophys. Res. Lett.*, **47**, e2020GL087736, https://doi.org/10.1029/2020GL087736.

Kim, D. H., and Y. Seo, 2013: Hydrodynamic analysis of storm movement effects on runoff hydrographs and loop-rating curves of a V-shaped watershed. *Water Resour. Res.*, **49**, 6613–6623, https://doi.org/10.1002/wrcr.20535.

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.

Leathers, D., B. Yarnal, and M. Palecki, 1991: The Pacific/North American teleconnection pattern and United Sates climate. Part I: Regional temperature and precipitation associations. *J. Climate*, **4**, 517–528, https://doi.org/10.1175/1520-0442(1991)004<0517:TPATPA>2.0.CO;2.

Lee, K., and J. Huang, 2007: Effect of moving rainstorms on attainment of equilibrium discharge. *Hydrol. Processes*, **21**, 3357–3366, https://doi.org/10.1002/hyp.6548.

Litzow, M. A., and Coauthors, 2020: The changing physical and ecological meanings of North Pacific Ocean climate indices. *Proc. Natl. Acad. Sci. USA*, **117**, 7665–7671, https://doi.org/10.1073/pnas.1921266117.

Liu, H., C. Wang, S.-K. Lee, and D. Enfield, 2015: Inhomogeneous influence of the Atlantic warm pool on United States precipitation. *Atmos. Sci. Lett.*, **16**, 63–69, https://doi.org/10.1002/asl2.521.

Lyons, S. W., 1990: Spatial and temporal variability of monthly precipitation in Texas. *Mon. Wea. Rev.*, **118**, 2634–2648, https://doi.org/10.1175/1520-0493(1990)118<2634:SATVOM>2.0.CO;2.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1080, https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Martens, B., W. Waegeman, W. A. Dorigo, N. E. C. Verhoest, and D. G. Miralles, 2018: Terrestrial evaporation response to modes of climate variability. *npj Climate Atmos. Sci.*, **1**, 43, https://doi.org/10.1038/s41612-018-0053-5.

McMahon, G., S. Gregonis, S. Waltman, J. Omernik, T. Thorson, J. Freeouf, A. Rorick, and J. Keys, 2001: Developing a spatial

framework of common ecological regions for the conterminous United States. *Environ. Manage.*, **28**, 293–316, https://doi.org/10.1007/s0026702429.

NOAA, 2019a: Atlantic Multidecadal Oscillation. NOAA Physical Sciences Laboratory, accessed 1 May 2019, https://www.esrl.noaa.gov/psd/data/timeseries/AMO/.

——, 2019b: East Pacific/North Pacific Oscillation. NOAA Climate Prediction Center, accessed 1 May 2019, https://www.cpc.ncep.noaa.gov/data/teledoc/ep.shtml.

——, 2019c: North Pacific Pattern. NOAA Physical Sciences Laboratory, accessed 1 May 2019, https://www.esrl.noaa.gov/psd/data/climateindices/list/.

——, 2019d: Pacific Decadal Oscillation. NOAA Physical Sciences Laboratory, accessed 1 May 2019, http://research.jisao.washington.edu/pdo/PDO.latest.

Roque-Malo, S., and P. Kumar, 2017: Patterns of change in high frequency precipitation variability over North America. *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.

Sadeghi, S., G. Tootle, E. Elliott, V. Lakshmi, M. Therrell, J. Kam, and B. Bearden, 2019: Atlantic Ocean sea surface temperatures and southeast United States streamflow variability: Associations with the recent multi-decadal decline. *J. Hydrol.*, **576**, 422–429, https://doi.org/10.1016/j.jhydrol.2019.06.051.

Schreiber, T., 2000: Measuring information transfer. *Phys. Rev. Lett.*, **85**, 461, https://doi.org/10.1103/PhysRevLett.85.461.

Seaber, P. R., K. F. Paul, 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/10.2307/2285891.

Sendrowski, A., and P. Passalacqua, 2017: Process connectivity in a naturally prograding river delta. *Water Resour. Res.*, **53**, 1841–1863, https://doi.org/10.1002/2016WR019768.

Seo, Y., and A. R. Schmidt, 2013: Network configuration and hydrograph sensitivity to storm kinematics. *Water Resour. Res.*, **49**, 1812–1827, https://doi.org/10.1002/wrcr.20115.

——, and ——, 2014: Evaluation of drainage networks under moving storms utilizing the equivalent stationary storms. *Nat. Hazards*, **70**, 803–819, https://doi.org/10.1007/s11069-013-0845-1.

——, ——, and M. Sivapalan, 2012: Effect of storm movement on flood peaks: Analysis framework based on characteristic timescales. *Water Resour. Res.*, **48**, W05532, https://doi.org/10.1029/2011WR011761.

Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 623–656, https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

Singh, V. P., 1997: Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph. *Hydrol. Processes*, **11**, 1649–1669, https://doi.org/10.1002/(SICI)1099-1085(19971015)11:12<1649::AID-HYP495>3.0.CO;2-1.

Stenseth, N. C., G. Ottersen, J. W. Hurrell, A. Mysterud, M. Lima, K.-s. Chan, N. G. Yoccoz, and A. Bjørn, 2003: Studying climate effects on ecology through the use of climate indices: The North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proc. Roy. Soc.*, **270B**, 2087–2096, https://doi.org/10.1098/rspb.2003.2415.

Ting, M., andH. Wang, 1997: Summertime U.S. precipitation variability and its relation to Pacific sea surface temperature. *J. Climate*, **10**, 1853–1873, https://doi.org/10.1175/1520-0442(1997)010<1853:SUSPVA>2.0.CO;2.

Trenberth, K., and C. Guillemot, 1995: Physical processes involved in the 1988 drought and 1993 floods in North America. *J. Climate*, **9**, 1288–1298, https://doi.org/10.1175/1520-0442(1996)009<1288:PPIITD>2.0.CO;2.

Volpi, E., M. Di Lazzaro, and A. Fiori, 2013: Analytical modeling of the hydrologic response under moving rainstorms: Storm-catchment interaction and resonance. *J. Hydrol.*, **493**, 132–139, https://doi.org/10.1016/j.jhydrol.2013.04.025.

Williams, P. L., and R. D. Beer, 2010: Nonnegative decomposition of multivariate information. arXiv, https://arxiv.org/abs/1004.2515.

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.

Zhu, Z., S. Piao, Y. Xu, A. Bastos, P. Ciais, and S. Peng, 2017: The effects of teleconnections on carbon fluxes of global terrestrial ecosystems. *Geophys. Res. Lett.*, **44**, 3209–3218, https://doi.org/10.1002/2016GL071743.