

Patterns of precipitation and soil moisture extremes in Texas, US: A complex network analysis

Alexander Y. Sun^{*,a}, Youlong Xia^b, Todd G. Caldwell^a, Zengchao Hao^c

^a Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78713, United States

^b I. M. Systems Group at Environmental Modeling Center, National Centers for Environmental Prediction, 5830 University Research Court, College Park, MD 207409, United States

^c College of Water Sciences, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Keywords:

Precipitation extremes
Flooding
Soil moisture
Event synchronization
Complex network theory
Community detection
Land surface models

ABSTRACT

Understanding of the spatial and temporal dynamics of extreme precipitation not only improves prediction skills, but also helps to prioritize hazard mitigation efforts. This study seeks to enhance the understanding of spatio-temporal covariation patterns embedded in precipitation (P) and soil moisture (SM) by using an event-based, complex-network-theoretic approach. Events concurrences are quantified using a nonparametric event synchronization measure, and spatial patterns of hydroclimate variables are analyzed by using several network measures and a community detection algorithm. SM–P coupling is examined using a directional event coincidence analysis measure that takes the order of event occurrences into account. The complex network approach is demonstrated for Texas, US, a region possessing a rich set of hydroclimate features and is frequented by catastrophic flooding. Gridded daily observed P data and simulated SM data are used to create complex networks of P and SM extremes. The uncovered high degree centrality regions and community structures are qualitatively in agreement with the overall existing knowledge of hydroclimate extremes in the study region. Our analyses provide new visual insights on the propagation, connectivity, and synchronicity of P extremes, as well as the SM–P coupling, in this flood-prone region, and can be readily used as a basis for event-driven predictive analytics for other regions.

1. Introduction

In late August 2017, Hurricane Harvey, a Category-4 tropical storm made landfall in the Gulf Coast of Texas, US. With peak accumulations of 51.88 in (1318 mm), Harvey became the wettest tropical cyclone on record in the contiguous United States (CONUS) (NCEP, 2017), and caused catastrophic flood damage and economic losses in southeastern Texas. Globally, flood is the most prevalent and damaging natural hazard, affecting more regions than any other types of natural hazard (Coumou and Rahmstorf, 2012; Easterling et al., 2000; Huntington, 2006; Kunkel et al., 2003; Seneviratne, 2012; Winsemius et al., 2015). The warming climate, steady growth of population, and rapid development of watersheds are likely to induce further changes in the distribution, severity and frequency of floods (Luke et al., 2017; Milly et al., 2002; Peduzzi, 2017), aggravating flood risks in flood-prone regions, especially along coastlines (Syvitski et al., 2009).

Extensive work has been done in the understanding of flood generation mechanisms (Berghuijs et al., 2016; Villarini and Smith, 2010), flood forecasting (Demargne et al., 2014; Komma et al., 2008; Restrepo

et al., 2012; Wu et al., 2014; Yuan et al., 2014), development of early warning systems (Alfieri et al., 2013), and flood risk management (Brouwer et al., 2007; Few, 2003). In particular, significant interests exist in using both process-based modeling and statistical analyses to identify mechanisms in the atmosphere, land surface, surface waters, build environment, and vadose zone that can potentially affect flood characteristics and distributions. Traditional flood frequency analyses typically extract events (e.g., maximum annual flood) from the site of interest and then fit a univariate probability distribution to the data (Reiss et al., 2001). Disparity among sites is mainly analyzed through qualitative examination. Villarini and Smith (2010), for instance, analyzed the flood peaks of more than 500 streamgauge stations in the eastern US, fitted a generalized extreme value distribution to annual flood peaks of each station, and quantified upper tail properties of flood distributions. They observed “striking spatial heterogeneity” in flood distributions over the region, which was attributed to the dependence of hydrologic and hydraulic responses on basin physiographic properties, as well as to the spatially varying mixtures of flood-generating mechanisms. Berghuijs et al. (2016) compared the regional differences

* Corresponding author.

E-mail address: alex.sun@beg.utexas.edu (A.Y. Sun).

in the process control of flooding responses by using daily streamflow and meteorological records from 420 Model Parameter Estimation Experiment (MOPEX) catchments for the period 1948–2001. Using mean and variance of the date of occurrence as statistics, they considered concurrence of maximum annual flow and its four predictors (i.e., maximum daily precipitation, maximum weekly precipitation, precipitation excess, and snowmelt), and found that precipitation excess, snowmelt, and rain-on-snow events are better predictors of the flooding responses than rainfall characteristics alone. Merz and Blöschl (2008) argued that in lieu of extracting flood peak samples obtained for a single site of interest, “much better use should be made of the wealth of hydrological knowledge gained in the past century” by expanding temporal, spatial, and causal information.

Indeed, various forms of information expansion and fusion have been adopted either explicitly or implicitly in recent hydrologic studies to quantify complex couplings among hydrometeorological variables and explain their roles in climate extreme events, such as concurrences of heatwaves and meteorological droughts (Hao et al., 2013; Mazdiyasi and AghaKouchak, 2015), the compound climate events and their societal impact (Leonard et al., 2014), and development of multivariate drought indices using precipitation (P), soil moisture (SM), or potential evapotranspiration time series (Hao and AghaKouchak, 2013; Liu et al., 2017b; Vicente-Serrano et al., 2010). The SM–P coupling belongs to one of such widely studied topics.

Antecedent SM and the storage of soil reservoirs at the beginning of a flood event are important for reliable flood forecasts (Komma et al., 2008). Soil moisture content, through its role in the evapotranspiration process, also serves as a key control in the partitioning of radiative energy into latent and sensible heat fluxes, thus affecting air temperature, boundary-layer stability, and possibly P. The nature of SM–P coupling, however, is still under debate (Seneviratne et al., 2010). Previous ensemble land surface modeling experiments (Koster et al., 2004) suggested that strong SM–climate coupling may exist in transitional zones between dry and wet climates. D’Odorico and Porporato (2004) provided theoretical and experimental evidence to support the hypothesis that in continental regions summer SM anomalies affect the probability of occurrence of the subsequent P. For each summer day (May 1–September 30) between 1980 and 2002, they calculated rainfall frequency in the subsequent 21 days and then related it to the regional average SM. Tuttle and Salvucci (2016) used Granger causality test, a multivariate linear regression analysis in essence, to estimate the causal relationship between SM and occurrence of the next-day P over CONUS. Their analysis was done using 9 years of remotely sensed SM and gauge-based P products at daily steps and a spatial resolution of 0.25°. They found that SM anomalies significantly influenced rainfall probabilities over 38% of CONUS. Liu et al. (2017a) evaluated the sensitivity of runoff to climate variables using Global Land Data Assimilation System (GLDAS) and a regional climate model. Again, they derived multiple hydroclimatic indices using only co-located series in each grid cell.

Because of the convection or other atmospheric dynamics, hydroclimatic covariability is often not limited to co-located events, necessitating the search for spatial patterns beyond single sites by looking at many-to-many correlation among all potentially connected sites. Such is a form of spatial information expansion referred to by Merz and Blöschl (2008). Geostatistical methods (e.g., kriging) have been widely used to analyze spatiotemporal distribution of variables, but they are parametric and require specification of covariance models. Empirical orthogonal decomposition (EOF) is another often used method for studying spatial and temporal patterns in gridded data sets, using which spatiotemporally varying climate variables are expanded as a linear combination of eigenmodes for explaining climate teleconnection phenomena, as well as for comparing climate observations to model simulations (Hannachi et al., 2007). An alternative approach for analyzing spatial and temporal processes is the complex network theory, which may be regarded as an extension of the classic graph theory to complex systems with high structural heterogeneity and inherently

dynamic properties (Newman, 2003). Complex network theory provides a powerful framework to dissect interdependence and connection strength of physical processes while retaining only the most significant network topology in the analyses, thus helping to reveal new visual evidence on embedded driving mechanisms that are otherwise not extractable through conventional analytical tools. Although the contemporary complex network theory bears many similarities to the classic graph theory that has long been used in hydrology, it provides a flexible computational framework for experimenting with many similarity/dissimilarity measures originated from fields like information theory and statistical physics. Importantly, more efficient algorithms have been developed to cope with large-scale networks arising from Big Data analytics.

In recent years, the complex network analysis has been applied to quantifying multivariate and multiscale connectivity in hydroclimate. Tsonis et al. (2006) and Donges et al. (2009b) constructed the so-called “climate networks” using climate reanalysis data. In their studies, the network nodes (or vertices) were identified with grid cells of climate data sets, and links between nodes (edges) were established using a similarity measure between times series associated with different grid cells; only the strongest links, filtered using either a predefined link density or correlation cutoff threshold, were retained in the final networks to help uncover the most important network characteristics. Using linear correlation analysis, Steinhäuser et al. (2011, 2012) constructed climate networks from reanalysis data and demonstrated their potential to capture regional and global dependence structures within and among climate variables. Sun et al. (2015) quantified spatial connectivity of observed terrestrial water storage anomalies from Gravity Recovery and Climate Experiment (GRACE) satellite data and suggested that network measures may be used as additional spatial constraints when calibrating global hydrological models, such that network statistics obtained using simulated data should match the same statistics obtained from observed data. Fan et al. (2017) constructed climate networks to identify regions that are most significantly affected by El Niño/La Niña events. In addition to linear correlation measures, nonlinear similarity measures have been used to explore uni- or multivariate pairwise relationships, such as mutual information (Donges et al., 2009b) and transfer entropy (Runge et al., 2015). Kriging, EOF, and complex network analysis all operate on certain similarity matrices (either covariance or adjacency matrix) constructed from gridded data sets. Thus, complex network analysis is applicable wherever the other two methods are applicable. However, a subtle but important difference is that even when linear similarity measures (e.g., Pearson correlation) are used, network-based analyses may still recover characteristics of nonlinear dynamic systems (e.g., scale free property) that are otherwise not identifiable using linear approaches such as EOF (Tsonis et al., 2006).

