



Research papers

Time-varying network-based approach for capturing hydrological extremes under climate change with application on drought

Riya Dutta, Rajib Maity*

Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

ARTICLE INFO

This manuscript was handled by Andras Barossy, Editor-in-Chief.

Keywords:

Hydrologic Extremes
Climate Change
Droughts
Time-varying approach
Bayesian Networks (BNs)

ABSTRACT

Hydrologic extremes often lead to droughts and floods that adversely affect the socio-economic development. Change in the characteristics and causes of hydrologic extremes due to climate variability and climate change poses a challenge for its reliable prediction. We propose a time-varying approach to capture such temporal changes, often gradual, in hydrologic extremes through temporal networks (a series of network structures). Graphical Modelling (GM) based networks are developed through Bayesian Model Averaging (BMA) to deal with the complexity between the causal variables and extreme events. A demonstration of the proposed time-varying approach is shown for 1-month and 3-month ahead hydrological drought prediction in terms of Standardized Streamflow Anomaly Index (SSAI), at basin scale, that has notably changed in the recent years in terms of its frequency and severity. The frequency and severity of below-normal flow events has increased, particularly during the monsoon season (high flow months). We hypothesize that time-varying cause-effect relationship is important to capture such gradual change in the characteristics of hydrologic extremes. The results indicate that SSAI values for the low flow months are strongly associated with streamflow whereas for the high flow months the dominant predictors are rainfall, precipitable water and relative humidity. Furthermore, the cause-effect relationship between hydroclimatic variables and extreme events needs to be updated every 2 years for high flow and 3 years for low flow months. The proposed model very well captures the above and below-normal flow events and can be used as a remedial measure to handle similar cases through a proper assessment of time-varying cause-effect relationship between hydroclimatic variables and extreme events.

1. Introduction

Spatio-temporal re-distribution of hydro-meteorological variables, due to changing climate and dynamic terrestrial environment, leads to changes in the characteristics and causes of extreme events. These changes negatively impact the socio-economic development and are often incomprehensible as it follows a complex mechanism (Van Lanen et al., 2013; Van Loon et al., 2016). Hydrologic extremes mostly originate from a deficit/excess of precipitation; however, hydrologists are more concerned with how this plays out through the different processes in the hydrologic cycle. The evolution of water deficit/excess through the different components of the hydrological cycle, like soil moisture and streamflow, is not instantaneous and is controlled by complex processes (Hao et al., 2018; Kiem et al., 2016). Other hydro-meteorological variables, such as temperature, potential evapotranspiration, relative humidity, precipitable water and pressure, also directly or indirectly influence the occurrence of hydrologic extremes (Cook

et al., 2014; Livneh and Hoerling, 2016; Luo et al., 2017). The aforementioned list of hydro-meteorological variables still remains incomplete and may vary with space and time. Overall, the development and evolution of an extreme event is dependent on multiple interacting factors, such as hydro-meteorological forcings, land-surface processes and human activities, and these factors are further accelerated under the impact of climate change (Cook et al., 2018; Mukherjee et al., 2018). Thereby, the complexity and uncertainty associated with the occurrence of such extremes make the investigation (modelling/analysis) of these events a challenging task.

Statistical modelling of hydrologic extremes, such as floods and droughts, has a long history, ranging from regression based models (Barros and Bowden, 2008; Liu and Negrón Juárez, 2001; Panu and Sharma, 2002; Sun et al., 2012) to artificial intelligence (AI), including machine learning (ML) and deep learning (DL) based techniques (Barua et al., 2012; Fung et al., 2020; Kaur and Sood, 2020; Khan et al., 2020; Maity et al., 2021; Mishra et al., 2007; Mishra and Desai, 2006; Santos

* Corresponding author.

E-mail address: rajib@civil.iitkgp.ac.in (R. Maity).<https://doi.org/10.1016/j.jhydrol.2021.126958>

Received 10 August 2021; Received in revised form 10 September 2021; Accepted 12 September 2021

Available online 20 September 2021

0022-1694/© 2021 Elsevier B.V. All rights reserved.

et al., 2014; Yang et al., 2015). Some of the probabilistic approaches are able to capture the nonlinear dependence among the variables and provide probabilistic prediction from the conditional distribution (Hao et al., 2016; Liu et al., 2015; Madadgar and Moradkhani, 2013; Svoboda et al., 2002; Wang et al., 2009; Wu et al., 2011; Yan et al., 2012; Zink et al., 2016). However, a major drawback of many modelling approaches is significant difficulty to build the joint distribution in higher dimensions when the pool of influencing variables is relatively large (Hao et al., 2018). Studies have utilized different techniques, such as correlation analysis, step-wise regression analysis, conditional independence structure based approach, model-free approach, self-organizing map, partial informational correlation, partial weights, wavelet-based techniques, to deal with such issues (Bowden et al., 2005; Dutta and Maity, 2020a, 2020b; Jiang et al., 2020; May et al., 2011; Sharma et al., 2016). New techniques for identifying important model inputs continue to emerge with each technique having its own advantages and limitations however no single method is best suited for all modelling purposes (Dutta and Maity, 2020a; Galelli et al., 2014; Maity and Kashid, 2011). Most of the existing techniques are either unable to avoid the redundant information from multiple associated variables or miss out important variables due to the complex nature of association (Schisterman et al., 2017). When a system like hydroclimatic system is composed of multiple interacting variables, a complete information on the conditional independence structure is helpful to obtain a well-defined set of input variables for a target variable. Only the directly associated variables may be picked out to use in the model, leaving out the effect of conditionally independent and independent variables (Ihler et al., 2007). This addresses the issues of high dimensionality due to large pool of influencing variables and effectively deals with redundant information from multiple variables. Here lies the advantage of Graphical Modelling (GM) approaches that provides a complete conditional independence structure that helps to further understand, predict and optimize the behavior of dynamical systems (Dutta and Maity, 2020c, 2020a, 2018).

Bayesian Networks (BNs), a class of GM approaches, are directed graphical models for representing probabilistic relationships among multiple interacting variables (Cooper, 1990; Heckerman et al., 1995; Witten et al., 2005). Formally, a BN is defined by a graphical structure, a family of (conditional) probability distributions and their parameters, which together specify a joint distribution over a set of random variables of interest. The graphical structure of a BN consists of a set of nodes and a set of directed edges. The nodes represent random variables, while the edges indicate conditional independence relations. A detailed literature review on the application of BNs in the field of hydrology and hydroclimatology can be found in Avilés et al., (2016) and Morrison and Stone, (2014). Recently, studies have utilized graphical modelling and Bayesian network based approaches to analyze primary hydrologic variables like precipitation (Das and Chanda, 2020; Dutta and Maity, 2020c; 2018), secondary hydrologic variables like streamflow (Dutta and Maity, 2020a; Ramadas et al., 2015) and tertiary hydrologic variables like drought (Avilés et al., 2016; Ramadas and Govindaraju, 2015). Whereas mostly the benefits of GM were realized in these studies, inherent non-stationarity may sometimes hinder the performance (Dutta and Maity, 2020a). Moreover, complexity in the graph structure may increase from primary to tertiary hydrologic variables. For example, identification of a single static graph structure assuming that the dependence among the variables remains constant over time may be questionable in a changing climate. Even keeping the effect of changing climate aside, moving from primary to tertiary hydrologic variables, such as droughts and floods, the uncertainty associated with the modelling framework substantially increases due to the complex interactions among the variables. In addition to this, climate variability and climate change leads to temporal redistribution of the hydro-meteorological variables causing intensification/alteration of the hydrologic cycle (Dutta and Maity, 2020a). This gradually changes the cause-effect relationship of the influencing variables and hydrologic

extremes and the characteristics of such events in terms of its frequency and severity. Such changes/fluctuations in the process may often lead to a non-stationary system (Betterle et al., 2017; Gibbs et al., 2018; Hwang et al., 2018; Milly et al., 2008; Wagener et al., 2010). Non-stationarity owing to gradual change in the association of the hydro-meteorological drivers and hydrologic extremes poses a challenge for modelling and development of the prediction models. Recent studies have used different techniques for modelling of extreme events in a non-stationary environment using copulas based techniques, detecting non-stationary hydrologic model parameters, time-varying model based on “generalized additive models for location, scale and shape” and break/change point analysis (Apurv and Cai, 2019; Chebana and Ouarda, 2021; Das et al., 2021; He et al., 2021; Hesarkazzazi et al., 2021; Machado et al., 2015; Pathiraja et al., 2016). However, in a time evolving process, the set of input variables may also change over time that most of the aforementioned approaches do not consider or need further development. In case of conventional GM approaches a single high scoring static model is utilized which may not represent the true time-varying association among the variables and there might be different models that explain the dependence reasonably well (Friedman and Koller, 2003; Tian et al., 2010). This forms the motivation of this study. There are two primary aspects to be considered. Firstly, the aspect of time-varying association among the influencing variables and hydrologic extremes can be dealt by updating/re-calibrating the model after a fixed time-interval. Such graph/network structures are referred to as temporal networks and have been successfully utilized to analyze primary and secondary hydrologic variables (Dutta and Maity, 2020a; 2020b). Secondly, in the context of tertiary hydrologic variables, like droughts, the aspect of uncertainty in identifying the robust networks can be dealt with by using Bayesian Model Averaging (BMA) for structural learning of Bayesian temporal network structures. BMA can be efficiently used to merge the information from multiple graph structures to truly understand the underlying process (Friedman & Koller, 2003). In some other context, BMA is utilized for merging forecast information from multiple models in order to improve the predictability of hydrologic variables such as, precipitation, streamflow and drought (Duan et al., 2007; Huo et al., 2019; Jiang et al., 2012; Lu et al., 2019; Meira Neto et al., 2018; Qu et al., 2017; Tian et al., 2018; Xu et al., 2018; Ye et al., 2004). However, utilization of combined potential of temporal networks along with BMA may be highly beneficial to deal with the inherent complexity in the tertiary hydrologic variables, being influenced by a large pool of variables and changing climatic and terrestrial conditions.

The objective of this study is to propose a BMA based temporal network approach to model the temporal evolution of the hydrologic extremes caused by changing climate and dynamic terrestrial environment. A demonstration is shown for 1-month ahead hydrological drought prediction defined by Standardized Streamflow Anomaly Index (SSAI), which has notably changed in recent years in terms of its frequency and severity (designated by SSAI magnitude). The two key components of the proposed modelling framework are, a) Bayesian framework, where BMA of BNs is used to learn the graph structures and b) the temporal networks, where the graph structure obtained using the Bayesian framework is re-iteratively updated along with the model parameters after a fixed time-interval. This fixed time-interval is optimized based on the model performance during model testing period. Month-wise SSAI values are estimated (twelve different series corresponding to each month) and analyzed to develop twelve different prediction models for each month of analysis. The performance of the proposed model is compared with its time-invariant counterpart and a commonly used machine learning based modelling frameworks.

2. Study area and data used

The upper Mahanadi river basin up to Basantpur gauging station is considered as the study area with an approximate spatial extent of 61,152 km², lying between 20°N to 23.5°N latitude and 80.5°E to 83°E

longitude. It may be noted that Mahanadi is one of the major rain-fed east flowing rivers in India. Different indices can be utilized to characterize the hydrological drought like Streamflow Drought Index (SDI), Standardized Streamflow Anomaly Index (SSAI), Palmer Hydrological Drought Index (PHDI), Standardized Standardized Reservoir Supply Index (SRSI), to name a few (<https://www.droughtmanagement.info/indices/>). First two indices are evaluated using streamflow data, PHDI is evaluated using precipitation, temperature and available water content and SRSI uses reservoir data. Each index has its own advantages and disadvantages, and selection of an index primarily depends on the application and the study area under consideration. Considering the strong seasonality in India, SSAI is used as the hydrological drought index to characterize the below/above-normal flow events. Detailed description on evaluation of SSAI is provided in the methodology section (Section 3.1). The time-period considered for the study is from January 1971 to December 2018 and the daily streamflow data at the outlet of the basin (Basantpur station) for the above time-period is obtained from the India-Water Resources Information System (India-WRIS; <https://indiawris.gov.in/wris/#/>). There are very few minor structures in the upstream, such as small check dams. However, given the large enough spatial extent of the basin, the streamflow values can be considered as unregulated.

In addition to streamflow, other input variables are temperature, precipitable water, potential evapotranspiration, pressure, relative humidity, soil moisture and rainfall, all from the previous time-step. These are selected based on the physical feedback mechanism of the hydroclimatic forcings, as established in different literature (Maity et al., 2010; Maity and Kashid, 2011; Meenu et al., 2013; Pichuka et al., 2017; Ramadas et al., 2015; Rehana and Mujumdar, 2014). Month-wise standardized anomaly values for each variable are obtained by subtracting each data series from their respective monthly mean and dividing by the respective standard deviation of that month. Daily rainfall data is obtained from India Meteorological Department (IMD) (Rajeevan et al., 2008) and data at each grid point is converted to monthly rainfall depth by accumulating it over the month. Temperature and potential evapotranspiration data are obtained from Climatic Research Unit (CRU) Time-Series (TS) gridded data (Harris et al., 2013). Soil moisture data is obtained from the Climate Prediction Centre (CPC) of the National Oceanic and Atmospheric Administration (NOAA) (Fan and van den Dool, 2004; CPC, 2014), evaluated using a land surface model (one-layer bucket water balance model) and rest of the variables are obtained from NCEP/NCAR Reanalysis 1 project using historical data to present (Kalnay et al., 1996). Gridded data is obtained from all the above-mentioned sources and the data are taken from the grid points lying within the study area. It may be noted that the NCEP/NCAR Reanalysis 2 product is an improvement over the NCEP/NCAR Reanalysis 1. Improvement in the quality of the input variables may have an improvement in the model performance. However, the primary reason for using Reanalysis 1 products is the availability of the data for a longer time period and a longer overlap with the available streamflow data for the study area. For development of a time-varying model, it is important to have a long enough data set so that the fixed time interval for model re-calibration can be effectively optimized. However, with the availability of longer streamflow data series it might be interesting to explore the uncertainty associated with each data source for different regions.