Many hydrologic analyses involve identification and extraction of discrete events (e.g., flood events) from continuous time series and then quantifying concurrence rates. Nonparametric event-based synchronicity measures, originally introduced in neuroscience to quantify event synchronicity in electroencephalogram (EEG) signals (Kreuz et al., 2013; Quiroga et al., 2002), have been combined with the complex network analysis to quantify spatial and temporal patterns of rainfall extremes. Given any two event series in binary form (i.e., 1 if an event occurs at a measurement time and 0 otherwise), event synchronization (ES) counts the number of concurrent events in a sliding window which, after being normalized by the total number of events, gives a simple nonparametric measure of synchronicity. ES circumvents some difficulty and/or ambiguity associated with the conventional correlation analysis, such as handling non-normality and nonlinearity in data. Malik et al. (2012) used ES to analyze extreme precipitation events in Indian Monsoon System, and Boers et al. (2013) applied ES and complex network analysis to characterize extreme precipitation synchronicity in South America Monsoon System. Konapala and Mishra (2017) used ES to study spatiotemporal evolution of droughts in 344 climatic

divisions of CONUS. Because ES depends on properties beyond the second statistical moment, it can be used to examine causal relationships (Quiroga et al., 2002). Recently, Donges et al. (2016) and Siegmund et al. (2017) introduced a similar event coincidence analysis (ECA) measure, using which precursor (trigger) coincidence rate may be calculated to reflect the probability of one type of events preceding (succeeding) another, thus complementing the ES measure which does not differentiate the order of events. Sun et al. (2017) applied ECA to quantifying concurrences of hydroclimate extremes in global river basins using GRACE data.

The main purposes of this study were to (a) quantify the spatial and temporal patterns of P and SM extremes with a focus on wet events, and (b) identify potential coupling between SM–P using ECA. Few network studies have analyzed the covariation pattern of hydrometeorological variables, which may span over multiple spatial and temporal scales. Knowledge of the SM–P covariation pattern is useful for flood mitigation and drought management, although we mainly focused on the former in this analysis. The selected area of study is the State of Texas in US, which is a geographic region possessing several unique and important attributes: (a) its climate divisions span a wide spectrum, from the semiarid west to the humid east; (b) it has produced some of the heaviest rainfall accumulations and flood magnitudes in the history of the US and the world (Smith et al., 2000); (c) the state is at the forefront of water-energy nexus studies, where increased energy production activities in semiarid west Texas are perceived to compete for limited water supplies that are already under stress due to drought, urbanization, and irrigated agriculture (Scanlon et al., 2014); and yet (4) seasonal predictability of rainfall and streamflow is generally low for most of Texas (Houborg et al., 2012; Sun et al., 2014; Zhang et al., 2006). This study utilized more than 30 years of daily precipitation data from the North American Land Data Assimilation System Phase 2 (NLDAS-2), which is a gauge-only precipitation product over the NLDAS grid (0.125°) and includes an orographic adjustment (Rui and Mocko, 2017). For SM, gridded model outputs from NLDAS-2 were used. This paper is organized as follows: Section 2 provides a brief description of the study area, Section 3 describes data and methodology, Section 4 shows the results of analyses, followed by discussion and conclusions.

2. Study area

Texas, covering an area of 690,000 km², is the largest state in CONUS by area and the second largest state by economic outputs (BEA, 2015) (Fig. 1(a)). Texas is divided into 10 climate subdivisions. Mean annual precipitation ranges from 355 mm in its semiarid west to 1187 mm in the humid east (PRISM, 2015), and the annual precipitation distribution (Fig. 1(b)) shows a clear north–south stripe pattern across the state. Average annual temperature gradually increases from 11° C in the northern Panhandle to 20° C in Lower Rio Grande that is located in the southmost part of the state. Gulf of Mexico is the predominant geographical feature affecting the state's climate, moderating seasonal temperatures along the Gulf Coast and providing the major source of precipitation for most of the state, except for the Trans-Pecos and Panhandle regions of Texas, where precipitation originates mostly from the eastern Pacific Ocean and from land-recycled moisture (TWDB, 1967).

Annual precipitation in Texas exhibits a distinctive bimodal temporal pattern. Spring is the wettest season in most of the Texas, with April and May the wettest months. A secondary peak of rainfall occurs in September and October (Slade and Patton, 2003). Tropical cyclones (hurricanes and tropic storms) typically occur in late summer or early fall, and those that made landfall have produced large depths of rainfall over wide areas of the state. Balcones Escarpment, an area of steep elevation gradient in central Texas at the boundary between the Edwards Plateau and the Gulf Coast Plain (labeled in Fig. 1(a)), assists in uplifting of air masses and formation of storms, and have produced many of the largest storms in the state (Slade and Patton, 2003; Smith et al., 2000).

Texas has suffered greatly from severe storms throughout its history. Flooding is common during spring and early fall. The maximum known discharges typically range from about 1.5–3 times greater than 100-year discharges for sites in the western and eastern parts of the state, but documented discharges for some sites along the Balcones Escarpment have been as much as 4 or 5 times greater than the 100-year peak discharges, causing severe losses to those establishments existing outside the 100-year floodplain but within floodplains of maximum floods (Slade and Patton, 2003). As a result, Texas has consistently suffered the most deaths and damage from flooding than any other state (Brody et al., 2008; Sharif et al., 2014). According to Federal Emergency Management Agency (FEMA) statistics on flood insurance payments from January 1978 to June 2017, Texas alone accounted for USD 6.9 billion in property losses, only second to Louisiana (<https://bsa.nfipstat.fema.gov/reports/1040.htm>). The recent Hurricane Harvey is expected to significantly increase that amount. Thus, the complex climatic and geographic settings of Texas represent a unique complex system with high structural heterogeneity, making it well suited for demonstrating the event-based complex network analysis. Insights gained from this study shall be beneficial not only to local emergency managers, but also to other coastal regions prone to flooding.

3. Data and methods

3.1. Data

Gridded P and SM data were acquired from NLDAS-2 (<https://ldas.gsfc.nasa.gov/nldas/NLDAS2model.php>), which includes four land surface models (i.e., Noah, Mosaic, SAC, and VIC) and covers CONUS with high spatial (0.125°) and temporal (hourly) resolution. Period of coverage includes a 30-year retrospective run from January 1, 1979 to December 31, 2008, augmented with a real-time extension from January 1, 2009 to the present (Xia et al., 2012). NLDAS precipitation field is synthesized using a number of rainfall products, including daily gauge, hourly radar, and global satellite observations. Non-precipitation forcing fields are derived from the analysis fields of North American Regional Reanalysis (Mesinger et al., 2006).

NLDAS forcing and output have been validated and assessed by using measurements collected from selected observation networks. For example, NLDAS-2 daily SM outputs were validated against the long-term in situ datasets (i.e., the North American Soil Moisture Database), including Illinois SM database, Oklahoma Mesonet SM dataset, and data from a total of 121 sites of the Soil Climate Analysis Network (SCAN) covering CONUS (Xia et al., 2014). The study of Xia et al. (2014) used in situ SM data from 1997–2002. Xia et al. (2015) further expanded the evaluation period to 13 years (1999–2012). Overall, all four NLDAS-2 land surface models were shown to capture the broad features of observed SM variations (i.e., the seasonal cycle and interannual variability), reproducing the daily, monthly, and annual SM anomalies well and capturing most of the wet and dry events at different soil depths (Xia et al., 2014). In their P and SM causality analysis, Tuttle and Salvucci (2016) compared the remotely sensed data (top 1 cm SM) from Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), a sensor deployed on NASA's Aqua satellite, to NLDAS-2 simulation results (0–10 cm SM simulated by Noah and Mosaic); they reported that the two products generally gave similar feedback patterns.

The period of study in this work ranges from January 1, 1979 to December 31, 2014. Previously, the model validation study by Xia et al. (2014) showed that for the top 10 cm, Noah (SAC) has the smallest (largest) mean absolute error and root-mean-square error, while Mosaic and VIC are in between. Noah is currently being used as the land component for National Centers for Environmental Prediction's (NCEP) operational model systems, including Global Forecast System, Climate Forecast System, North American Modeling System, and other regional numerical model systems. It is important to understand the

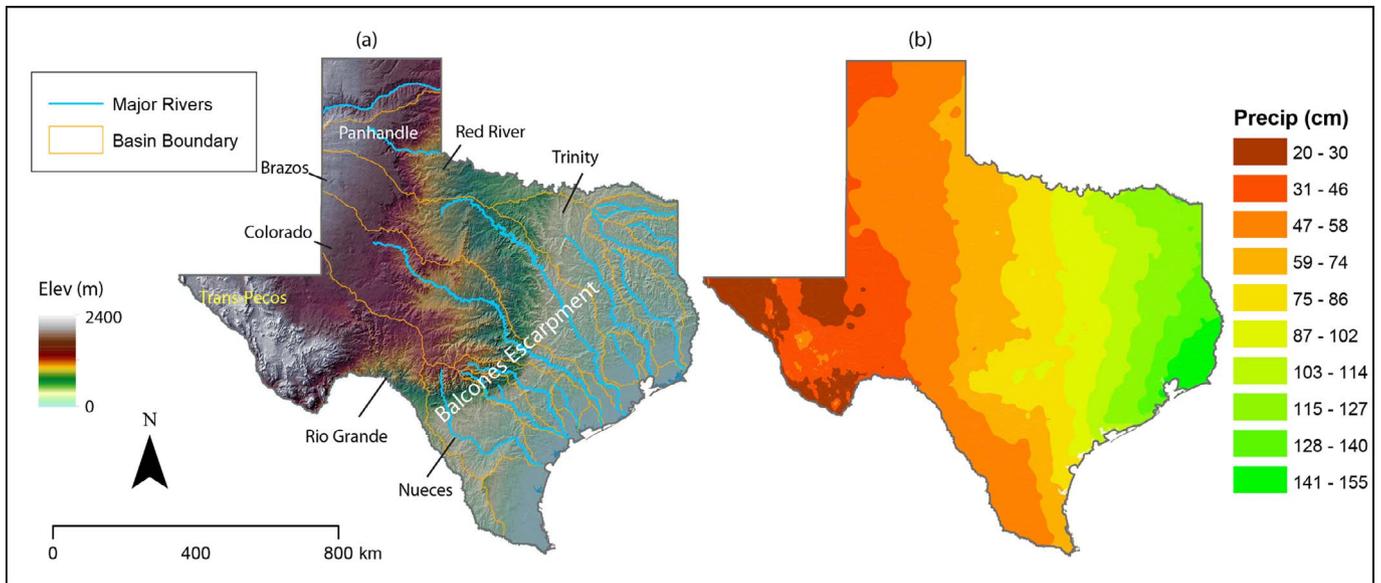


Fig. 1. (a) Map of Texas. Elevation decreases from northwest towards southeast. All major rivers (blue) originate from within the state. Major basins (basin boundary in golden) mentioned in the discussion are labeled. Balcones Escarpment (labeled in white), a narrow area of steep elevation changes in central Texas, is historically associated with extreme flooding events; (b) Annual precipitation pattern in Texas exhibits a distinctive north–south stripe pattern (source: PRISM Climate Group). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

limitations and strengths of Noah for improving weather and climate prediction skills. Thus, simulated 0–10 cm SM data from Noah were mainly used for quantifying SM–P coupling. Results from VIC and Mosaic are provided as part of the supporting materials. The total number of valid NLDAS grid cells for Texas is 4112. Downloaded hourly P and SM data were first aggregated into daily steps. The daily step is short enough for identification of low-frequency, high-magnitude wet events. To explore seasonal patterns, the daily time series corresponding to each grid cell were split into four seasons, December–February (DJF), March–May (MAM), July–August (JJA), and September–November (SON). Data processing was done using Python scripts on a computer cluster operated by the Texas Advanced Computing Center at The University of Texas.

3.2. Event-based complex network analysis

3.2.1. Event synchronization (ES)

General workflow of the event-based complex network analysis used in this study consists of event extraction, event synchronization quantification, and network construction (Fig. 2). During event extraction, events of interest are identified according to certain criteria (e.g., the top 10% daily rainfalls in a season). Manual event identification has also been used, as done in Ford et al. (2015). After forming event series, pairwise event synchronization is then performed.

Let $S^{(A)} = \{t_i\}$, ($t_i < t_{i+1}$, $i = 1, \dots, N_A$) denote an ordered event series A containing N_A events, where t_i denotes the time index (an integer) of the i th event. Similarly, $S^{(B)} = \{t_j\}$, ($j = 1, \dots, N_B$) denotes another ordered event series B. The two event series may represent either events of the same type (e.g., P) but occurred at different locations, or co-located events of different types (e.g., P and SM). A coincidence indicator $C_{ij}^{(A,B)}$ between $S^{(A)}$ and $S^{(B)}$ is defined as (Kreuz et al., 2013)

$$C_{ij}^{(A,B)} = \begin{cases} 1, & \text{if } |t_i^{(A)} - t_j^{(B)}| \leq \tau_{ij}^{(A,B)} \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

in which $t_i^{(A)} - t_j^{(B)}$ is the time delay between any event pair from each of the two event series, and $\tau_{ij}^{(A,B)}$ defines the width of a sliding window for determining whether the two events are associated with each other

$$\tau_{ij}^{(A,B)} = \frac{1}{2} \min(t_{i+1}^A - t_i^A, t_i^A - t_{i-1}^A, t_{j+1}^B - t_j^B, t_j^B - t_{j-1}^B), \quad (2)$$

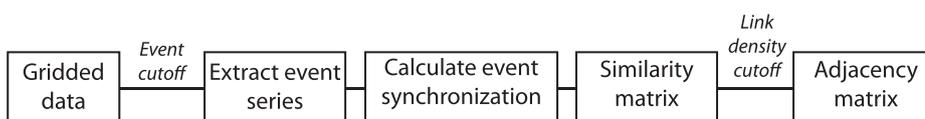
in which the factor 1/2 is used to avoid double counting if two events in $S^{(A)}$ are close to the same event in $S^{(B)}$ (Quiroga et al., 2002). Because intervals between events indirectly reflect the event occurrence rate, $\tau_{ij}^{(A,B)}$ allows the coincidence indicator to adapt to the local rate of event occurrence. By allowing such dynamic delays between events, the ES measure avoids specification of a fixed lag between time series, which can often be ambiguous for climate series (Boers et al., 2013). An upper bound τ_{\max} can be imposed such that the window width is $\hat{\tau}_{ij}^{(A,B)} = \min\{\tau_{ij}^{(A,B)}, \tau_{\max}\}$. The ES measure is then simply calculated by summing over $C_{ij}^{(A,B)}$ and normalizing the result by the total number of events from both series,

$$ES = \frac{1}{M} \sum_{ij} C_{ij}^{(A,B)} \quad (3)$$

where $M = N_A + N_B$ is the total number of events. The nonlinear ES measure as given in Eq. (1) is symmetric.

In this work, ES was used to quantify similarity between all pairs of grid cells, yielding an $N_c \times N_c$ square similarity matrix, where N_c is the total number of grid cells (because of symmetry, only half of the calculations are necessary). For each P time series, an event series was extracted by considering the top 10% (5%) of all daily values, resulting in about 328 (165) events per grid cell per season. The similarity matrix was obtained by using the Python package, pypike (Mulansky, 2015), with a τ_{\max} value of 3 days.

Fig. 2. Workflow of network construction.



3.2.2. Network measures

Climate networks can be constructed from the similarity matrix, in which grid cells are treated as network nodes, and a pair of nodes are said to be connected if there is a significant link between them. For the latter purpose, a preset link density (e.g., 0.02) is usually set to help uncover the most significant features in the network (Donges et al., 2009a). In this approach, the ES values of all pairs of event series are first sorted in an increasing order to generate an empirical cumulative distribution function (ECDF) of ES, and a cutoff ES value (denoted as ES_T) is then found for the corresponding percentile (e.g., 98th percentile for a link density of 0.02). Only links having ES values greater than the ES_T are retained for network construction. The higher the ES_T is (or equivalently, the smaller the link density), the less likely that two nodes are linked due to spurious correlation. To test the statistical significance of the chosen link density, a procedure similar to that in Boers et al. (2013) and Konapala and Mishra (2017) was used. Specifically, 5000 independent event series were generated with 328 (165) independent and uniformly distributed random events in each; the ES values of all pairs of event series were then calculated, and used to derive a test ECDF for comparison with the original input data. The null hypothesis is that the randomly generated test set has a similar ES value to ES_T at the same link density. If ES_T is significantly larger, then the null hypothesis is rejected. The link trimming process described here leads to a very sparse, binary adjacency matrix, with elements equal to 1 if the corresponding grid cells are linked and 0 otherwise. Calculation and storage of similarity matrices is often the most time-consuming step in network construction.

The following climate network measures were used in this study to characterize the topology of networks.

Degree centrality (DC) of a node is the number of directly connected neighbors (also known as the first neighbors in literature) the node has,

$$DC_i = \sum_{j \in n_i} a_{ij}, \quad (4)$$

where n_i is the set of directly connected neighbors of node i and a_{ij} are elements of the adjacency matrix. Nodes with a high DC are called super nodes and clusters of such super nodes are referred to as hubs.

Betweenness centrality (BN) of a node is the fraction of shortest paths connecting any two nodes that also pass through the node,

$$BN_i = \frac{\sum_{k,l \neq i} \sigma_{kl}(i)}{\sum_{k,l \neq i} \sigma_{kl}} \quad (5)$$

where σ_{kl} denotes the number of shortest paths between any pair of nodes k and l , and $\sigma_{kl}(i)$ includes only the shortest paths that pass through i . A node with a high BN value indicates the importance of the node in mediating transport of network attributes (e.g., traffic, information, or mass). In climate networks, regions exhibiting high BN values may represent pathways for the long-ranged, directed propagation of extreme climate events (Boers et al., 2013).

Mean link distance (LD) is the average length of links that each node has, which is normalized by node degree centrality

$$LD_i = \frac{1}{DC_i} \sum_{j=1}^{N_c} a_{ij} dist(i, j) \quad (6)$$

where DC_i is the DC of node i , N_c is the number of nodes, and $dist(i, j)$ is the geographical distance between nodes i and j , and is defined as

$$dist(i, j) = R \cos^{-1}(\sin \lambda_i \sin \lambda_j + \cos \lambda_i \cos \lambda_j \cos \Delta_{ij}) \quad (7)$$

in which λ_i and λ_j are latitudes of the nodes, Δ_{ij} is the absolute difference in longitudes, and R is Earth radius. High LD values indicate regions of high synchronicity with distant nodes. Unlike DC which is a local measure of the directly connected neighborhood, both BN and LD are global measures of network topology that rely on shortest link lengths.