3. Methodology

Following sub-sections illustrate the main steps involved in the proposed time-varying approach to develop the prediction model. In brief, Section 3.1 deals with drought characterization. Next, a network/graph structure is developed through BMA of BNs in Section 3.2 among the associated variables (input and target variables). In Section 3.3, the concept of temporal networks is developed by imparting time-varying characteristics. Lastly, Section 3.4 provides details on the other existing approaches utilized for comparing the performance of the proposed

BMA based temporal network approach.

3.1. Hydrological drought characterization

A streamflow based drought index, referred to as Standardized Streamflow Anomaly Index (SSAI), is utilized to characterize the hydrologic drought. First, month-wise anomaly values of streamflow are obtained following eqn. (1), as follow:

$$x_{aij} = x_{ij} - x_i \quad (1)$$

where, x_{aij} is the anomaly value for the i^{th} month of the j^{th} year, x_{ij} is the observed streamflow for the i^{th} month of the j^{th} year and x_i is the long-term mean for the i^{th} month. Next, these anomaly values are fitted to a best-fit probability distribution, identified using the Chi-square goodness of fit test considering 5% significance level. The parameters of the fitted distribution are used to estimate the Cumulative Distribution Function (CDF) of the anomaly values, represented by $F(X_a) = P(X_a < x_a)$. These values range between zero to one and are referred to as reduced variates of streamflow anomaly. Next, these reduced variates are transformed to standard normal variates as follows,

$$Z_a = \Phi^{-1}(F(X_a)) \quad (2)$$

where, Φ^{-1} is the inverse of cumulative standard normal distribution. These values are the standardized streamflow anomaly values and can range between $-\infty$ to $+\infty$. Positive values of this index indicate above-normal flow events and the negative values indicate below-normal flow events or droughts resulting from streamflow excess and deficit, respectively. These SSAI values define the severity of extremes and are used as the target variable in developing the BMA based temporal network model. Initially, these SSAI values are used as the target variable in developing the BMA based temporal network model. Additionally, the frequency of occurrence of below/above-normal flow events is evaluated as the total number of such events occurring over a particular time period divided by the length of the respective time period. It may also be noted that for effective implementation of the proposed methodology the streamflow data should be de-trended before evaluating the SSAI values.

Next, different categories of droughts, e.g., moderate, severe, extreme, are considered in the analysis, depending on the numerical values (also referred to as severity) of the index. Following Maity et al., (2013), following near normal flow category (N : near normal events including normal, abnormally dry and abnormally wet conditions), below-normal flow/drought categories ($D1$: moderately dry, $D2$: severely dry, $D3$: extremely dry and $D4$: exceptionally dry events) and above-normal flow categories ($W1$: moderately wet, $W2$: severely wet, $W3$: extremely wet and $W4$: exceptionally wet events) are considered. In this categorization, a value in the range of $[-0.7, 0.7]$ are considered as near normal events, as shown in Figs. 2 and 3.

3.2. Development of the BMA based network structure and the prediction model

BNs are a class of graphical models that represent the association between different variables by means of Directed Acyclic Graphs (DAGs; Cooper, 1990; Heckerman et al., 1995; Witten et al., 2005). Given the causes (causal/input variables), a BN can be used to compute the effect (target variable). Development of a BN involves learning the network structure, also referred to as the graph structure, development of the probabilistic model (parameter estimation of conditional probabilities) and prediction of the target variable given the directly influencing input variables.

In traditional BNs, two different learning/model selection algorithms, namely score based algorithm and constraint based algorithm are utilized to learn the graph structure (Scutari, 2017). These algorithms provide a single high-scoring model that is selected as the final

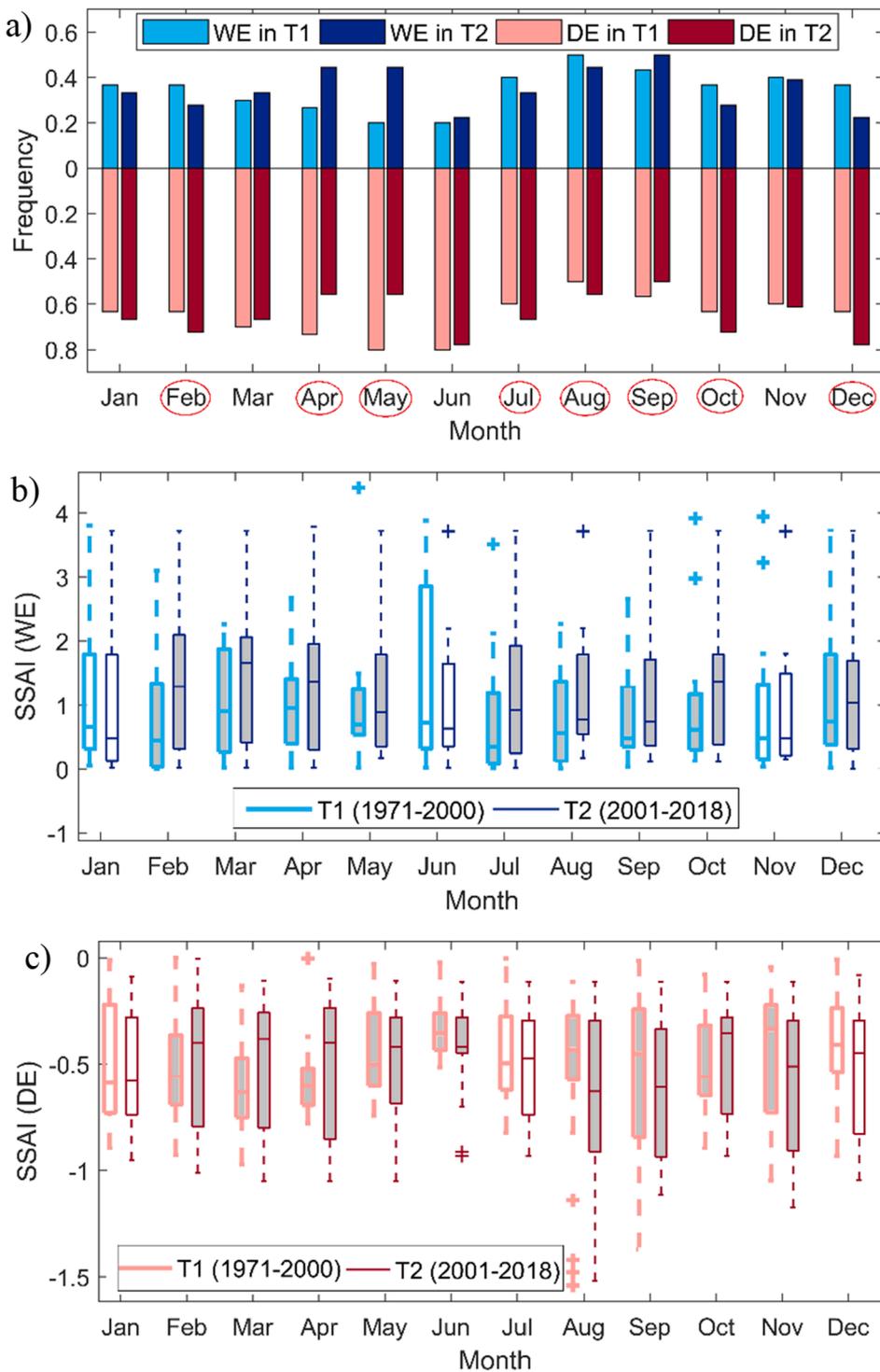


Fig. 1. Temporal change in the a) frequency of below-normal (designated by DE) and above-normal (designated by WE) flow events b) severity (designated by SSAI magnitude) of above-normal flow events and c) severity of below-normal flow events. The months with significant change (at 5% significance) in frequency are encircled in red and the months with significant change (at 5% significance) in severity of above and below normal events (in terms of mean) are filled in grey. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

graph structure (Heckerman et al., 1995). However, considering the uncertainty associated with tertiary hydrologic variables such as drought and flood, there may be multiple graph structures (in addition to the single high scoring model) that represent the dependency among the associated variables equally well. Furthermore, other structures may present some vital information related to the conditional independence among the variables that may get overlooked while selecting the single high scoring model. Owing to the complexity of the hydroclimatic forcings behind hydrologic droughts it may be advantageous to identify multiple graph structures that fit the data reasonably well and derive the combined information from all these structures to obtain the association

among the variables. In this study, the order-Markov Chain Monte Carlo (order-MCMC) algorithm, based on the BMA of BNs, is used for precisely learning the graph structure (Friedman and Koller, 2003) by averaging the information from multiple graph structures. To start with, all the graph structures that fit the data comparatively well based on the likelihood-equivalent Bayesian score, also referred to as the BDe score (detailed description for evaluation of BDe score is provided in Appendix A), are selected. The number of selected structures is optimized so that the highly scoring graph structures are not excluded and the computational cost is minimized. Next, the selected graph structures are grouped based on their topological order (designated by θ). These orders are

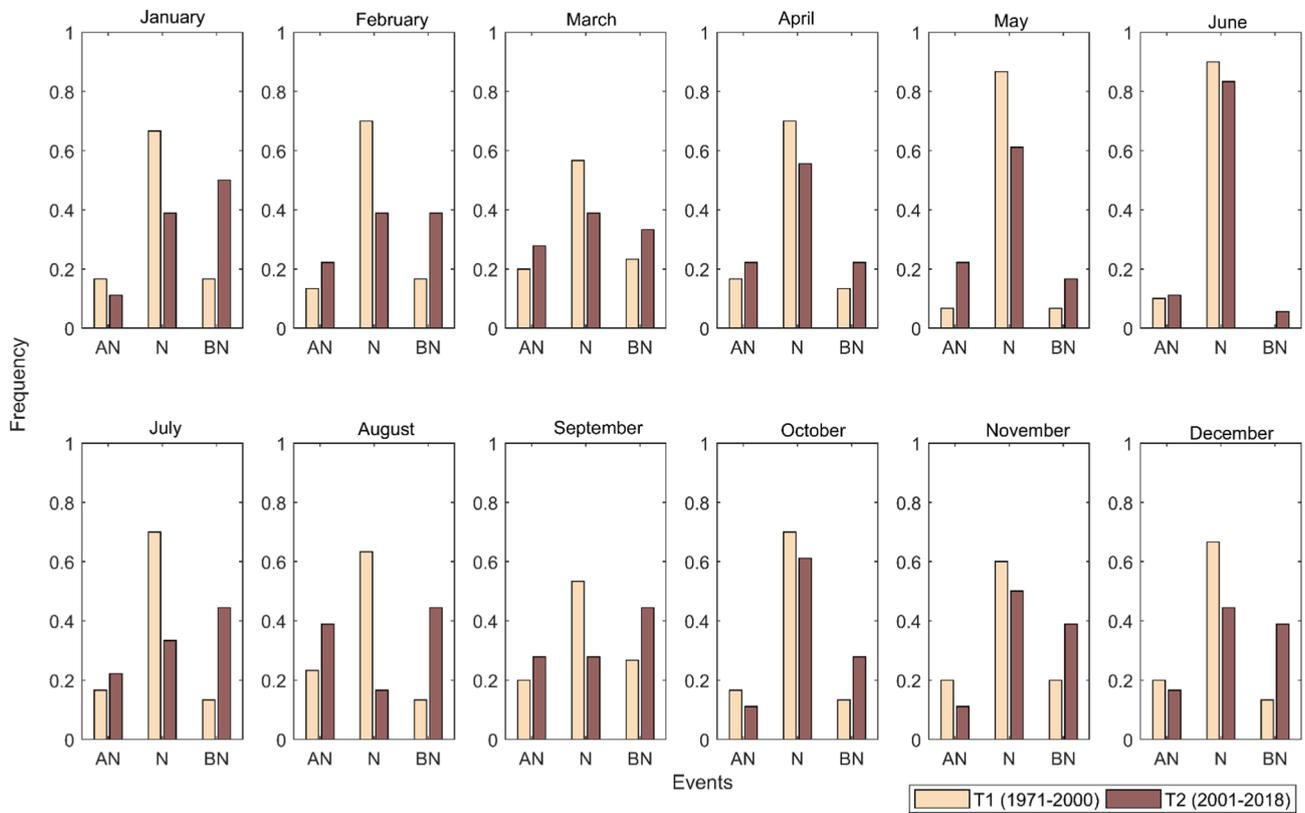


Fig. 2. Temporal change in the frequency of extreme events considering the first model development period of 1971–2000 and the contiguous testing period of 2001–2018. The below/above-normal flow events are categorized as near normal ($[-0.7, 0.7]$, designated by N), above-near normal (>0.7 , designated by AN) and below-near normal (<-0.7 , designated by BN) events. The severity of drought events has changed in the recent years considering both the high and low flow months. The change in frequency of near normal events is significant (at 5% significance) for all the months except for the month of June.