The artificially imposed boundary (i.e., the Texas state border in the current case) may affect the accuracy of calculated network measures because connections to nodes outside the boundary are cut off. For DC

measure, we simply expanded the study area to include significantly larger areas outside Texas, whereas for BN and LD measures, we adopted a more sophisticated procedure suggested in Rheinwalt et al. (2012) and recently in Wiedermann et al. (2016): for each season an ensemble of random surrogate networks are generated that preserve the global and local link distributions between nodes, hence explicitly taking into account the spatial embedding of the network in some metric space. The underlying assumption is that links that are clipped due to the boundary effect can be randomly “re-wired” inside the boundary using network statistics derived from the existing network. The ensemble mean of all surrogate networks is then used to correct for the boundary effect. For example, the corrected BN is expressed as the ratio between BN of the input network and the ensemble mean BN. In this work, an ensemble of 300 surrogate networks were used for each season. Network measures were computed using the pyunicorn package developed by Donges et al. (2015) which, in turn, calls the open-source graph package, igraph (Csardi and Nepusz, 2006), for network calculations.

3.2.3. Community detection

The problem of community detection refers to the partition of a network into communities of densely connected nodes, with the nodes belonging to different communities being only sparsely connected (Blondel et al., 2008). In this study, we were interested in combining community detection with the aforementioned network measures to understand the spatial structures of covariation regions.

Community detection has gained significant interests since the seminal paper by Girvan and Newman (2002). Although a wide array of algorithms are available for community detection (Fortunato, 2010), very few of those are suitable for large networks having more than several hundreds of nodes. The particular algorithm adopted in this work is a multilevel partition algorithm proposed by Blondel et al. (2008). It optimizes a network measure called modularity (Q), an indicator of “partition correctness” in the sense that a good partition should have many edges within communities but only a few between them (Clauset et al., 2004):

$$Q = \frac{1}{2N_l} \sum_{ij} \left[a_{ij} - \frac{DC_i DC_j}{2N_l} \right] \delta(\mathcal{C}_i - \mathcal{C}_j), \quad (8)$$

where N_l is the total number of edges in a network; a_{ij} are elements of adjacency matrix, DC_i and DC_j are the degree centrality of nodes i and j , respectively; \mathcal{C}_u is the community membership of a node u and is equal to 1 if u is part of it; and the delta-function $\delta(\mathcal{C}_u - \mathcal{C}_v)$ is 1 if nodes u and v are in the same community and 0 otherwise. Eq. (8) suggests that Q is closely related to degree centrality, with higher Q values indicating better community partitions. To maximize the value of Q , the multilevel partition algorithm operates iteratively in two stages. In the first stage, every node in the network is assigned a separate community and the nodes are moved among communities iteratively in a way that maximizes each node’s local contribution to the overall modularity. In the second stage, a new network is built whose vertices are communities identified from the first stage. The two-stage process is iterated until there are no more changes and a maximum modularity is attained. The algorithm is very efficient, making it suitable for analyzing large networks within reasonable time. We used an implementation of the multi-level partition algorithm from the igraph (Csardi and Nepusz, 2006) for this work.

3.2.4. Soil moisture–precipitation coupling

In addition to constructing extreme precipitation networks, the SM–P coupling was quantified using an event-based analysis measure, ECA. In particular, we were interested in a precursor coincidence rate that is defined between two event series, $S^{(A)}$ and $S^{(B)}$, as (Siegmund et al., 2017)

$$r_p = \frac{1}{N_A} \sum_{i=1}^{N_A} H \left(\sum_{j=1}^{N_B} I_{[0, \Delta T]}((t_i^A - \tau) - t_j^B) \right), \tag{9}$$

in which $S^{(A)}$ and $S^{(B)}$ are as defined before, $H(\cdot)$ is Heaviside function; $I_{[0, \Delta T]}(\cdot)$ is an indicator function, which is equal to 1 when its argument falls within $[0, \Delta T]$ and 0 otherwise. The parameter ΔT defines the width of a sliding window, and parameter τ allows to explicitly take into account lagged relationships between event series. By definition, r_p is in the $[0,1]$ range.

Thus, if $S^{(A)}$ represents P events and $S^{(B)}$ represents SM anomalies, then r_p is the fraction of P events that are preceded by at least one SM anomaly event. Here SM anomalies are obtained by detrending time series for each season by subtracting the seasonal means. The maximum slide window length, ΔT , was set to 7 days and the lag τ to 0 for the SM–P analysis.

4. Results and discussion

Extreme precipitation networks were generated from gridded P data from NLDAS-2, using the 90th percentile of individual grid time series as event cutoff threshold and with a link density of 0.02 (see description under Section 3.2.1). For DJF, for example, the resulting network has 171,000 links (out of $4112^2/2$) at the specified link density, which corresponds to an ES value of 0.79. The procedure described in Section 3.2.2 was used to test the significance of the specified link density by using 5000 uniformly distributed random event series. The ES value of the test set is only 0.27 at the 0.02 link density and the maximum ES value of the test set is 0.33. Thus, the retained links are considered statistically significant. Unless otherwise specified, all results reported below pertain to this set of networks.

Fig. 3 shows the resulting seasonal DC maps, which are normalized by the maximum nodal degree of each season. Significant variations in spatial DC patterns can be observed across seasons. Node clusters exhibiting high DC values (hot spots) tend to synchronize with a relatively large number of adjacent nodes during extreme rainfalls (note:

adjacency is in a similarity sense, not necessarily in a geographical sense). Consequently, those hot spot regions may be used as a proxy of high flooding risks. In winter (DJF), high-value clusters (DC > 0.4) cover the northeastern half of Texas, which extends further to the north to Oklahoma. In spring (MAM), the high-value clusters become significantly smaller in size and are mainly located in southeastern Texas. In summer (JJA), the extreme P clusters cover the entire central Texas. Finally, in fall (SON), a high-value (>0.8) DC cluster is observed near the four-state corner of Texas–Oklahoma–Arkansas–Louisiana (i.e., near the right border in Fig. 3(d)).

These DC patterns may be further interpreted using Texas’ climatology. In winter and spring, precipitation is formed when cool air from the north and from high plains of northern Mexico meet the humid air from Gulf of Mexico. In early fall, tropical cyclones from both the eastern Pacific and Atlantic Basins contribute to some of the heaviest rainfalls near the Texas Gulf Coast. In the summer, orographic ascent along the Balcones Escarpment acts an important forcing mechanism for the extreme thunderstorms observed in central Texas (Nielsen et al., 2016). The lack of strongly synchronized P clusters in MAM, the wettest season in Texas, is intriguing. This may be attributed to the lack of a dominant precipitation generation mechanism in MAM.

As a sensitivity study, DC maps generated using the 95th-percentile event cutoff is shown in Supporting Information (SI) Fig. S1. In general, the DC clusters become weaker when the P event cutoff threshold is increased, because of the smaller number of P events and subsequently less spatial synchronicity. This is especially true for DJF and MAM. In SON, however, we see that more smaller-sized extreme rainfall hot spots appear in the Brazos river basin.

In Fig. 4, BN maps derived for the same P networks are shown. Recall that BN is a measure of the importance of a network node in mediating propagation of a certain network attribute which, in the current case, may be interpreted as air mass associated with extreme P. As described under Section 3.2.2, the BN maps were corrected by using the ensemble mean of 300 surrogate networks that were generated separately for each season to compensate for the boundary effect. In

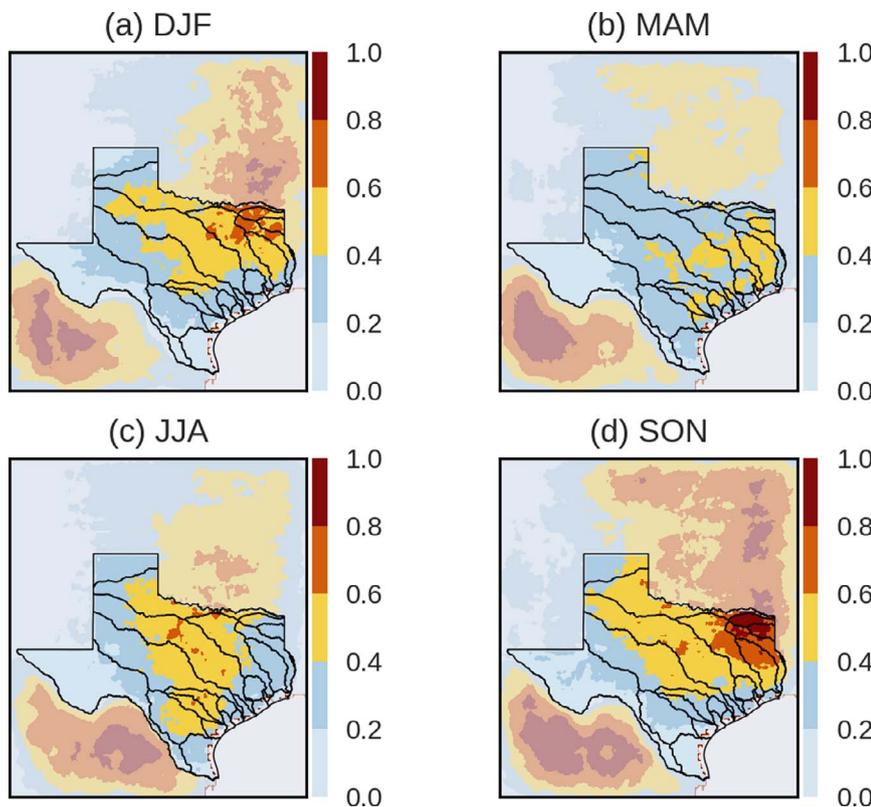


Fig. 3. Precipitation degree centrality maps for (a) DJF, (b) MAM, (c) JJA (d) and SON. At each valid grid cell, the 90th percentile value of precipitation time series (from 1979 to 2014) was used as event cutoff threshold. All P networks were constructed using a link density of 0.02.

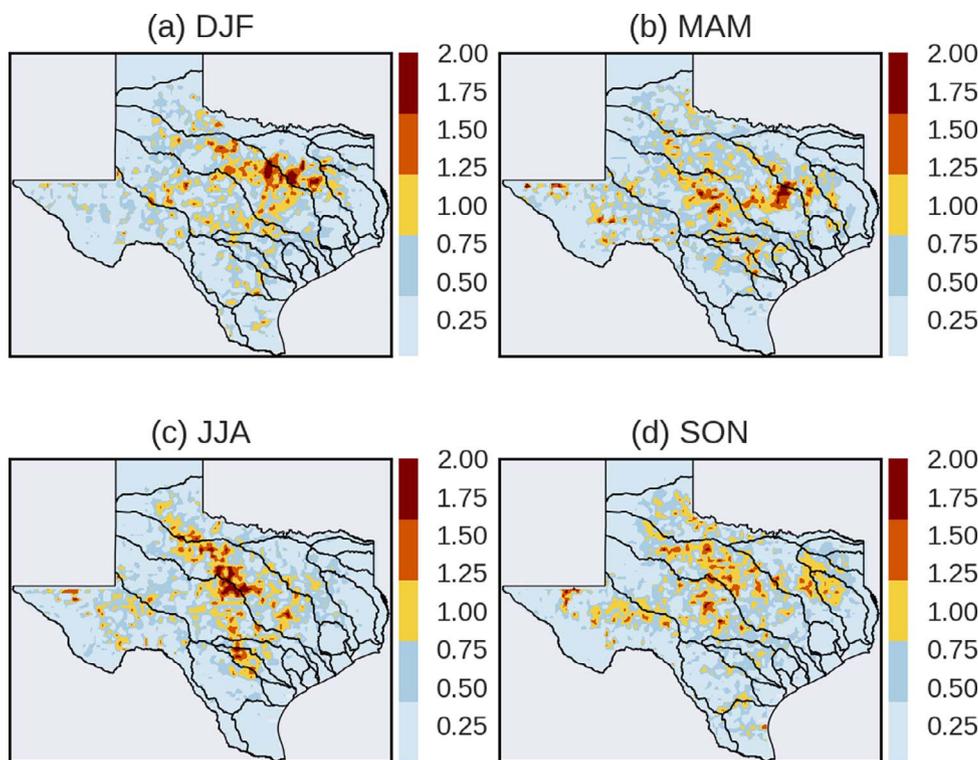


Fig. 4. Precipitation betweenness centrality maps for (a) DJF, (b) MAM, (c) JJA, and (d) SON.

DJF, strong BN values can be observed in northeastern Texas near the junction of Brazos/Trinity river basins. The pattern is rotated clockwise during JJA, with one cluster seen in north central Texas and another trace of clusters aligned with Balcones Escarpment. BNs in MAM and SON are more scattered, suggesting the randomness in heavy rainfalls in those seasons. Therefore, flooding risks in spring and fall tend to be quite widespread, an observation that is consistent with what Slade and Patton (2003) had observed on the basis of 256 major storms in the history of Texas (1850–2002). In general, many of the high-valued BN clusters appear on or near the major river basin boundaries. Specifically, the Brazos river basin acts as a major P transport pathway in most of the seasons. BN maps derived from networks that were constructed using the 95th-percentile cutoff (Fig. S2) agree well with those in Fig. 4, showing Brazos river basin as a major pathway.

The LD maps, shown in Fig. 5, measure the average shortest spatial distance between each node and its first neighbors. All high value LD clusters are concentrated in central Texas, in the middle of Brazos and Colorado river basins. Small sized clusters are observed in Balcones Escarpment in JJA and near east Texas in SON. The nodal average link distance is smaller in JJA than in other seasons, probably because of the local nature of many storms. Comparing to Fig. 4, we see co-existence of high-valued LD and BN clusters in central Texas, indicating that the region acts as a central hub for facilitating the long-range, directed moisture transport processes from different parts of the state.

At last, we identified the P community structure using the 90th percentile extreme P networks. All seasons have similar numbers of community structures, as shown in Fig. 6. The community structure shapes are relatively fixed in the north Panhandle area and in west Texas across all seasons, but vary significantly in central Texas and the Gulf Coast region. A community border is formed around the Balcones Escarpment in all seasons, except in SON. In the latter case, a golden color community covering the Houston area is seen, with its longer axis parallel to the Gulf Coast. On the other hand, an elongated north–south green color community is observed that spans from the southern most corner all the way to central Texas. These are the communities that are often influenced by tropical cyclones during SON. For example, the golden color community includes the area that was severely flooded

during the landfall of Hurricane Harvey in late August 2017.

SM patterns were examined using the same complex network analysis. Although SM does not transport as dynamically as air masses per se, its spatial distribution is influenced by multiple atmospheric, surface, and subsurface processes, including P, evapotranspiration, and infiltration. Thus, the spatial distribution of SM and its network topology inherently reflect not only the soil properties (e.g., texture and hydraulic conductivity), but also signatures of multiple physical processes. Soil property data used in NLDAS were derived from 1-km State Soil Geographic (STATSGO) database and have 11 layers and 16 texture classes. Fig. S3 shows the soil texture data for Texas.

We constructed the SM networks using the 0–10 cm SM data from Noah model of NLDAS-2. The event cutoff was again set to 90th percentile of cell-wise SM anomaly values and the link density to 0.02. The resulting SM DC maps are shown in Fig. 7. Overall, the spatial patterns of SM DC bear many similarities with the rainfall DC maps shown in Fig. 3, especially in the summer and fall seasons. In addition, high-value soil moisture DC clusters are found near the southern High Plains area (the lower Texas Panhandle), which is a region of relatively deep soils, comprising of sandy loams and clay loams. Human alterations of the surface water and energy balance have been massive in the High Plains area, as already reported in numerous previous studies (e.g., Moore and Rojstaczer, 2002; Scanlon et al., 2012). Irrigation water withdrawal at the region is over $6 \times 10^9 \text{ m}^3$ annually, an amount that is so massive that it can cool the surface and raise the latent heat flux, as well as the convective available potential energy (Moore and Rojstaczer, 2002). Although NLDAS-2 land surface models do not simulate human activities directly, a number of in situ SCAN sites are located near the lower Texas Panhandle area. Noah outputs were shown to have relatively high correlation (0.78) with in situ observations in Great Plains (Xia et al., 2014). For comparison, the SM degree maps constructed using the 0–10 cm SM data from Mosaic and VIC are given in Figs. S5 and S6, respectively. Overall, the spatial patterns of Mosaic are similar to those from Noah (except for DJF), although the DC cluster sizes obtained from Mosaic are smaller. The SM patterns from VIC, however, are significantly different from those given by Noah and Mosaic. The latter observation is consistent with Xia et al. (2014), who showed that the

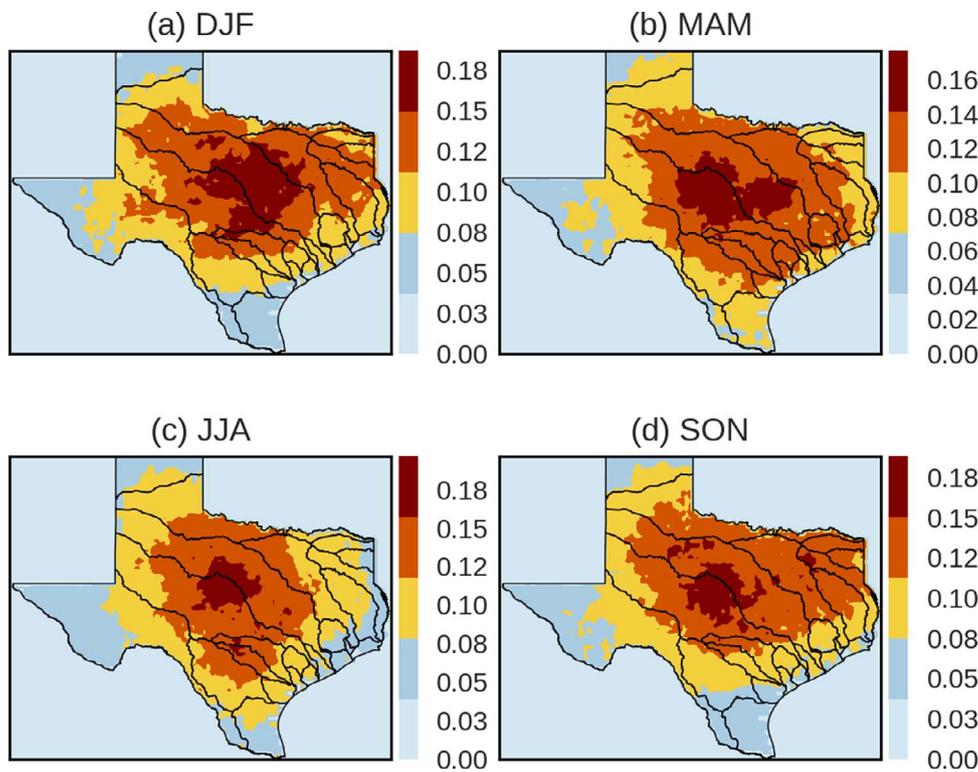


Fig. 5. Precipitation mean local distance maps for (a) DJF, (b) MAM, (c) JJA, and (d) SON.

VIC top 10 cm soil moisture has less seasonal variation when compared with the observations and the other three models, and a likely cause was attributed to the limited bare soil fractions assigned in VIC grid cells.