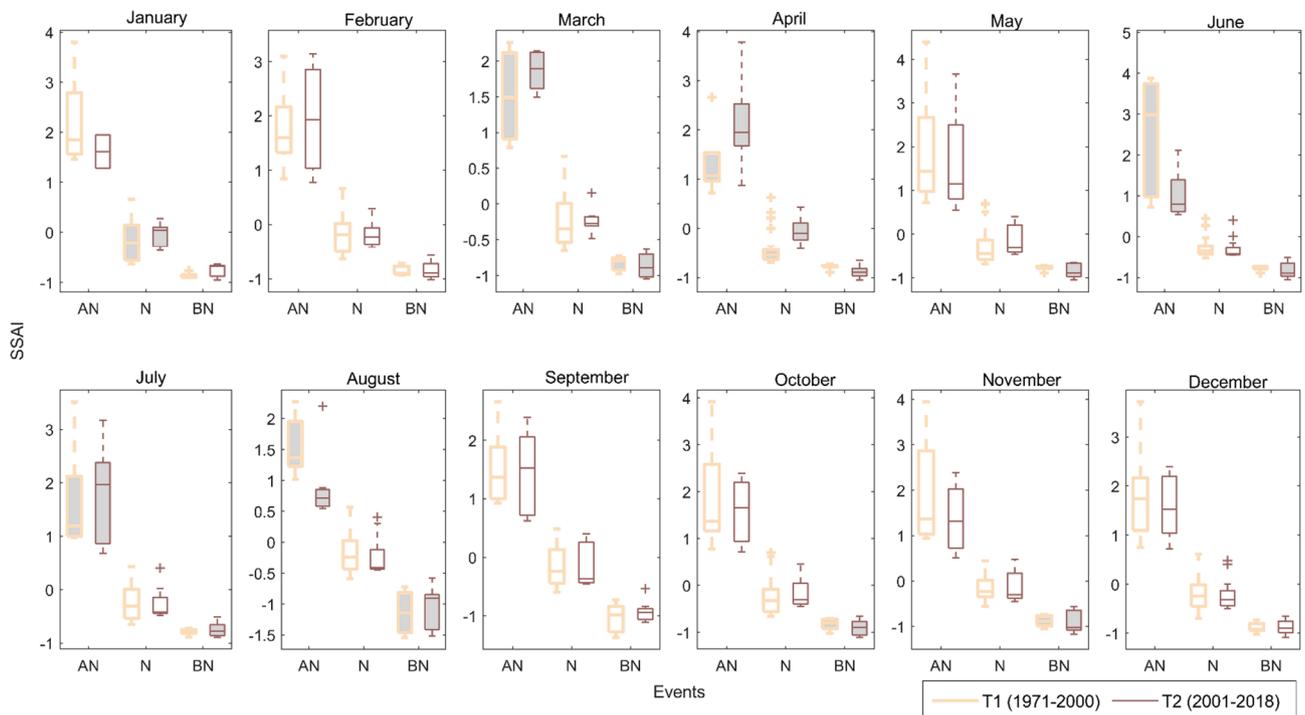


Fig. 3. Temporal change in the severity of drought events (designated by SSAI magnitude) in terms of mean, 25th and 75th percentile and the range considering the first model development period of 1971–2000 and the contiguous testing period of 2001–2018. The below/above-normal flow events are categorized as near normal ($[-0.7, 0.7]$, designated by N), above-near normal (>0.7 , designated by AN) and below-near normal (<-0.7 , designated by BN) events. The months for which the change in the mean, considering the above categories, is significant at 5% significance level are filled with grey.

arranged in such a way that each node may only have parents from further up the chain or following it in the ordering. All the graph structures consistent with an order are combined to reduce the search to smaller space and a score is assigned to each group which is equal to the sum of the scores of all graph structures in the group. Next, a Markov chain is constructed considering all the node orders rather than all the graph structures. Detailed description on evaluation of the group score and construction of the Markov chain is provided in [Appendix A](#). By grouping together and averaging the score over so many graph structures, an optimized structure is obtained that represents the association among the input variables and the target variable. After identification of the final graph structure that represents the association among the variables, the degree of association between the input variables and target variable is evaluated. This degree of association is measured as score gained/lost (BDe score), when a particular edge, for which the strength is being evaluated, is included/excluded from the structure. That is, the BDe score for two graph structures, one with and another without the edge, are evaluated. The difference between the score of the graph is considered as the edge strength. Using the obtained graph structure, the prediction model is developed by factorization of the graph and parameter learning.

The prediction model is the joint probability distribution associated with a given graph structure. It is obtained as a product of functions associated with a subset of the nodes (variables). The function turns out to be the conditional probability of a variable given its parent variables ([Ihler et al., 2007](#)). Identification of the conditional probability distribution also referred to as parameter learning is carried out using the Maximum Likelihood Estimation (MLE) approach (details provided in [Appendix A](#)). The prediction model becomes ready once the parameters are estimated. Next, the prediction of the target variable is carried out by plugging in the new values for the “parents of the target variable” (variables directly associated with the target variable as obtained from the graph structure), in the conditional probability distribution of the target variable. It may be noted that the above methodology is used for the model calibration considering a particular model development period.

3.3. Development of the temporal networks and re-calibration of prediction model

The time-varying characteristics are imparted to the prediction model by gradually updating the network structures over time, also referred to as temporal networks ([Dutta and Maity, 2020a](#)). The model development period is considered as a moving window of 30 years and the model is re-calibrated iteratively after a fixed time-interval (say n years) in terms of model inputs and parameters. The value of n needs to be optimized and this optimized value is designated by τ and referred to as the Optimum Recursive Interval (ORI) for model re-calibration ([Dutta and Maity, 2020a, 2018](#)). In order to obtain τ , different values of n , starting from $n = 1$ to $n = 5$ years, are considered. As the time-period of the study is from 1971 to 2018, the first model development period is considered from 1971 to 2000 and the model testing period is from 2001 to 2001+($n-1$). As the model is updated after n years the next model development period is shifted by n years and the second model development period is considered from 1971 + n to 2000 + n . The process continues over the entire time period of the study. To identify the value of τ , this procedure is repeated for different values of n and the model performance during all the contiguous model testing periods, is evaluated to identify the ORI of model re-calibration.

3.4. Comparison with existing approaches

Performance of the proposed BMA based temporal network approach is compared with three commonly used modelling concepts in the field of hydrology. To start with, a time-invariant counterpart of the BMA based temporal network approach is used. Next, two Machine Learning

(ML) techniques, namely Support Vector Regression (SVR) and Artificial Neural Network (ANN) are utilized in a time-varying framework. Details on all the three models are as follows. Firstly, for development of the BMA based time-invariant network approach, the procedure explained in the previous sub-section remains the same but only one graph structure is developed using 30-years data. Next, the developed prediction model is used for the entire testing period without the concept of time-varying models, explained before. Lastly two commonly used ML based approaches namely, SVR and ANN are developed based on the time-varying concept but the concept of conditional independence structure is not utilized. SVR and ANN are common machine learning techniques utilized in different hydroclimatic studies ([Ardabili et al., 2019](#); [Barua et al., 2012](#); [Cristianini and Shawe-Taylor, 2000](#); [Khan et al., 2020](#); [Maity et al., 2010](#); [Prasad et al., 2017](#); [Raghavendra and Deka, 2014](#)). Inputs for the SVR model are identified through correlation analysis as followed traditionally. In this comparison, the ML based models are developed with the aforementioned time-varying concept and thus, the inputs and the parameters of the models are updated after n years, as in the proposed BMA based temporal network approach.

4. Results and discussions

Broadly the results are presented to show the temporal changes in the characteristics of hydrologic drought and the ability of the proposed model to capture temporal change in the association among the hydroclimatic variables and predict the occurrence of below/above normal flow events. Section-wise presentation of results and discussion is as follows: Section 4.1 shows the change in the frequency and severity of observed extreme events considering the first model development period and the contiguous model testing period. Next, the hydroclimatic forcings directly influencing the hydrologic drought as revealed by the network structures (developed using the proposed model) is presented in [Section 4.2](#). Section 4.3 shows the temporal change in the network structures (developed using the proposed model) and establishes the fact the association between the hydroclimatic forcings and extreme events is gradually changing over time. Section 4.4 shows the ability of the proposed BMA based temporal network approach to capture the below/above normal flow events. Lastly, [Section 4.5](#) compares the performance of the proposed model with other well-established modelling approaches.

4.1. Temporal change in the frequency and severity of hydrologic drought

At the outset, the change in the characteristics of extreme events is ascertained in terms of change in the frequency and severity of below/above-normal flow events. [Fig. 1a](#) shows the temporal change in the frequency of the below/above-normal flow events, evaluated individually for each month, considering the time periods of 1971–2000 (henceforth denoted as T1) and 2001–2018 (henceforth denoted as T2). Additionally, the significant changes in the frequency, identified using the two-sample z-test of proportions considering 5% significance level is also shown. It is clearly noticed that the frequency of below-normal flow events has increased for the post-monsoon season (October and November with significant increase in October), winter season (December-February with significant increase in December and February) and two monsoon months of July and August (significant change in both the months). Furthermore, the pre-monsoon season (March-May with significant increase in April and May) and other monsoon months, i.e., June and September (significant change in September) exhibit an increase in the frequency of above-normal flow events. The higher frequency of below/above-normal flow events especially considering the months in the pre-monsoon (March-May), monsoon (June-September) and post-monsoon (October-November) seasons, may arise from the change in monsoon intensity and shift in monsoon pattern ([Sahu et al., 2020](#)).

Secondly, the temporal change (considering the time periods of T1

and T2) in the severity of below/above-normal flow events, assessed by considering the negative and positive values of SSAI as two different data series, is shown in Fig. 1b and 1c, respectively, through boxplots. The figures show the change in the severity of below/above-normal flow events in terms of either increase or decrease in the mean, range, 25th quartile (lower) and 75th quartile (upper) values of SSAI for each month of analysis. The significant changes in the mean, identified using the two-sample *t*-test considering 5% significance level, for the below/above normal flow events are shown in grey. Considering the months of July–October, (months falling in the monsoon and post-monsoon season) an increase in the mean, maximum and upper quartile value is observed, thereby showing an increase in the severity of above-normal flow events (Fig. 1b). Similar observations can be made for the comparatively low flow months of February, March, April and May that exhibit significant change (at 5% significance) in its mean. It is interesting to note that in the pre-monsoon season both the frequency and severity of above-normal flow events have increased in the recent time period, i.e., T2 as compared to T1. Further, the severity of below-normal flow events in terms of mean, lower quartile and minimum value has increased considering the monsoon season (June–September). Similar observations are made for the months of November, December and January too. A decrease in the severity in terms of mean is observed for the months of February–May (significant at 5% significance). However, an increase in the severity in terms of the lower quartile value is observed for these months. It is interesting to note that the months in the monsoon season show change in the frequency and severity of below/above-normal flow events depending on the month of analysis.

In order to ascertain the temporal change in the frequency and severity of below/above-normal events, the SSAI values have been divided into three categories (as shown in the contingency tables in Fig. 9; First Table). Fig. 2 shows the temporal change (considering time periods T1 and T2) in the frequency of different categories of below/above-normal flow events for each month of analysis. It may be noted that for all months starting from January to December, the frequency of below-normal events has increased. Similarly, the frequency of above-normal events has also increased for the months of February–September. Thereby considering the months in the monsoon season (June–September), the frequency of near normal events has decreased and the frequency of both above- and below-normal events have increased. Moreover, the severity of above-normal and below-normal events has increased (decreased) either in terms of mean, (upper quartile) lower quartile or (maximum) minimum values for almost all the months (Fig. 3). The significant changes in the mean, identified using the two-sample *t*-test considering 5% significance level, for all the three categories of below/above-normal flow events are shown in grey. The pre-monsoon months of March and April show significant increase in the mean severity of above-normal events and significant decrease in the mean severity of below-normal events. The month of May shows a significant decrease in the mean severity of below-normal events. Next, considering the monsoon months of June, July and August significant change is observed in the mean severity of both above- and below-normal events. For instance, the months of June and August show significant decrease in the mean severity of above-normal events and July shows significant increase in the mean severity of above-normal events. Furthermore, the post-monsoon season (October and November) shows significant decrease in the mean severity of below-normal events. Thereby, alterations of the hydrologic cycle strongly impact the characteristics of the extreme events over time, and significant temporal changes are noticed in the recent decades (T2) as compared to past (T1).

4.2. Hydroclimatic forcings behind hydrologic droughts as revealed by the network structures

The proposed approach is based on the concept of network structures which provides the complete conditional independence structure that

reveals the hydroclimatic forcings behind the hydrologic droughts. Figs. 5 (panel#1) and 6 (panel#1) show the network structures for the first model development period (1971–2000) considering one typical low flow (January) and high flow (July) month, respectively. The input variables (temperature-TE, precipitable water-PW, potential evapotranspiration-PE, pressure-PR, relative humidity-RH, soil moisture-SM, rainfall-RA, streamflow-ST; all inputs from the previous month) and the target variable (SSAI; current month of analysis) are represented by differently colored nodes and the association between any two variables is shown by a directed edge. Absence of an edge can be interpreted as the variables being independent or conditionally independent and presence of an edge can be interpreted as the variables being directly dependent. Considering the month of January (a typical high flow month), SSAI (target variable) is directly dependent on streamflow and conditionally independent of all the other input variables. Streamflow is directly influenced by soil moisture, which is in turn directly dependent on rainfall and temperature. Further, relative humidity is associated with rainfall, precipitable water and potential evapotranspiration. A possible reason for the indirect connection between rainfall and precipitable water could be the multiple factors affecting precipitation quantity like degree of saturation, atmospheric water vapor, and the presence of dynamic mechanisms which provide the cooling necessary to produce saturation. Similar network structures are identified for each month and the input variables directly influencing SSAI are shown in Fig. 7. In Fig. 7, input variables are represented by colored squares; black color indicates no association or conditionally independent association and different shades of red indicates significant association with varying values of edge strengths (strength of association). Considering other low flow months of December, February, March, April and May and the first model development period (Fig. 7a), streamflow emerges as the primary influencing variable for 1-month ahead prediction. For the months of May and December, in addition to streamflow, rainfall also shows direct association with SSAI. A probable reason can be the higher rainfall magnitude in the months of May (falling in the pre-monsoon season) and December (falling in the winter monsoon season) as compared to the other low flow months for the considered study area.

Next, considering the month of July (a typical high flow month) and the first model development period (Fig. 6, panel#1), SSAI (target variable) is found to be directly dependent on rainfall and precipitable water. Furthermore, the streamflow variations show direct dependence on soil moisture and rainfall, and relative humidity is directly associated with temperature, potential evapotranspiration and precipitable water. During high flow months, it is expected that the moderately sized catchment will almost be in a stationary state as the precipitation is quite continuous over time. Thus, the variation in extreme events will respond mainly to precipitation variation in time. Fig. 7b shows that precipitable water directly influences SSAI for August and September (other monsoon months) during the first model development period. Streamflow shows a direct association with SSAI in case of June (first monsoon month) and November. Thereby the complete conditional independence structure, as identified by the proposed approach, helps to identify the directly influencing variables and given these variables below/above-normal flow events are conditionally independent or independent of the other hydroclimatic forcings.