The combination of short-duration extreme rainfalls, high soil moisture conditions, and low-permeability soil classes may significantly aggravate the flooding risk. Thus, regions with stronger covariability in

P and SM should be evaluated during hazard mitigation planning. A question is then whether there is also coupling between P and SM in these regions, especially in the sense of SM–P feedback mechanism.

The ES measure we have used so far does not differentiate the order of events, as mentioned before. It is trivial that higher P leads to higher SM. However, SM–P coupling in the other direction (i.e., due to SM feedback) is more intriguing and is still subject to active investigations,

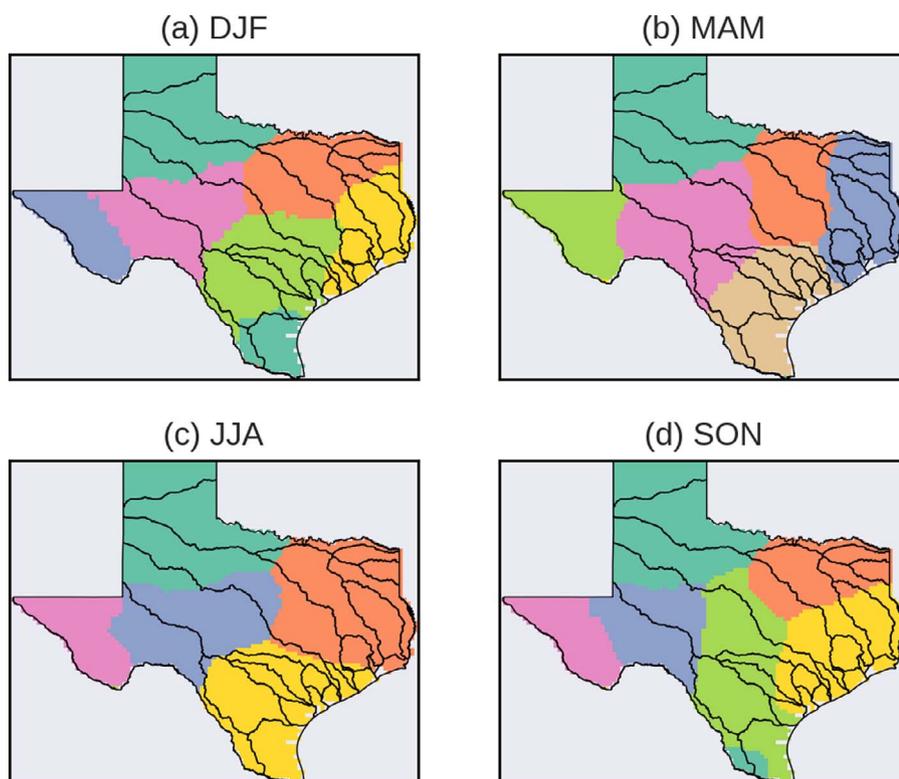


Fig. 6. Communities identified from precipitation networks for (a) DJF, (b) MAM, (c) JJA, and (d) SON. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

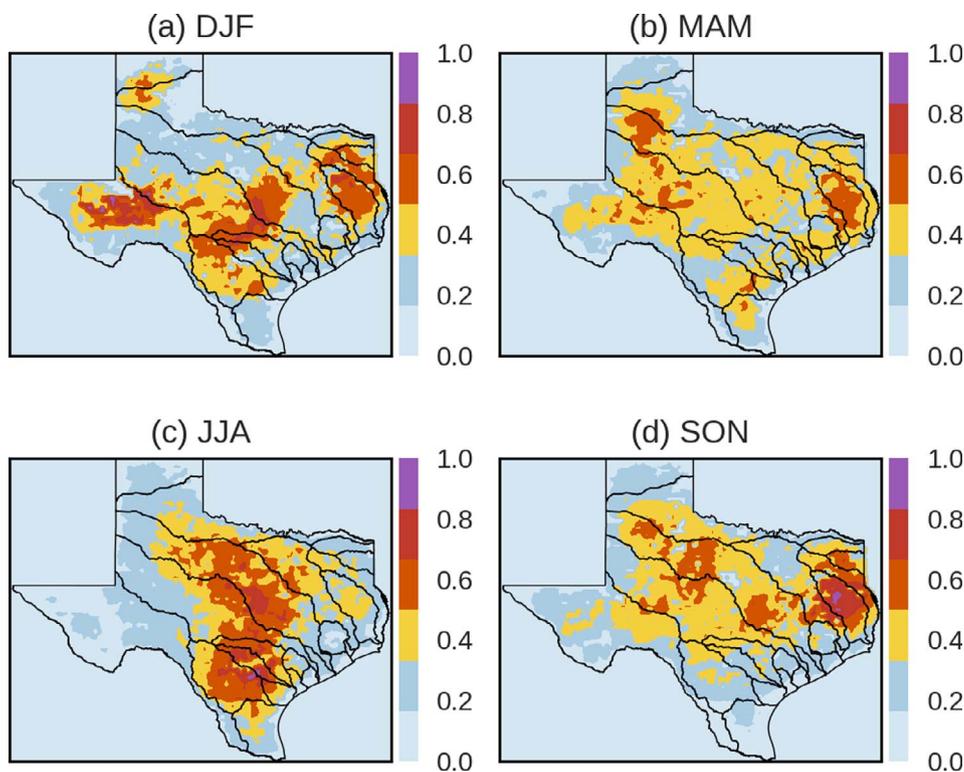


Fig. 7. Soil moisture degree centrality maps for (a) DJF, (b) MAM, (c) JJA, and (d) SON.

as mentioned in the introduction. Here we applied the ECA measure described under Section 3.2.4, specifically the precursor coincidence rate, to quantify cell-wise SM–P coupling with SM anomalies being the precursor. The absolute values of SM anomalies were used so that contributions of both dry and wet SM events to subsequent P events could be included. A statistical significance test was used to ensure that the coincidence is not purely random. The null hypothesis of the significance test is that the observed number of coincidences can be explained by two independent series of randomly distributed events following a Poisson process. If the p -value is smaller than a user-defined confidence level α , the null hypothesis can be rejected (Donges et al., 2016).

Results are presented in Fig. 8. All colored areas are significant at $\alpha = 0.05$ level. In winter, SM–P coupling is sparse; strong precursor coincidence rate (> 0.7) is mainly observed along the Texas Gulf Coast. In spring and fall, the coupling becomes more widespread, with mild coupling regions (0.5–0.6) covering almost the entire state. However, strong coupling is still limited to the coastal area. SM–P coupling seems to be the strongest in summer, the hottest and driest season in Texas, where significant coupling is seen in Lower Rio Grande river basin near the US–Mexico border and in the adjacent Nueces river basin. A previous study showed that summer precipitation in the US may be significantly contributed by local recycling, through moisture returning to the atmosphere by evapotranspiration from the same region (D’Odorico and Porporato, 2004). A more recent study tried to link in situ observations of soil moisture from more than 100 stations in Oklahoma to subsequent unorganized afternoon convective precipitation (Ford et al., 2015); the authors found a statistically significant preference for convective precipitation initiation over drier than normal soils, with over 70% of events initiating over soil moisture below the long-term median. In our case, the hot summer evapotranspiration may directly drive moisture recycling, leading to initiation of local convective precipitation. The coupling in the Texas Panhandle, part of the High Plains, is weak or none in all seasons. Interestingly, many areas exhibiting high extreme rainfall synchronicity (e.g., central Texas and Balcones Escarpment) do not exhibit strong SM–P coupling, suggesting the alternative forcing mechanisms behind the extreme P events

occurred in those areas (e.g., sharp elevation changes). The main exception is near the Texas–Louisiana border in SON, where we see high precipitation DC, soil moisture DC, and strong SM–P coupling. Indeed, that is the region known to be highly prone to flooding. As a sensitivity study, the same analysis was repeated using the 95th percentile event threshold (Fig. S4), in which case the higher threshold tends to weaken the coupling pattern, as we already saw in other network measures.

Our SM–P coupling results are generally in agreement with the findings of Tuttle and Salvucci (2016), who used observed P and SM data. For example, those authors identified the Texas Gulf Coast area as a region of strong SM–P coupling. They did not observe significant SM–P feedback in Great Plains, one of the hot spots identified in the original study of Koster et al. (2004). Our study, which uses a non-parametric event causality measure, also confirmed that little coupling exists in the Texas High Plains, which is a subregion of Great Plains. Nevertheless, model uncertainties in land surface models may affect the simulated SM, which in turn may affect the SM patterns and SM–P coupling. Future works may use in situ data and/or remotely sensed soil moisture data to further confirm the spatiotemporal patterns reported here.

5. Summary and conclusions

Extreme event covariability has received significant attention in recent years. From the perspective of flood risk and vulnerability assessment, a strong need exists to identify flooding risks in flood-prone areas, to support local decision making (Balica et al., 2013; Kauffeldt et al., 2016; Leskens et al., 2014) and mitigate the negative impact of climatic variability (Mazdiyasn and AghaKouchak, 2015). This study examined covariation patterns embedded in extreme P and its land surface covariate, SM, using an event-based, complex-network-theoretic approach. The nonlinear, nonparametric ES and ECA measures are useful for quantifying the coincidence of extreme events, circumventing limiting assumptions behind conventional linear correlation analyses. These network measures were demonstrated using high-resolution NLDAS data for Texas, which is a US state exemplifying strong water-energy nexus. Results reveal that extreme event hot spot patterns and

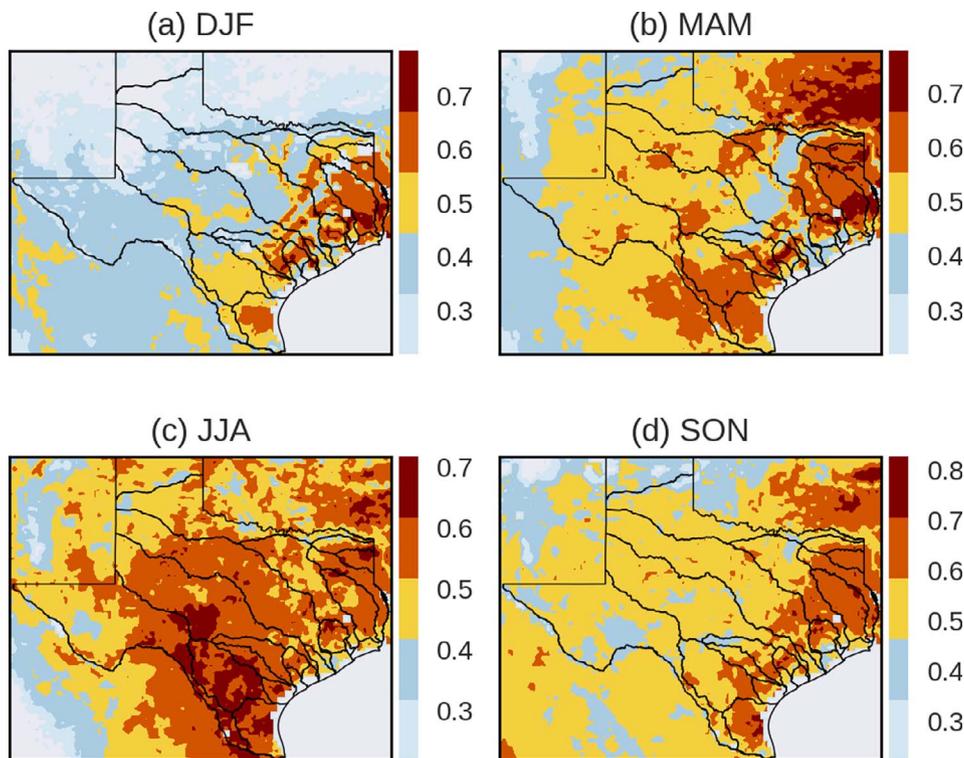


Fig. 8. Precursor coincidence rate between soil moisture and precipitation events (defined as the fraction of P events that are preceded by at least one SM event in each sliding window) for (a) DJF, (b) MAM, (c) JJA, and (d) SON.

community structures show strong seasonal patterns in Texas. Within each season, strong spatial variability also exists because of the wide range of climate divisions in the study area. The high-resolution extreme event patterns uncovered in this study may be used as a reference to direct future hazard mitigation efforts.

The capability to quickly visualize embedded connections in massive spatiotemporal data is critical as environmental scientists start to embrace Big Data technologies (Granell et al., 2016). Built on a number of network measures, we show that the complex network analysis represents a promising tool for extracting interesting features out of large gridded data sets, providing new visual analytics and supplementary information to emergency managers. Finally, it is emphasized that although the methodologies are demonstrated for Texas, they are general and can be applied to studying any other regions or other types of extreme event synchronicity.

Acknowledgment

NLDAS-2 data were downloaded from NASA Goddard Earth Sciences Data and Information Services Center (NASA/GES DISC) and have been described clearly in Section 2 and the support information. The authors are grateful to the AE and two anonymous reviewers for their comments. The first author is grateful to insightful discussions with Drs. Thomas Kreuz and Mario Mulansky (pyspike developers) and to Dr. Jonathan Donges (pyunicorn).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2017.12.019.

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* 17 (3), 1161.

Balica, S., Popescu, I., Beevers, L., Wright, N.G., 2013. Parametric and physically based modelling techniques for flood risk and vulnerability assessment: a comparison.

Environ. Model. Softw. 41, 84–92.

BEA, 2015. List of U.S. States by GDP, Bureau of Economic Analysis. accessed Jan 18, 2016.

Berghuijs, W.R., Woods, R.A., Hutton, C.J., Sivapalan, M., 2016. Dominant flood generating mechanisms across the United States. *Geophys. Res. Lett.* 43 (9), 4382–4390.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008 (10), P10008.

Boers, N., Bookhagen, B., Marwan, N., Kurths, J., Marengo, J., 2013. Complex networks identify spatial patterns of extreme rainfall events of the South American monsoon system. *Geophys. Res. Lett.* 40 (16), 4386–4392.

Brody, S.D., Zahran, S., Highfield, W.E., Grover, H., Vedlitz, A., 2008. Identifying the impact of the built environment on flood damage in Texas. *Disasters* 32 (1), 1–18.

Brouwer, R., Akter, S., Brander, L., Haque, E., 2007. Socioeconomic vulnerability and adaptation to environmental risk: a case study of climate change and flooding in Bangladesh. *Risk Anal.* 27 (2), 313–326.

Clauset, A., Newman, M.E., Moore, C., 2004. Finding community structure in very large networks. *Phys. Rev. E* 70 (6), 066111.

Coumou, D., Rahmstorf, S., 2012. A decade of weather extremes. *Nat. Clim. Chang.* 2 (7), 491–496.

Csardi, G., Nepusz, T., 2006. The iGraph software package for complex network research. *InterJournal, Complex Syst.* 1695 (5), 1–9.

Demargne, J., Wu, L., Regonda, S.K., Brown, J.D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H.D., Fresch, M., et al., 2014. The science of NOAA's operational hydrologic ensemble forecast service. *Bull. Am. Meteorol. Soc.* 95 (1), 79–98.

D'Odorico, P., Porporato, A., 2004. Preferential states in soil moisture and climate dynamics. *Proc. Natl. Acad. Sci. USA* 101 (24), 8848–8851.

Donges, J.F., Heitzig, J., Beronov, B., Wiedermann, M., Runge, J., Feng, Q.Y., Tupikina, L., Stolbova, V., Donner, R.V., Marwan, N., et al., 2015. Unified functional network and nonlinear time series analysis for complex systems science: the pyunicorn package. *Chaos Interdiscip. J. Nonlinear Sci.* 25 (11), 113101.

Donges, J.F., Schleussner, C.-F., Siegmund, J.F., Donner, R.V., 2016. Event coincidence analysis for quantifying statistical interrelationships between event time series. *Eur. Phys. J. Spec. Top.* 225 (3), 471–487.

Donges, J.F., Zou, Y., Marwan, N., Kurths, J., 2009. The backbone of the climate network. *Europhys. Lett. (EPL)* 87 (4), 48007.

Donges, J.F., Zou, Y., Marwan, N., Kurths, J., 2009. Complex networks in climate dynamics. *Eur. Phys. J. Spec. Top.* 174 (1), 157–179.

Easterling, D.R., Evans, J., Groisman, P.Y., Karl, T., Kunkel, K.E., Ambenje, P., 2000. Observed variability and trends in extreme climate events: a brief review. *Bull. Am. Meteorol. Soc.* 81 (3), 417–425.

Fan, J., Meng, J., Ashkenazy, Y., Havlin, S., Schellnhuber, H.J., 2017. Network analysis reveals strongly localized impacts of El Niño. *Proceedings of the 2017 National Academy of Sciences of the United States of America*. pp. 201701214.

Few, R., 2003. Flooding, vulnerability and coping strategies: local responses to a global threat. *Progr. Dev. Stud.* 3 (1), 43–58.

Ford, T.W., Rapp, A.D., Quiring, S.M., Blake, J., 2015. Soil moisture-precipitation coupling: observations from the Oklahoma mesonet and underlying physical mechanisms. *Hydrol. Earth Syst. Sci.* 19 (8), 3617–3631.