4.3. Temporal change in the hydroclimatic forcings behind hydrologic droughts

The time-varying association between hydroclimatic forcings and SSAI is assessed through the proposed BMA based temporal network approach. To start with, the first 30-year period is considered as the first model development period and the initial network structure is developed. The network structure is recursively re-developed after each τ years interval, denoted as ORI, as defined in the methodology section (Section 3.3). The value of τ is ascertained by assuming different values

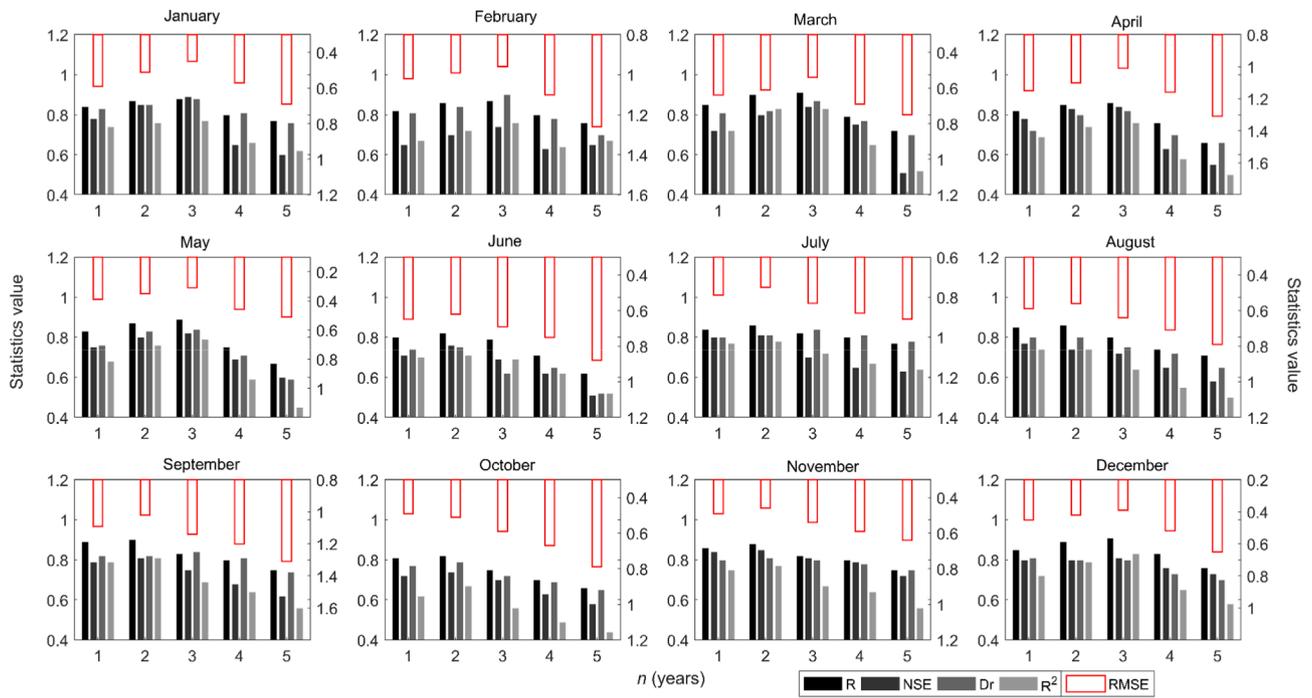


Fig. 4. Month-wise comparison of the different statistics value to identify the value of τ , considering the contiguous model testing period (2001–2018), best performance is obtained considering the value of n as 2 years for the monsoon and post-monsoon season (June–November) and 3 years for the rest of the months (December to May).

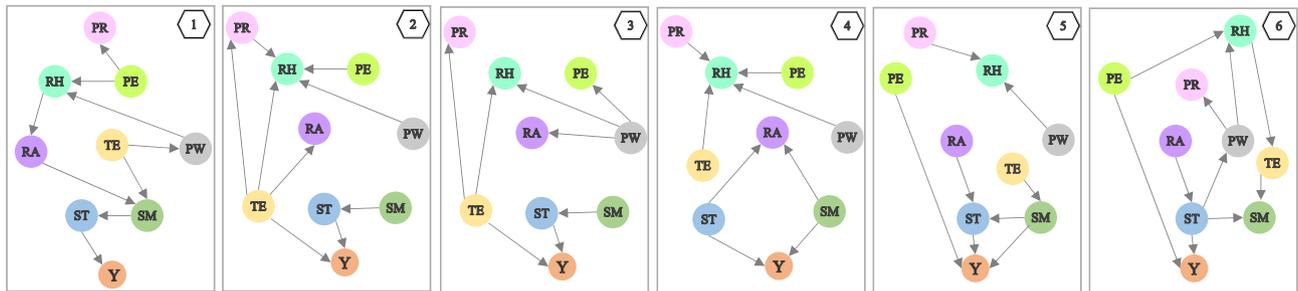


Fig. 5. Time-varying association of the hydroclimatic forcings (input) variables and SSAI (target variable; indicated as Y) over the years for the month of January (typical low flow month). The model development period is considered as a moving window of 30 years and considering the ORI of model calibration as 3 years the six model development periods are shown as 1 (1971–2000) to 6 (1986–2015).

($n = 1, 2, 3, \dots$) and comparing the model performance for the contiguous model testing period. Four performance metrics namely, Correlation coefficient (R), Nash-Sutcliffe Efficiency (NSE), Index of agreement (Dr), Coefficient of determination (R^2) and Root Mean Square Error ($RMSE$) are utilized for assessment. Fig. 4 shows the month-wise performance of the models considering the different values of n (1, 2 ... 5 years). It is observed that a value of $n = 2$ years and $n = 3$ years provides the best possible performance considering the months in the monsoon and post-monsoon season (June–November) and the months in winter, and pre-monsoon season (December–May), respectively. Thereby the prediction model needs to be re-calibrated after 2 and 3 years, considering the comparatively low flow and high flow months, respectively in order to capture the time-varying association among the hydro-meteorological variables.

Considering the value of τ as 2 years, the corresponding model development periods (testing periods) are 1971–2000 (2001–2002), 1973–2002 (2003–2005) ... 1987–2016 (2017–2018) and considering the same as 3 years, the corresponding model development periods (testing periods) are 1971–2000 (2001–2003), 1974–2003 (2004–2006)

... 1986–2015 (2016–2018) Thereby, the model is calibrated 9 times for the high flow months and 6 times for the low flow months, considering the entire time period (1971–2018) of the study. Figs. 5 and 6 show the time-varying temporal networks obtained for the 6 model development periods considering a typical low flow (January) and a typical high flow (July) month, respectively. For example, considering the month of January and the first model development period SSAI is directly dependent on streamflow, whereas for the second model development period (Fig. 5, panel#2), in addition to streamflow, SSAI also shows a direct dependence on temperature. Further, it is interesting to note a distinct change in the interaction among the other input variables also. For instance, soil moisture is conditionally independent of rainfall and temperature. Similar observations can be made for the third model development period also. Such changes in the interaction among the variables are accounted for by the modified rainfall pattern for the considered study area and a change in the terrestrial environment. Considering the fifth and sixth model development period, SSAI shows direct dependence on the streamflow and potential evapotranspiration in addition to other input variables. Change in climatic factors like high/

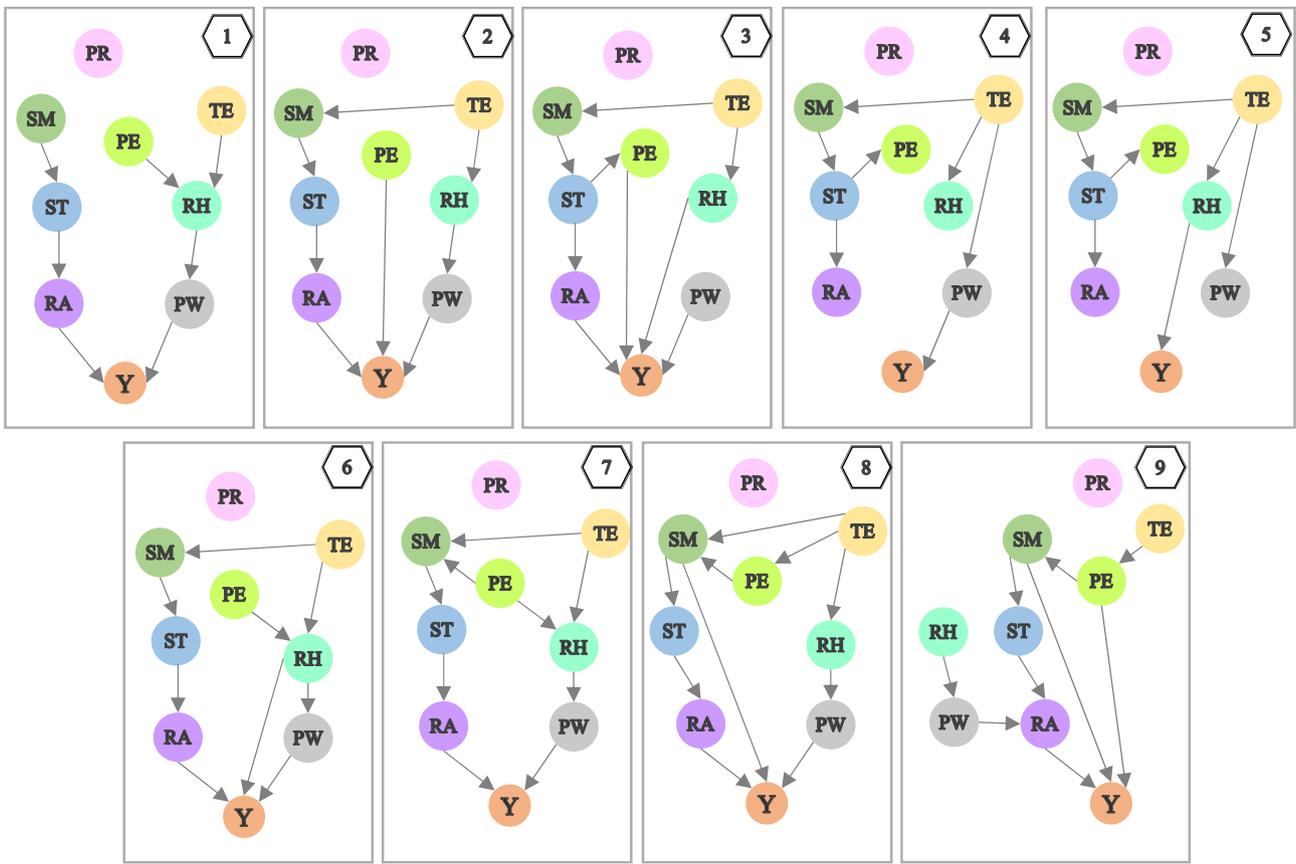


Fig. 6. Time-varying association of the hydro-meteorological (input) variables and SSAI (target variable; indicated as Y) over the years for the month of July (typical high flow month). The model development period is considered as a moving window of 30 years and considering the ORI of model calibration as 2 years the nine model development periods are shown as 1 (1971–2000) to 9 (1987–2016). The input variables used are Streamflow (ST), Rainfall (RA), Soil Moisture (SM), Relative Humidity (RH), Pressure (PR), Potential Evapotranspiration (PE), Precipitable Water (PW) and Temperature (TE), all from the previous time-step.

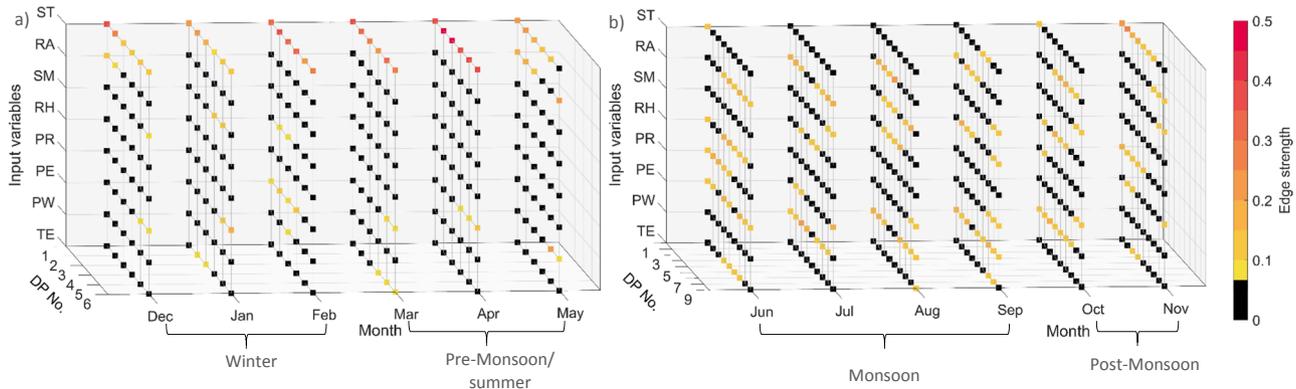


Fig. 7. Time-varying association of the input variables and the drought index (SSAI). The input variables used are Streamflow (ST), Rainfall (RA), Soil Moisture (SM), Relative Humidity (RH), Pressure (PR), Potential evapotranspiration (PE), Precipitable Water (PW) and Temperature (TE), all from the previous time-step. Based on the obtained value ORI of model calibration of a) 3 years for the dry flow months, the six model development periods are shown as 1 (1971–2000) to 6 (1981–2010) and b) 2 years for the high flow months, the nine model development periods are shown as 1 (1971–2000) to 9 (1987–2016). SSAI for the low flow months show strong association with streamflow as compared to other input variables. However, considering the low flow months SSAI is strongly influenced by variables like precipitation, precipitable water and relative humidity.

low vapor pressure, increased/decreased radiation in the surrounding region and change in terrestrial environment may affect the influence of potential evapotranspiration over time. Next, considering the month of July and the first model development period SSAI is found to be directly dependent on rainfall and precipitable water. It is interesting to note that considering the second model development period (Fig. 6, panel#2), in addition to rainfall and precipitable water, potential

evapotranspiration also shows a direct association with SSAI. The interaction between the other input variables is more or less similar with pressure remaining independent of all the other variables. This result contradicts the findings obtained for a typical low flow month, whereas the interaction between the variables drastically changes over the time. Considering the sixth model development period (1981–2010), SSAI shows a direct dependence on rainfall, relative humidity and

precipitable water. A gradual shift in the directly dependent input variables is observed considering the last two model development periods with increasing dependence of SSAI on soil moisture and potential evapotranspiration.