- Fortunato, S., 2010. Community detection in graphs. *Phys. Rep.* 486 (3), 75–174.
- Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* 99 (12), 7821–7826.
- Granel, C., Havlik, D., Schade, S., Sabeur, Z., Delaney, C., Pielorz, J., Usländer, T., Mazzetti, P., Schleidt, K., Kobernus, M., et al., 2016. Future internet technologies for environmental applications. *Environ. Model. Softw.* 78, 1–15.
- Hannachi, A., Jolliffe, I., Stephenson, D., 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review. *Int. J. Climatol.* 27 (9), 1119–1152.
- Hao, Z., AghaKouchak, A., 2013. Multivariate standardized drought index: a parametric multi-index model. *Adv. Water Resour.* 57, 12–18.
- Hao, Z., AghaKouchak, A., Phillips, T.J., 2013. Changes in concurrent monthly precipitation and temperature extremes. *Environ. Res. Lett.* 8 (3), 034014.
- Houborg, R., Rodell, M., Li, B., Reichle, R., Zaitchik, B.F., 2012. Drought indicators based on model-assimilated gravity recovery and climate experiment (grace) terrestrial water storage observations. *Water Resour. Res.* 48 (7), W07525.
- Huntington, T.G., 2006. Evidence for intensification of the global water cycle: review and synthesis. *J. Hydrol. (Amst.)* 319 (1), 83–95.
- Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., Thielen, J., 2016. Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environ. Model. Softw.* 75, 68–76.
- Komma, J., Blöschl, G., Reszler, C., 2008. Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting. *J. Hydrol. (Amst.)* 357 (3), 228–242.
- Konapala, G., Mishra, A., 2017. Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental usa. *J. Hydrol. (Amst.)* 555, 600–620.
- Koster, R.D., Dirmeyer, P.A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C., Kanae, S., Kowalczyk, E., Lawrence, D., et al., 2004. Regions of strong coupling between soil moisture and precipitation. *Science* 305 (5687), 1138–1140.
- Kreuz, T., Chicharro, D., Houghton, C., Andrzejak, R.G., Mormann, F., 2013. Monitoring spike train synchrony. *J. Neurophysiol.* 109 (5), 1457–1472.
- Kunkel, K.E., Easterling, D.R., Redmond, K., Hubbard, K., 2003. Temporal variations of extreme precipitation events in the United States: 1895–2000. *Geophys. Res. Lett.* 30 (17), 1900.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., Stafford-Smith, M., 2014. A compound event framework for understanding extreme impacts. *Wiley Interdiscip. Rev. Clim. Change* 5 (1), 113–128.
- Leskens, J., Brugnach, M., Hoekstra, A., Schuurmans, W., 2014. Why are decisions in flood disaster management so poorly supported by information from flood models? *Environ. Model. Softw.* 53, 53–61.
- Liu, D., Mishra, A.K., Zhang, K., 2017a. Runoff sensitivity over asia: role of climate variables and initial soil conditions. *J. Geophys. Res. Atmos.* 122 (4), 2218–2238.
- Liu, M., Xu, X., Xu, C., Sun, A.Y., Wang, K., Scanlon, B.R., Zhang, L., 2017b. A new drought index that considers the joint effects of climate and land surface change. *Water Resour. Res.* 53 (4), 3262–3278.
- Luke, A., Vrugt, J.A., AghaKouchak, A., Matthew, R., Sanders, B.F., 2017. Predicting nonstationary flood frequencies: evidence supports an updated stationarity thesis in the United States. *Water Resour. Res.* 53 (7), 5469–5494.
- Malik, N., Bookhagen, B., Marwan, N., Kurths, J., 2012. Analysis of spatial and temporal extreme monsoonal rainfall over south asia using complex networks. *Clim. Dyn.* 39 (3–4), 971–987.
- Mazdiyasi, O., AghaKouchak, A., 2015. Substantial increase in concurrent droughts and heatwaves in the United States. *Proc. Natl. Acad. Sci.* 112 (37), 11484–11489.
- Merz, R., Blöschl, G., 2008. Flood frequency hydrology: I. Temporal, spatial, and causal expansion of information. *Water Resour. Res.* 44 (8), W08432.
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E.H., et al., 2006. North American regional re-analysis. *Bull. Am. Meteorol. Soc.* 87 (3), 343–360.
- Milly, P.C.D., Wetherald, R.T., Dunne, K., Delworth, T.L., 2002. Increasing risk of great floods in a changing climate. *Nature* 415 (6871), 514–517.
- Moore, N., Rojstaczer, S., 2002. Irrigation's influence on precipitation: Texas high plains, USA. *Geophys. Res. Lett.* 29 (16).
- Mulansky, M., 2015. PySpike, <https://github.com/mariomulansky/pyspike> (accessed Jan 5, /2016).
- NCEP, 2017. Preliminary storm total rainfall in inches from 800 pm CDT Thu Aug 24 through 900 pm CDT Wed Aug 30, (<http://www.wpc.ncep.noaa.gov/discussions/nfdsccl.html>). accessed Sept 1, 2017.
- Newman, M.E., 2003. The structure and function of complex networks. *SIAM Rev.* 45 (2), 167–256.
- Nielsen, E.R., Schumacher, R.S., Kecklik, A.M., 2016. The effect of the balcones escarpment on three cases of extreme precipitation in central texas. *Mon. Weather Rev.* 144 (1), 119–138.
- Peduzzi, P., 2017. Flooding: prioritizing protection? *Nat. Clim. Chang* 7 (9), 625.
- PRISM, 2015. Us annual rainfall patterns, <http://www.prism.oregonstate.edu/gallery/>. accessed Jan 11, 2016.
- Quiroga, R.Q., Kreuz, T., Grassberger, P., 2002. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* 66 (4), 041904.
- Reiss, R.-D., Thomas, M., Reiss, R., 2001. *Statistical Analysis of Extreme Values*. Springer.
- Restrepo, P., Rakovec, O., Noh, S., He, M., Clark, M., Hendricks Franssen, H., Weerts, A., Liu, Y., Kumar, S., Moradkhani, H., et al., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrol. Earth Syst. Sci.* 16 (10), 3863–3887.
- Rheinwalt, A., Marwan, N., Kurths, J., Werner, P., Gerstengarbe, F.-W., 2012. Boundary effects in network measures of spatially embedded networks. *Europhys. Lett. (EPL)* 100 (2), 28002.
- Rui, H., Mocko, D., 2017. README Document for North American Land Data Assimilation System Phase 2 (NLDAS-2) Products. Technical Report. NASA Goddard Space Flight Center, (<https://hydro1.gesdisc.eosdis.nasa.gov/data/nldas/readme.nldas2.pdf>).
- Runge, J., Petoukhov, V., Donges, J.F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M., Kurths, J., 2015. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Commun.* 6.
- Scanlon, B.R., Faunt, C.C., Longuevergne, L., Reedy, R.C., Alley, W.M., McGuire, V.L., McMahon, P.B., 2012. Groundwater depletion and sustainability of irrigation in the us high plains and central valley. *Proc. Natl. Acad. Sci.* 109 (24), 9320–9325.
- Scanlon, B.R., Reedy, R.C., Nicot, J.P., 2014. Will water scarcity in semiarid regions limit hydraulic fracturing of shale plays? *Environ. Res. Lett.* 9 (12), 124011.
- Seneviratne, S.I.C.-a., 2012. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge University Press, pp. 109–230.
- Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orłowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci. Rev.* 99 (3), 125–161.
- Sharif, H.O., Jackson, T.L., Hossain, M.M., Zane, D., 2014. Analysis of flood fatalities in texas. *Nat. Hazard. Rev.* 16 (1), 04014016.
- Siegmund, J.F., Siegmund, N., Donner, R.V., 2017. CoinCalc – a new R package for quantifying simultaneities of event series. *Comput. Geosci.* 98, 64–72.
- Slade, R.M., Patton, J.M., 2003. Major and catastrophic storms and floods in Texas: 215 major and 41 catastrophic events from 1953 to September 1, 2002, USGS Open-File Report 2003–193.
- Smith, J.A., Baeck, M.L., Morrison, J.E., Sturdevant-Rees, P., 2000. Catastrophic rainfall and flooding in texas. *J. Hydrometeorol.* 1 (1), 5–25.
- Steinhauser, K., Chawla, N.V., Ganguly, A.R., 2011. Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science. *Stat. Anal. Data Min. ASA Data Sci. J.* 4 (5), 497–511.
- Steinhauser, K., Ganguly, A.R., Chawla, N.V., 2012. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Clim. Dyn.* 39 (3–4), 889–895.
- Sun, A., Chen, J., Donges, J., 2015. Global terrestrial water storage connectivity revealed using complex climate network analyses. *Nonlinear Process. Geophys. Discuss.* 2, 781–809.
- Sun, A.Y., Scanlon, B.R., AghaKouchak, A., Zhang, Z., 2017. Using grace satellite gravimetry for assessing large-scale hydrologic extremes. *Remote Sens. (Basel.)* 9 (12), 1287.
- Sun, A.Y., Wang, D., Xu, X., 2014. Monthly streamflow forecasting using gaussian process regression. *J. Hydrol. (Amst.)* 511, 72–81.
- Syvitski, J.P., Kettner, A.J., Overeem, L., Hutton, E.W., Hannon, M.T., Brakenridge, G.R., Day, J., Vörösmarty, C., Saito, Y., Giosan, L., et al., 2009. Sinking deltas due to human activities. *Nat. Geosci.* 2 (10), 681–686.
- Tsonis, A.A., Swanson, K.L., Roebber, P.J., 2006. What do networks have to do with climate? *Bull. Am. Meteorol. Soc.* 87 (5), 585–595.
- Tuttle, S., Salvucci, G., 2016. Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States. *Science* 352 (6287), 825–828.
- TWDB, 1967. *The Climate and Physiography of Texas: Texas Water Development Board Report 53*. Technical Report. TWDB.
- Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim.* 23 (7), 1696–1718.
- Villarini, G., Smith, J.A., 2010. Flood peak distributions for the eastern united states. *Water Resour. Res.* 46 (6).
- Wiedermann, M., Donges, J.F., Kurths, J., Donner, R.V., 2016. Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes. *Phys. Rev. E* 93 (4), 042308.
- Winsemius, H.C., Aerts, J., van Beek, L.P., Bierkens, M.F., Bouwman, A., Jongman, B., Kwadijk, J.C., Ligtoet, W., Lucas, P.L., van Vuuren, D.P., et al., 2015. Global drivers of future river flood risk. *Nat. Clim. Chang* 6 (4), 381–385.
- Wu, H., Adler, R.F., Tian, Y., Huffman, G.J., Li, H., Wang, J., 2014. Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* 50 (3), 2693–2717.
- Xia, Y., Ek, M.B., Wu, Y., Ford, T., Quiring, S.M., 2015. Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. part I: comparison and analysis. *J. Hydrometeorol.* 16 (5), 1962–1980.
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., et al., 2012. Continental-scale water and energy flux analysis and validation for the North American land data assimilation system project phase 2 (NLDAS-2): 1. intercomparison and application of model products. *J. Geophys. Res. Atmos.* 117, D03109.
- Xia, Y., Sheffield, J., Ek, M.B., Dong, J., Chaney, N., Wei, H., Meng, J., Wood, E.F., 2014. Evaluation of multi-model simulated soil moisture in NLDAS-2. *J. Hydrol. (Amst.)* 512, 107–125.
- Yuan, X., Wood, E.F., Liang, M., 2014. Integrating weather and climate prediction: toward seamless hydrologic forecasting. *Geophys. Res. Lett.* 41 (16), 5891–5896.
- Zhang, F., Odins, A.M., Nielsen-Gammon, J.W., 2006. Mesoscale predictability of an extreme warm-season precipitation event. *Weather Forecast.* 21 (2), 149–166.