Similar temporal networks are identified for each month and the input variables directly influencing SSAI are shown in Fig. 7. For example, considering the month of February, during the first model development period (1971–2000), SSAI is directly dependent on streamflow and potential evapotranspiration but conditionally independent of precipitable water, pressure, temperature, relative humidity, soil moisture, and rainfall. Moreover, the strength of association between SSAI and streamflow is stronger considering the first three model development periods, i.e., 1971–2000, 1974–2003, and 1977–2006. Gradually the strength reduces over time as noticed for the fourth (1980–2009), fifth (1983–2012) and sixth (1986–2015) development periods. A closure observation reveals that the high variation in February streamflow series is noticed during the entire time-period with a significantly increasing trend till late 1990s and a decreasing trend afterwards. This leads to a changing below/above-normal flow events and alteration in the association of SSAI with streamflow over time.

Based on the obtained results, it is interesting to notice that SSAI for the low flow months, namely, January, February, March, April, May, and December with the streamflow range of 1.83 to 249.55 cumec, is strongly associated with streamflow (Fig. 7a). Contrary to this, during the high flow months of June, July, August, September, October and November where the range of streamflow lies between 247.88 and 6572.32 cumec, the dominant predictors are rainfall, precipitable water and relative humidity (Fig. 7b). Furthermore, the number of input variables directly influencing SSAI, considering a particular model development period for the high flow months is more as compared to the low flow months. As streamflow is the dominant predictor considering the SSAI of the low flow months with high strength of association, most of the information on the below/above-normal flow events can be extracted from streamflow itself making the contribution of other variables insignificant. However, considering the high flow months, most of the significant input variables show similar association with SSAI and information from all these variables is vital to explain the below/above-normal flow events. It can be clearly stated that the causal factors (input variables) of extreme events change temporally for a particular season as

well as from one season to another. Thereby, it is vital to iteratively re-calibrate the model to successfully capture the below/above-normal flow events.

4.4. Performance of the proposed time-varying approach

The observed and predicted SSAI values obtained using the BMA based temporal network approach along with the different categories of below/above-normal flow events depicted with different color gradations of red/blue are shown in Fig. 8. The near-normal events are shown in grey. The proposed approach shows satisfactory performance in capturing the events falling near normal to severe dry/wet flow events. However, considering the extreme and exceptional dry/wet flow events the model shows better performance in capturing these two wet flow categories as compared to these dry flow categories. Further, considering four different contingency tables, Fig. 9 shows the different categorizations of SSAI (maintaining the range of near-normal events as $-0.7, 0.7$) and the efficacy of the proposed model to capture the same. The values mentioned in the diagonals (shown in grey) of the contingency tables show that the observed and predicted values of SSAI fall in the same range, i.e., a near normal/below-near normal/above-near normal events have been correctly captured along with its severity range. For instance, considering the third type of categorization, 23 (summation of last column) above-near normal flow events have been observed to be greater than 1.2. Out of these events, 18 (diagonal element) events have been correctly predicted to lie in this range. The remaining five events have been under-predicted as less severe above-near normal flow events (4) and a near normal flow event (1). Similarly, considering the second type of categorization, 57 (summation of third column) below-normal flow events have been observed to lie in the range $[-1.2, -0.7]$. 24 events have been correctly predicted to fall in this range and the rest of the events are predicted as either less or more severe above-normal flow events. It can be observed that as the range of SSAI is made smaller for categorization of below/above-normal flow events the model error gradually increases; however, the model performance is satisfactory considering all the four categories. Overall, the SSAI as well the range are well captured by the proposed BMA based temporal network approach, due to its ability to re-calibrate the model and capture the time-varying characteristics among the variables.

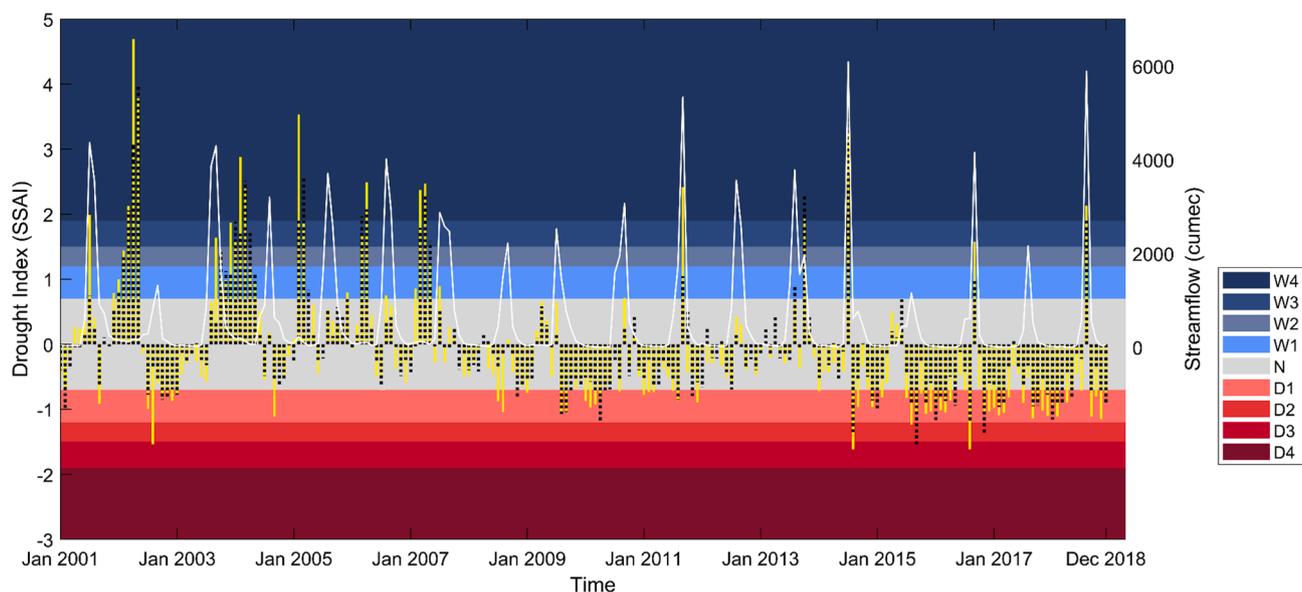


Fig. 8. Comparison of the observed (solid yellow) and predicted (dotted black) SSAI obtained using the proposed BMA based temporal network approach for the month-wise analysis. The positive and negative values of SSAI (above and below-normal flow events) are divided into nine categories as near-normal (N), moderately wet/dry (W1/D1), severely wet/dry (W2/D2), extremely wet/dry (W3/D3) and exceptionally wet/dry (W4/D4). The streamflow values for the contiguous testing period are also plotted in white. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Next, month-wise prediction performance of the proposed BMA based temporal network approach is assessed by comparing the observed and the predicted SSAI values (Fig. 10). The results show that the predicted values closely follow the observed values for many months, such as January, March, May, October and November. For the months of February, April, July and September, positive errors are noticed and for the months of June, August and December negative errors are observed. It is also interesting to note that as compared to the high flow months (the statics value lie in the following range: $R = 0.85-0.91$, $RMSE = 0.50-1.13$, $NSE = 0.75-0.82$, and $Dr = 0.75-0.83$), the performance of the model is comparatively better considering the low flow months (the statics value lie in the following range: $R = 0.86-0.92$, $RMSE = 0.32-1.01$, $NSE = 0.75-0.87$ and $Dr = 0.80-0.91$). A probable reason is the high variations in streamflow during the high flow months leads to extreme and exceptional above-normal flow

events that are complex in nature. Further, the uncertainty range considering the predictions at the 5th and 95th percentile is also shown in Fig. 10. It may be noted that all the observed values fall within the uncertainty range, except one observed SSAI value in the months of February and April.

4.5. Comparison of the proposed approach with the other modelling approaches

Performance of the BMA based temporal network approach is compared with three other modelling approaches: i) BMA based time-invariant network approach, ii) SVR based time-varying approach, and iii) ANN based time-varying approach. The performance statistics obtained using all the three models are compared in Table 1 (the models are designated as M1, M2, M3 and M4 and the data series for month-wise

		Observed								
	<-1.9	0	0	0	0	0	0	0	0	0
	[-1.9,-1.5)	0	0	0	1	0	0	0	0	0
	[-1.5,-1.2)	0	1	0	2	0	0	0	0	0
	[-1.2,-0.7)	0	1	1	24	9	0	0	0	0
	[-0.7,0.7]	0	1	0	30	111	5	0	1	0
	(0.7,1.2]	0	0	0	0	2	5	0	2	2
	(1.2,1.5]	0	0	0	0	0	0	2	0	0
	(1.5,1.9]	0	0	0	0	0	0	1	1	2
	>1.9	0	0	0	0	0	0	0	2	10
			<-1.9	[-1.9,-1.5)	[-1.5,-1.2)	[-1.2,-0.7)	[-0.7,0.7]	(0.7,1.2]	(1.2,1.5]	(1.5,1.9]
Predicted	<-1.5	0	0	1	0	0	0	0	0	
	[-1.5,-1.2)	1	0	2	0	0	0	0	0	
	[-1.2,-0.7)	1	1	24	9	0	0	0	0	
	[-0.7,0.7]	1	0	30	111	5	0	1	0	
	(0.7,1.2]	0	0	0	2	5	0	4	0	
	(1.2,1.5]	0	0	0	0	0	2	0	0	
	>1.5	0	0	0	0	0	1	15	0	
			<-1.5	[-1.5,-1.2)	[-1.2,-0.7)	[-0.7,0.7]	(0.7,1.2]	(1.2,1.5]	>1.5	
	<-1.2	1	3	0	0	0				
	(-1.2,-0.7]	2	24	9	0	0				
	[-0.7,0.7]	1	30	111	5	1				
	[0.7,1.2]	0	0	2	5	4				
	>1.2	0	0	0	0	18				
		<-1.2	[-1.2,-0.7)	[-0.7,0.7]	(0.7,1.2]	>1.2				
	<-0.7	30	9	0						
	[-0.7,0.7]	31	111	6						
	>0.7	0	2	27						
		<-0.7	[-0.7,0.7]	>0.7						

Fig. 9. Compilation of four contingency tables comparing the observed versus predicted SSAI values divided into different categories based on the severity of the below/above-normal flow events. The first table shows the nine categories namely, near-normal i.e. [-0.7, 0.7], moderately wet/dry, severely wet/dry, extremely wet/dry and exceptionally wet/dry and the efficiency of the prediction model in capturing the extreme events. For the following three tables the near-normal band is kept the same; however, the bands defining the below/above-normal events are gradually changed in order to provide a detailed analysis of the model performance.

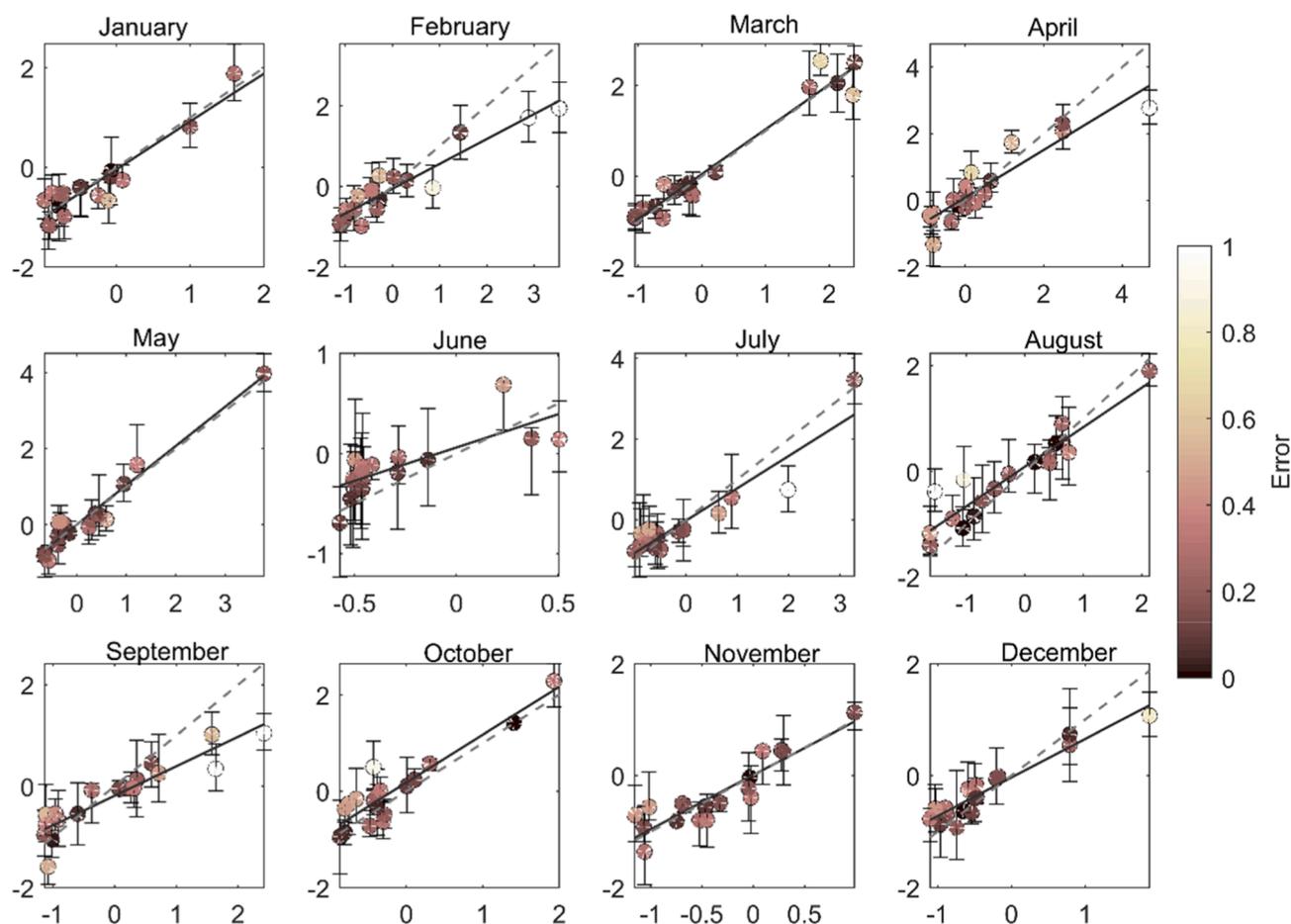


Fig. 10. Month-wise performance of the BMA based temporal network approach in capturing SSAI for all the contiguous testing periods, i.e. 2001–2018. The upper and lower bar shows the uncertainty range of prediction considering the 95th and 5th percentile respectively. Deviation of the predicted value from the corresponding observed value is shown in terms of error as depicted by the color gradient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analysis is designated as S1). Furthermore, Fig. 11 shows that the BMA based temporal network approach successfully captures the mean and range of the SSAI values considering all the months of analysis. Moving our attention to the low flow months, it is clearly evident that the BMA based time-invariant network approach gives sub-power performance due to its inability to identify the temporal change in the causes-effect relationship. Moving on to the ML based time-varying approaches, the results indicate that for the low flow months, the model is able to reasonably capture the mean and range of SSAI, except for the months of January and May. However, considering the high flow months, performance of the proposed model is better as compared to the ML based time-varying approaches. Month-wise analysis is also carried out with non-detrended series and the results are presented in Table 1 (data series for month-wise analysis is designated as S1_ndt) for all the four models. It may be observed that the analysis carried out after removing the trend shows a slightly improved performance as compared to using the non-detrended series. Overall, the ability of the proposed approach to capture the true dependence structure between the large pool of hydro-meteorological variables and drought index improves the prediction skill of the proposed model.

Next, additional analysis is carried out considering SSAI for all the months as a single series. Having established the efficacy of the time-varying approach over the time-invariant counterpart results obtained using the three time varying approaches (temporal network, SVR and ANN based) considering all the months as a single series is shown in Fig. 12a. While carrying out the month-wise analysis it was observed that a unique set of predictors are identified for each month considering

a particular time-period. When SSAI for each month is analyzed as a single series, it is observed that the set of predictors identified are dominated by the fact that most of the months (November–May) are low flow months as compared to the monsoon season (June–September). Thereby, the model performance deteriorates (the static value using the proposed approach are: $R = 0.69$, $RMSE = 0.70$, $NSE = 0.56$, and $Dr = 0.62$; Table 1) in comparison to the month-wise analysis. The model is unable to specially capture the below/above-normal flow events for the high flow months due to the inability of the model to appropriately identify the predictor set. Fig. 12a also compares the performance of the three time-varying models utilizing the box-plots. In terms of capturing the mean and range of the observed data, the performance of all the models are comparable. However, when comparing the performance statics (considering the SVR (ANN) based time-varying approach and monthly drought index as a single series are: $R = 0.50$ (0.45), $RMSE = 0.89$ (0.99), $NSE = 0.40$ (0.38), and $Dr = 0.43$ (0.41); Table 1) it is observed that the proposed model out-performs the ML based time-varying approaches due to its ability to appropriately identify the complex mechanism associated with hydrologic drought.

The prediction models developed so far can be utilized to provide 1-month ahead forecast of below/above normal flow events. In order to provide an outlook of the model performance with a longer lead time, results for 3-month ahead forecast are also presented in Table 1 and Fig. 12b. Considering a longer lead time (the static value using the proposed approach are: $R = 0.75$, $RMSE = 0.88$, $NSE = 0.66$, and $Dr = 0.73$; Table 1), the model performance gradually deteriorates. The primary reason being the increase in the lag time of the influencing hydro-

Table 1
Performance statistics for the different models and the four different data series.

Lead time	Data Series	Model	Performance statistics				
			R	RMSE	NSE	Dr	R ²
1 month	S1	M1	0.90	0.67	0.80	0.83	0.762
		M2	0.75	0.81	0.55	0.69	0.521
		M3	0.70	0.82	0.52	0.63	0.471
		M4	0.68	0.85	0.51	0.63	0.442
	S1_ndt	M1	0.89	0.67	0.79	0.82	0.760
		M2	0.73	0.82	0.54	0.69	0.510
		M3	0.69	0.86	0.51	0.63	0.461
		M4	0.65	0.89	0.50	0.61	0.420
	S2	M1	0.69	0.70	0.56	0.62	0.458
		M2	0.50	0.89	0.40	0.43	0.239
		M3	0.45	0.99	0.38	0.41	0.186
		M4	0.43	0.99	0.35	0.42	0.180
	S2_ndt	M1	0.67	0.71	0.55	0.62	0.450
		M2	0.49	0.91	0.39	0.42	0.230
		M3	0.43	0.98	0.36	0.40	0.185
		M4	0.41	1.01	0.35	0.40	0.170
3 months	S1	M1	0.75	0.88	0.66	0.73	0.521
		M2	0.63	1.02	0.47	0.54	0.402
		M3	0.63	1.10	0.42	0.53	0.396
		M4	0.54	1.15	0.32	0.45	0.245
	S1_ndt	M1	0.75	0.89	0.64	0.73	0.522
		M2	0.62	1.06	0.45	0.54	0.395
		M3	0.61	1.10	0.41	0.52	0.391
		M4	0.50	1.21	0.29	0.43	0.242
	S2	M1	0.53	1.02	0.44	0.49	0.251
		M2	0.42	1.25	0.32	0.40	0.204
		M3	0.43	1.28	0.33	0.43	0.185
		M4	0.30	1.40	0.27	0.31	0.162
	S2_ndt	M1	0.51	1.09	0.42	0.48	0.247
		M2	0.42	1.24	0.35	0.40	0.190
		M3	0.41	1.29	0.34	0.41	0.182
		M4	0.29	1.45	0.25	0.28	0.150

(i) M1: BMA based temporal network approach, M2: SVR based time-varying approach, M3: ANN based time-varying approach, M4: BMA based time-invariant network approach.

(ii) S1: Month-wise data, S2: All months as single series, S1_ndt: Month-wise non-detrended data, S2_ndt: All months as single series non-detrended data.

meteorological variable and SSAI. However, given the complex nature of tertiary hydrological variables, such as floods and droughts, the model very well captures the extreme flow events for the considered study area. With the availability of the data for the hydroclimatic forcings, that influence the hydrologic droughts, from different sources like the

ground observations and reanalysis products, effective forecast with a lead time of 1 to 3-months can be provided using the developed model. Further on availability of the observed streamflow data the model should be re-calibrated at span of 2–3 years based on the identified ORI of model re-calibration in order to capture time-varying association among the hydroclimatic variables and extreme events. This time span may change based on the study area considered.

5. Summary and conclusions

Inherent non-stationarity owing to the impact of climate and terrestrial changes causes a gradual change in the characteristics of hydrological extremes. This study proposes a BMA based temporal network approach as an efficient modelling technique for capturing such slow moving changes and prediction of hydrologic extremes by assessing the time-varying cause-effect relationship between hydroclimatic variables and extreme events. Changing climate and dynamic terrestrial environment suggests a change in the characteristics of extreme events like flood and drought. Considering the river-basin used for this study, results indicate temporal change in the characteristics of extreme events. An increase in the frequency of occurrence during the time period of 2001–2018 (with respect to the first model development period of 1971–2000) of below-normal flow events can be observed with an increased severity especially during the monsoon season (June–September). Studies have established that change in climatic factors, such as high temperature, precipitation deficit, increased evaporation, reduced runoff and infiltration, caused due to combined effect of climate variability and human activities has reduced streamflow which might be a primary cause for increase in hydrologic drought (Kim and Jehanzaib, 2020). Similar observations of change in frequency and severity are also made considering the non-monsoon months, hence establishing the temporality associated with the characteristics of extremes. Further, the benefit of applying the time-varying concept in prediction of extreme events is established by carrying out basin scale prediction. In the proposed time-varying approach BMA is used to develop the network/graph structures that are re-iteratively updated to obtain a series of networks (temporal networks). This helps to establish the change in the causality of extreme events. In this technique, instead of learning a single static high scoring graph structure, the information from multiple probable graph structures is used to obtain the final graph structure, that are updated after a fixed time interval, which reduces the uncertainty while dealing with tertiary hydrologic variables like flood and drought and capture the temporal change in the causality of the extreme events in a

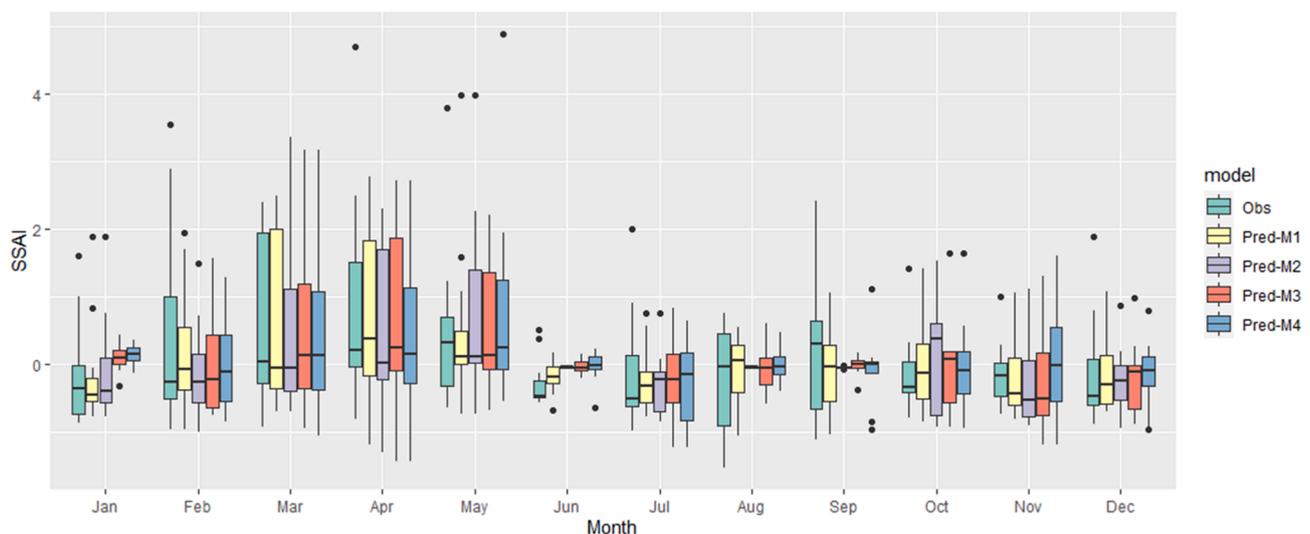


Fig. 11. Comparison of the observed SSAI (Obs) and predicted SSAI obtained using the proposed BMA based temporal network approach (Pred-M1), BMA based time-invariant network approach (Pred-M2), SVR based time-varying approach (Pred-M3) and ANN based time-varying approach (Pred-M4).

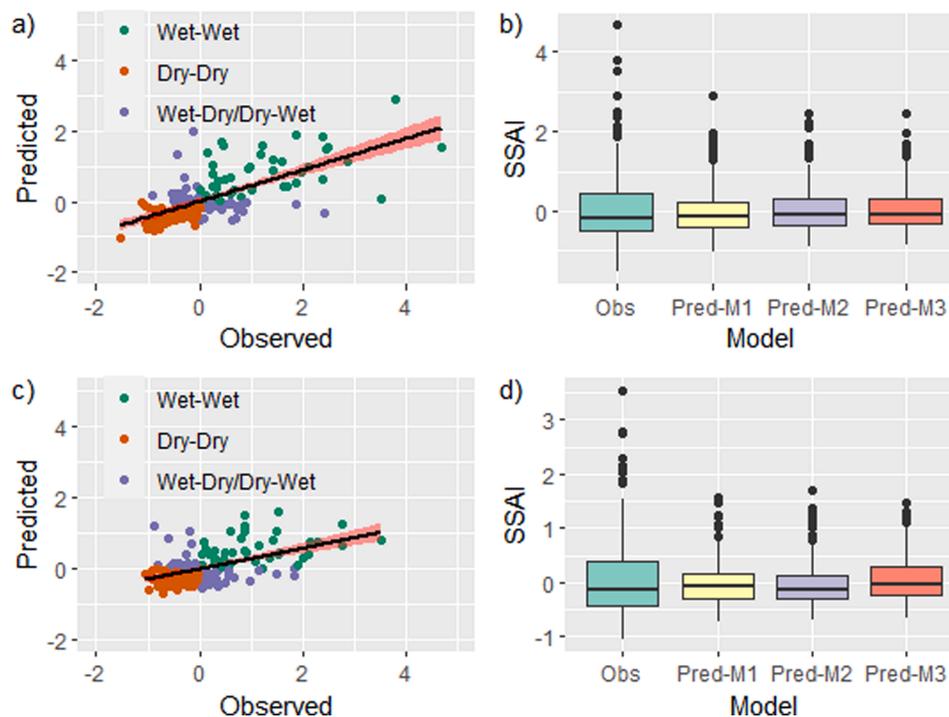


Fig. 12. Comparison of the observed and predicted SSAI obtained using the proposed BMA based temporal network approach (Pred-M1), SVR based time-varying approach (Pred-M2) and ANN based time-varying approach (Pred-M3) for the drought index for all the months as a single series, a) & b) 1-month lead and c) & d) 3-month lead. The scatter plot shows the results obtained using the proposed model.

non-stationary environment.

As a typical example, 1-month ahead hydrologic drought prediction defined by a streamflow based index (SSAI) is carried out to study the efficacy of the proposed approach. Applying the concept of BMA based temporal networks, the results indicate that the proposed prediction model, in terms of model inputs and parameters, needs to be recalibrated every 2 years and 3 years considering the high and low flow months respectively, in order to appropriately capture the time-varying association between the input and target variables. Based on the time-varying dependence structure obtained among the variables, drought index (SSAI) for the low flow months are strongly associated with streamflow. Contrary to this, considering the high flow months the dominant predictors are rainfall, precipitable water and relative humidity. Results also show that the BMA based temporal network approach successfully captures the extreme events associated with both low and high flows. Further, dividing the extreme events into different classes based on the severity of below/above-normal flow events, the proposed model shows satisfactory agreement with the observed events. That is, the predicted SSAI values very well captures the severity of the below/above-normal flow events, with higher efficacy considering the above-normal flow events. The ability of the proposed approach to identify the conditional independence structure among a large pool of associated variables, considering the complexity associated with tertiary hydrologic variables and capture the time-varying association among the variables can be effective for analyzing complex hydrologic processes. Overall, the findings of this study establish (i) the change in the drought frequency and severity over time, and (ii) benefit of the time-varying concept as an efficient approach for such complex hydrologic extremes exhibiting gradual change in its characteristics. Whereas the

first issue is being realized/established for many hydrological variables at different places around the world, the second aspect is a valuable contribution in terms of a remedial measure to handle such cases through a proper assessment of time-varying cause-effect relationship between hydroclimatic variables and extreme events.

CRediT authorship contribution statement

Riya Dutta: Methodology, Investigation, Writing - original draft.
Rajib Maity: Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the Ministry of Earth Science, Government of India through a sponsored project. Daily streamflow data are collected from the Water Resources Information System in India at <http://www.india-wris.nrsc.gov.in/>. Daily rainfall data are obtained from India Meteorological Department (IMD). The climatic variables are obtained from the Climate Prediction Centre (CPC) of the National Oceanic and Atmospheric Administration (NOAA) at <https://www.cpc.ncep.noaa.gov/>.

Appendix A. Mathematical details for development of the BMA based network approach

The following section gives the mathematical details for evaluation of the BDe score, evaluation of the group score, development of the Markov-chain and factorization of the final graph structure.

Evaluation of the BDe score

The score of each graph structure, also known as the BDe score, (Geiger and Heckerman, 2002) is the posterior probability of the structure evaluated as,

$$P(G/D) \propto P(D/G)P(G) = \prod_{i=1}^n S(X_i, Pa_i/D) \tag{1}$$

where, G refers to the network structure among the variables, D represents the data set, S is a score function dependent on variable X_i , and Pa_i is the parent set of variable X_i .

Scoring of the topological orders (Θ)

Each order Θ receives a score $R(\Theta/D)$ equal to the sum of the scores of all graph structures in the order.

$$R(\Theta/D) = \sum_{G \in \Theta} P(G/D) \tag{2}$$

The score of each node's possible parent sets is computed and efficiently used as the sum of the scores of all the graph structures compatible with a particular order (Friendman and Koller, 2003). The score of the order is evaluated as the product of the node score sums over possible parent sets.

$$R(\Theta/D) = \sum_{G \in \Theta} P(G/D) \propto \prod_{i=1}^n \sum_{Pa \in \Theta_i} S(X_i, Pa_i/D) \tag{3}$$

Construction of the Markov chain

A chain with stationary distribution proportional to $R(\Theta/D)$ is produced by a Metropolis-Hastings algorithm with acceptance probability,

$$\rho = \min \left\{ 1, \frac{q(\Theta_j/\Theta^*)R(\Theta^*/D)}{q(\Theta^*/\Theta_j)R(\Theta_j/D)} \right\} \tag{4}$$

where $q(\Theta^*/\Theta_j)$ is the probability of proposing a move to Θ^* from Θ_j , and can be any move in the space of orders. The simplest move is flipping two nodes in the order while leaving the position of others unchanged. Upon convergence, order-MCMC provides a sample of an order Θ^* from a distribution proportional to the score $R(\Theta/D)$ over the space of possible orders of the nodes of the graph structure. Given a sampled order, a network structure is selected by sampling the parents of each node independently according to the scores of its permissible parent sets.

Factorization and parameter learning of the final graph structure

Let X_1, X_2, \dots, X_N be N random variables with a known graph structure (directed graph). The joint probability distribution, also known as the global probability distribution, depends on a set of local probability distributions, one for each node/variable. It can be expressed as follows,

$$P(X_1, X_2, \dots, X_N) = \prod_{i=1}^n p(X_i/pa_i) \tag{5}$$

where $p(X_i/pa_i)$ is the local conditional probability associated with node i . A specific form of the factorization, given by the Markov property of BNs (Korb and Nicholson, 2004), states that every random variable X_i , directly depends only on its parents. Thereby the conditional probability of any variable X_i given rest of the variables can be written as,

$$p(X_i/rest) = p(X_i/pa_i) \tag{6}$$

In order to identify the local probability distribution two methods namely, Bayesian parameter estimation and maximum likelihood parameter estimation can be utilized (Scutari, 2017, 2010). In this study, MLE is used to learn the parameter set Φ of the joint probability distribution function of the variables that can be represented as,

$$P(X_1, X_2, \dots, X_N/\Phi) = \prod_{i=1}^n p(X_i/pa_i, \theta_i) \tag{7}$$

where θ_i is the vector of parameters for the conditional distribution of x_i and $\Phi = (\theta_1, \dots, \theta_N)$. Given the training data $X = \{x_1, \dots, x_m\}$ where $x_l = (x_{l1}, \dots, x_{ln})'$ and M is the number of observations for each variable, the log-likelihood of Θ with rest to X is computed as,

$$\log \left(\frac{\Phi}{X} \right) = \sum_{l=1}^M \sum_{i=1}^N \log p(X_{li}/pa_i, \theta_i) \tag{8}$$

The likelihood function as given in equation 5, which decomposes according to the network structure, thereby, the MLEs for each node is computed independently (Neapolitan and Jiang, 2007). The local probability distribution is extracted from the fitted probabilistic graphical model (Scutari, 2017). The probabilistic distribution is used to predict the target variables given the directly influencing input variables.

References

- Apurv, T., Cai, X., 2019. Evaluation of the stationarity assumption for meteorological drought risk estimation at the multidecadal scale in contiguous United States. *Water Resour. Res.* 55, 5074–5101. <https://doi.org/10.1029/2018WR024047>.
- Ardabili, S., Mosavi, A., Dehghani, M., Varkonyi-koczy, A.R., 2019. Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review 1–24. <https://doi.org/10.20944/preprints201908.0166.v1>.
- Avilés, A., Celleri, R., Solera, A., Paredes, J., 2016. Probabilistic forecasting of drought events using Markov chain- and Bayesian network-based models: A case study of an Andean regulated river basin. *Water (Switzerland)* 8 (2), 37. <https://doi.org/10.3390/w8020037>.
- Barros, A.P., Bowden, G.J., 2008. Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling River Basin. *J. Hydrol.* 357 (3–4), 349–367. <https://doi.org/10.1016/j.jhydrol.2008.05.026>.
- Barua, S., Ng, A.W.M., Perera, B.J.C., 2012. Artificial neural network-based drought forecasting using a nonlinear aggregated drought index. *J. Hydrol. Eng.* 17 (12), 1408–1413. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000574](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000574).
- Betterle, A., Radny, D., Schirmer, M., Botter, G., 2017. What do they have in common? Drivers of streamflow spatial correlation and prediction of flow regimes in ungauged locations. *Water Resour. Res.* 53 (12), 10354–10373. <https://doi.org/10.1002/2017WR021144>.
- Bowden, G.J., Dandy, G.C., Maier, H.R., 2005. Input determination for neural network models in water resources applications. Part 1 - Background and methodology. *J. Hydrol.* 301 (1–4), 75–92. <https://doi.org/10.1016/j.jhydrol.2004.06.021>.
- Chebana, F., Ouara, T.B.M.J., 2021. Multivariate non-stationary hydrological frequency analysis. *J. Hydrol.* 593, 125907. <https://doi.org/10.1016/j.jhydrol.2020.125907>.
- Cook, B.I., Mankin, J.S., Anchukaitis, K.J., 2018. Climate change and drought: from past to future. *Curr. Clim. Chang. Reports* 4 (2), 164–179. <https://doi.org/10.1007/s40641-018-0093-2>.
- Cook, B.I., Smerdon, J.E., Seager, R., Coats, S., 2014. Global warming and 21st century drying. *Clim. Dyn.* 43 (9–10), 2607–2627. <https://doi.org/10.1007/s00382-014-2075-y>.
- Cooper, G.F., 1990. The computational complexity of probabilistic inference using bayesian belief networks. *Artif. Intell.* 42 (2–3), 393–405. [https://doi.org/10.1016/0004-3702\(90\)90060-D](https://doi.org/10.1016/0004-3702(90)90060-D).
- CPC. (2014). Soil Moisture (V2), NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. Retrieved November 1, 2014, from <http://www.esrl.noaa.gov/psd/data/gridded/data.cpcsoil.html>.
- Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge.
- Das, P., Chanda, K., 2020. Bayesian Network based modeling of regional rainfall from multiple local meteorological drivers. *J. Hydrol.* 591, 125563. <https://doi.org/10.1016/j.jhydrol.2020.125563>.
- Das, S., Das, J., Umamahesh, N.V., 2021. Nonstationary Modeling of Meteorological Droughts: Application to a Region in India. *J. Hydrol. Eng.* 26 (2), 05020048. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002039](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002039).
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* 30 (5), 1371–1386. <https://doi.org/10.1016/j.advwatres.2006.11.014>.
- Dutta, R., Maity, R., 2020a. Temporal networks based approach for non-stationary hydroclimatic modelling and its demonstration with streamflow prediction. *Water Resour. Res.* 0–1. <https://doi.org/10.1029/2020wr027086>.
- Dutta, R., Maity, R., 2020b. Identification of potential causal variables for statistical downscaling models: effectiveness of graphical modeling approach. *Theor. Appl. Climatol.* 142 (3–4), 1255–1269. <https://doi.org/10.1007/s00704-020-03372-4>.
- Dutta, R., Maity, R., 2020c. Spatial variation in long-lead predictability of summer monsoon rainfall using a time-varying model and global climatic indices. *Int. J. Climatol.* 40 (14), 5925–5940. <https://doi.org/10.1002/joc.v40.1410.1002/joc.6556>.
- Dutta, R., Maity, R., 2018. Temporal evolution of hydroclimatic teleconnection and a time-varying model for long-lead prediction of Indian summer monsoon rainfall. *Sci. Rep.* 8, 10778. <https://doi.org/10.1038/s41598-018-28972-z>.
- Fan, Y., van den Dool, H., 2004. Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. *Journal of Geophysical Research D: Atmospheres* 109 (10), 1–8. <https://doi.org/10.1029/2003JD004345>.
- Friedman, N., Koller, D., 2003. Being Bayesian about network structure. *Mach. Learn.* 50, 95–125. <https://doi.org/10.1023/A:1020249912095>.
- Fung, K.F., Huang, Y.F., Koo, C.H., Mirzaei, M., 2020. Improved svr machine learning models for agricultural drought prediction at downstream of langat river basin. *Malaysia. J. Water Clim. Chang.* 11, 1383–1398. <https://doi.org/10.2166/wcc.2019.295>.
- Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. *Environ. Model. Softw.* 62, 33–51. <https://doi.org/10.1016/j.envsoft.2014.08.015>.
- Gibbs, M.S., McInerney, D., Humphrey, G., Thyer, M.A., Maier, H.R., Dandy, G.C., Kavetski, D., 2018. State updating and calibration period selection to improve dynamic monthly streamflow forecasts for an environmental flow management application. *Hydrol. Earth Syst. Sci.* 22 (1), 871–887. <https://doi.org/10.5194/hess-22-871-2018>.
- Hao, Z., Hao, F., Singh, V.P., 2016. A general framework for multivariate multi-index drought prediction based on Multivariate Ensemble Streamflow Prediction (MESP). *J. Hydrol.* 539, 1–10. <https://doi.org/10.1016/j.jhydrol.2016.04.074>.
- Hao, Z., Singh, V.P., Xia, Y., 2018. Seasonal drought prediction: advances, challenges, and future prospects. *Rev. Geophys.* 56 (1), 108–141. <https://doi.org/10.1002/rog.v56.110.1002/2016RG000549>.
- He, B., Chang, J., Wang, Y., Wang, Y., Zhou, S., Chen, C., 2021. Spatio-temporal evolution and non-stationary characteristics of meteorological drought in inland arid areas. *Ecol. Indic.* 126, 107644. <https://doi.org/10.1016/j.ecolind.2021.107644>.
- Heckerman, D., Mamdani, A., Wellman, M.P., 1995. Real-world applications of Bayesian networks. *Commun. ACM - CACM.* doi 38 (3), 24–26.
- Hesarkazzazi, S., Arabzadeh, R., Hajibabaei, M., Rauch, W., Kjeldsen, T.R., Prodocimi, I., Castellarin, A., Sitzfreni, R., 2021. Stationary vs non-stationary modelling of flood frequency distribution across northwest England. *Hydrol. Sci. J.* 66 (4), 729–744. <https://doi.org/10.1080/02626667.2021.1884685>.
- Huo, W., Li, Z., Wang, J., Yao, C., Zhang, K.E., Huang, Y., 2019. Multiple hydrological models comparison and an improved Bayesian model averaging approach for ensemble prediction over semi-humid regions. *Stoch. Environ. Res. Risk Assess.* 33 (1), 217–238. <https://doi.org/10.1007/s00477-018-1600-7>.
- Hwang, T., Martin, K.L., Vose, J.M., Wear, D., Miles, B., Kim, Y., Band, L.E., 2018. Nonstationary hydrologic behavior in forested watersheds is mediated by climate-induced changes in growing season length and subsequent vegetation growth. *Water Resour. Res.* 54 (8), 5359–5375. <https://doi.org/10.1029/2017WR022279>.
- Ihler, A.T., Kirshner, S., Ghil, M., Robertson, A.W., Smyth, P., 2007. Graphical models for statistical inference and data assimilation. *Phys. D Nonlinear Phenom.* 230 (1–2), 72–87. <https://doi.org/10.1016/j.physd.2006.08.023>.
- Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., Ma, M., 2012. Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *J. Hydrol.* 452–453, 213–225. <https://doi.org/10.1016/j.jhydrol.2012.05.055>.
- Jiang, Z., Sharma, A., Johnson, F., 2020. Refining predictor spectral representation using wavelet theory for improved natural system modeling. *Water Resour. Res.* 56, 1–17. <https://doi.org/10.1029/2019WR026962>.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kaur, A., Sood, S.K., 2020. Deep learning based drought assessment and prediction framework. *Ecol. Inform.* 57, 101067. <https://doi.org/10.1016/j.ecoinf.2020.101067>.
- Khan, N., Sachindra, D.A., Shahid, S., Ahmed, K., Shire, M.S., Nawaz, N., 2020. Prediction of droughts over Pakistan using machine learning algorithms. *Adv. Water Resour.* 139, 103562. <https://doi.org/10.1016/j.advwatres.2020.103562>.
- Kiem, A.S., Johnson, F., Westra, S., van Dijk, A., Evans, J.P., O'Donnell, A., Rouillard, A., Barr, C., Tyler, J., Thyer, M., Jakob, D., Woldemeskel, F., Sivakumar, B., Mehrotra, R., 2016. Natural hazards in Australia: droughts. *Clim. Change* 139 (1), 37–54. <https://doi.org/10.1007/s10584-016-1798-7>.
- Kim, T.W., Jehanzaib, M., 2020. Drought risk analysis, forecasting and assessment under climate change. *Water (Switzerland)* 12, 1–7. <https://doi.org/10.3390/w12071862>.
- Korb, K., Nicholson, A., 2004. *Bayesian Artificial Intelligence*. Chapman & Hall/CRC Press, London.
- Liu, W.T., Juárez, R.I.N., 2001. ENSO drought onset prediction in northeast Brazil using NDVI. *Int. J. Remote Sens.* 22 (17), 3483–3501. <https://doi.org/10.1080/01431160010006430>.
- Liu, Z., Zhou, P., Chen, X., Guan, Y., 2015. A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *J. Geophys. Res. Atmos.* 120, 10116–10129. <https://doi.org/10.1002/2015JD023787>.
- Livneh, B., Hoerling, M.P., 2016. The physics of drought in the U.S. Central Great Plains. *J. Clim.* 29, 6783–6804. <https://doi.org/10.1175/JCLI-D-15-0697.1>.
- Lu, Y., Jiang, S., Ren, L., Zhang, L., Wang, M., Liu, R., Wei, L., 2019. Spatial and Temporal variability in precipitation concentration over mainland China, 1961–2017. *Water (Switzerland)* 11. <https://doi.org/10.3390/w11050881>.
- Luo, L., Apps, D., Arcand, S., Xu, H., Pan, M., Hoerling, M., 2017. Contribution of temperature and precipitation anomalies to the California drought during 2012–2015. *Geophys. Res. Lett.* 44 (7), 3184–3192. <https://doi.org/10.1002/2016GL072027>.
- Machado, M.J., Botero, B.A., López, J., Francés, F., Díez-Herrero, A., Benito, G., 2015. Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrol. Earth Syst. Sci.* 19, 2561–2576. <https://doi.org/10.5194/hess-19-2561-2015>.
- Madadgar, S., Moradkhani, H., 2013. A Bayesian framework for probabilistic seasonal drought forecasting. *J. Hydrometeorol.* 14, 1685–1705. <https://doi.org/10.1175/JHM-D-13-010.1>.
- Maity, R., Bhagwat, P.P., Bhatnagar, A., 2010. Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrol. Process.* 24 (7), 917–923. <https://doi.org/10.1002/hyp.v24:710.1002/hyp.7535>.
- Maity, R., Kashid, S.S., 2011. Importance analysis of local and global climate inputs for basin-scale streamflow prediction. *Water Resour. Res.* 47, 1–17. <https://doi.org/10.1029/2010WR009742>.
- Maity, R., Khan, M.I., Sarkar, S., Dutta, R., Maity, S.S., Pal, M., Chanda, K., 2021. Potential of Deep Learning in drought assessment by extracting information from hydrometeorological precursors. *J. Water Clim. Chang.* <https://doi.org/10.2166/wcc.2021.062>.
- May, R., Dandy, G., Maier, H., 2011. Review of input variable selection methods for artificial neural networks. *Artif. Neural Networks - Methodol. Adv. Biomed. Appl.* <https://doi.org/10.5772/16004>.

- Meenu, R., Rehana, S., Mujumdar, P.P., 2013. Assessment of hydrologic impacts of climate change in Tunga-Bhadra river basin, India with HEC-HMS and SDSM. *Hydrol. Process.* 27, 1572–1589. <https://doi.org/10.1002/hyp.9220>.
- Meira Neto, A.A., Oliveira, P.T.S., Rodrigues, D.B.B., Wendland, E., 2018. Improving streamflow prediction using uncertainty analysis and Bayesian model averaging. *J. Hydrol. Eng.* 23 (5), 05018004. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001639](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001639).
- Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: whither water management? *Science* (80-) 319 (5863), 573–574. <https://doi.org/10.1126/science.1151915>.
- Mishra, A.K., Desai, V.R., 2006. Drought forecasting using feed-forward recursive neural network. *Ecol. Modell.* 198 (1–2), 127–138. <https://doi.org/10.1016/j.ecolmodel.2006.04.017>.
- Mishra, A.K., Desai, V.R., Singh, V.P., 2007. Drought forecasting using a hybrid stochastic and neural network model. *J. Hydrol. Eng.* 12 (6), 626–638. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:6\(626\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:6(626)).
- Morrison, R., Stone, M., 2014. Spatially implemented Bayesian network model to assess environmental impacts of water management. *Water Resour. Res.* 50, 8107–8124. <https://doi.org/10.1002/2014WR015600>. Received.
- Mukherjee, S., Mishra, A., Trenberth, K.E., 2018. Climate change and drought: a perspective on drought indices. *Curr. Clim. Chang. Reports* 4 (2), 145–163. <https://doi.org/10.1007/s40641-018-0098-x>.
- Neapolitan, R.E., Jiang, X., 2007. Bayesian Networks. *Probabilistic Methods Financ. Mark. Informatics* 53–110. <https://doi.org/10.1016/b978-012370477-1.50020-7>.
- Panu, U.S., Sharma, T.C., 2002. Défis de la recherche sur les sécheresses: Quelques perspectives et directions futures. *Hydrol. Sci. J.* 47, S19–S30. <https://doi.org/10.1080/02626660209493019>.
- Pathiraja, S., Marshall, L., Sharma, A., Moradkhani, H., 2016. Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. *Adv. Water Resour.* 94, 103–119. <https://doi.org/10.1016/j.advwatres.2016.04.021>.
- Pichuka, S., Prasad R, R., Maity, R., Kunstmann, H., 2017. Development of a method to identify change in the pattern of extreme streamflow events in future climate: Application on the Bhadra reservoir inflow in India. *J. Hydrol. Reg. Stud.* 9, 236–246. <https://doi.org/10.1016/j.ejrh.2016.12.084>.
- Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmos. Res.* 197, 42–63. <https://doi.org/10.1016/j.atmosres.2017.06.014>.
- Qu, B.o., Zhang, X., Pappenberger, F., Zhang, T., Fang, Y., 2017. Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian model averaging. *Water (Switzerland)* 9 (2), 74. <https://doi.org/10.3390/w9020074>.
- Raghavendra, S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* J. 19, 372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>.
- Rajeevan, M., Bhate, J., Jaswal, A.K., 2008. Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett.* 35, 1–6. <https://doi.org/10.1029/2008GL035143>.
- Ramadas, M., Govindaraju, R.S., 2015. Probabilistic assessment of agricultural droughts using graphical models. *J. Hydrol.* 526, 151–163. <https://doi.org/10.1016/j.jhydrol.2014.09.026>.
- Ramadas, M., Maity, R., Ojha, R., Govindaraju, R.S., 2015. Predictor selection for streamflows using a graphical modeling approach. *Stoch. Environ. Res. Risk Assess.* 29 (6), 1583–1599. <https://doi.org/10.1007/s00477-014-0977-1>.
- Rehana, S., Mujumdar, P.P., 2014. Basin scale water resources systems modeling under cascading uncertainties. *Water Resour. Manag.* 28 (10), 3127–3142. <https://doi.org/10.1007/s11269-014-0659-2>.
- Sahu, N., Panda, A., Nayak, S., Saini, A., Mishra, M., Sayama, T., Sahu, L., Duan, W., Avtar, R., Behera, S., 2020. Impact of indo-pacific climate variability on high streamflow events in Mahanadi River Basin, India. *Water (Switzerland)* 12 (7), 1952. <https://doi.org/10.3390/w12071952>.
- Santos, J.F., Portela, M.M., Pulido-Calvo, I., 2014. Spring drought prediction based on winter NAO and global SST in Portugal. *Hydrol. Process.* 28 (3), 1009–1024. <https://doi.org/10.1002/hyp.v28.3.1002/hyp.9641>.
- Schisterman, E.F., Perkins, N.J., Mumford, S.L., Ahrens, K.A., Mitchell, E.M., 2017. Collinearity and causal diagrams. *Epidemiology* 28, 47–53. <https://doi.org/10.1097/EDE.0000000000000554>.
- Scutari, M., 2017. Bayesian network constraint-based structure learning algorithms: parallel and optimised implementations in the bnlearn R package. *J. Stat. Softw.* VV. <https://doi.org/10.18637/jss.v077.i02>.
- Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* 35, 1–22. <https://doi.org/10.18637/jss.v035.i03>.
- Sharma, A., Mehrotra, R., Li, J., Jha, S., 2016. A programming tool for nonparametric system prediction using Partial Informational Correlation and Partial Weights. *Environ. Model. Softw.* 83, 271–275. <https://doi.org/10.1016/j.envsoft.2016.05.021>.
- Sun, L., Mitchell, S.W., Davidson, A., 2012. Multiple drought indices for agricultural drought risk assessment on the Canadian prairies. *Int. J. Climatol.* 32 (11), 1628–1639. <https://doi.org/10.1002/joc.2385>.
- Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., Stephens, S., 2002. The drought monitor. *Bull. Am. Meteorol. Soc.* 83 (8), 1181–1190.
- Tian, J., He, R., Ram, L., 2010. Bayesian model averaging using the κ -best Bayesian network structures, in: Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, UAI 2010. pp. 589–597.
- Tian, Y., Xu, Y.P., Wang, G., 2018. Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin. *Sci. Total Environ.* 622–623, 710–720. <https://doi.org/10.1016/j.scitotenv.2017.12.025>.
- Van Lanen, H.A.J., Wanders, N., Tallaksen, L.M., Van Loon, A.F., 2013. Hydrological drought across the world: Impact of climate and physical catchment structure. *Hydrol. Earth Syst. Sci.* 17, 1715–1732. <https://doi.org/10.5194/hess-17-1715-2013>.
- Van Loon, A.F., Gleeson, T., Clark, J., Van Dijk, A.I.J.M., Stahl, K., Hannaford, J., Di Baldassarre, G., Teuling, A.J., Tallaksen, L.M., Uijlenhoet, R., Hannah, D.M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., Rangelcroft, S., Wanders, N., Van Lanen, H.A.J., 2016. Drought in the anthropocene. *Nat. Geosci.* 9 (2), 89–91. <https://doi.org/10.1038/ngeo2646>.
- Wagner, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao, P.S.C., Basu, N.B., Wilson, J.S., 2010. The future of hydrology: An evolving science for a changing world. *Water Resour. Res.* 46, 1–10. <https://doi.org/10.1029/2009WR008906>.
- Wang, Q.J., Robertson, D.E., Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resour. Res.* 45, 1–18. <https://doi.org/10.1029/2008WR007355>.
- Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, Amsterdam.
- Wu, L., Seo, D.-J., Demargne, J., Brown, J.D., Cong, S., Schaake, J., 2011. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. *J. Hydrol.* 399 (3–4), 281–298. <https://doi.org/10.1016/j.jhydrol.2011.01.013>.
- Xu, L., Chen, N., Zhang, X., Chen, Z., 2018. An evaluation of statistical, NMME and hybrid models for drought prediction in China. *J. Hydrol.* 566, 235–249. <https://doi.org/10.1016/j.jhydrol.2018.09.020>.
- Yan, J., Liao, G.Y., Gebremichael, M., Shedd, R., Vallee, D.R., 2012. Characterizing the uncertainty in river stage forecasts conditional on point forecast values. *Water Resour. Res.* 48, 1–11. <https://doi.org/10.1029/2012WR011818>.
- Yang, T., Zhou, X., Yu, Z., Krysanova, V., Wang, B.o., 2015. Drought projection based on a hybrid drought index using Artificial Neural Networks. *Hydrol. Process.* 29 (11), 2635–2648. <https://doi.org/10.1002/hyp.v29.11.1002/hyp.10394>.
- Ye, M., Neuman, S.P., Meyer, P.D., 2004. Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. *Water Resour. Res.* 40, 1–17. <https://doi.org/10.1029/2003WR002557>.
- Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schäfer, D., Marx, A., 2016. The German drought monitor. *Environ. Res. Lett.* 11 (7), 074002. <https://doi.org/10.1088/1748-9326/11/7/074002>.