





Bayesian network based procedure for regional drought monitoring

Ali, Zulfiqar; Hussain, Ijaz; Grzegorczyk, Marco Andreas; Ni, Guangheng; Faisal, Muhammad; Qamar, Sadia; Shoukry, Alaa Mohamd; Wahab Sharkawy, Mohammed Abdel; Gani, Showkat; Al-Deek, Fares Fawzi

Published in: Journal of Environmental Management

DOI: 10.1016/j.jenvman.2020.111296

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version Publisher's PDF, also known as Version of record

Publication date: 2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Ali, Z., Hussain, I., Grzegorczyk, M. A., Ni, G., Faisal, M., Qamar, S., Shoukry, A. M., Wahab Sharkawy, M. A., Gani, S., & Al-Deek, F. F. (2020). Bayesian network based procedure for regional drought monitoring: The Seasonally Combinative Regional Drought Indicator. *Journal of Environmental Management, 276*, [111296]. https://doi.org/10.1016/j.jenvman.2020.111296

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverneamendment.

Take-down policy If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: http://www.elsevier.com/locate/jenvman



Research article

Bayesian network based procedure for regional drought monitoring: The Seasonally Combinative Regional Drought Indicator

Zulfiqar Ali^{a,b,*}, Ijaz Hussain^b, Marco Andreas Grzegorczyk^c, Guangheng Niⁱ, Muhammad Faisal^{d,e}, Sadia Qamar^f, Alaa Mohamd Shoukry^{g,h}, Mohammed Abdel Wahab Sharkawy^g, Showkat Gani^j, Fares Fawzi Al-Deek^g

^a School of Statistics, Minhaj University, Lahore, 54000, Pakistan

^j College of Business Administration, King Saud University Muzahimiyah, Saudi Arabia

ARTICLE INFO

Keywords: Drought Standardized precipitation temperature index (SPTI) Bayesian networks Geospatial variation

ABSTRACT

Drought is a complex natural hazard. It occurs due to a prolonged period of deficient in rainfall amount in a certain region. Unlike other natural hazards, drought hazard has a recurrent occurrence. Therefore, comprehensive drought monitoring is essential for regional climate control and water management authorities. In this paper, we have proposed a new drought indicator: the Seasonally Combinative Regional Drought Indicator (SCRDI). The SCRDI integrates Bayesian networking theory with Standardized Precipitation Temperature Index (SPTI) at varying gauge stations in various month/seasons. Application of SCRDI is based on five gauging stations of Northern Area of Pakistan. We have found that the proposed indicator accounts the effect of climate variation within a specified territory, accurately characterizes drought by capturing seasonal dependencies in geospatial variation scenario, and reduces the large/complex data for future drought monitoring. In summary, the proposed indicator can be used for comprehensive characterization and assessment of drought a certain region.

1. Introduction

In many regions across the world, the risk of drought hazard has been increasing due to climate change and global warming (Turral et al., 2011; Marquina, 2010). Unlike other natural hazards, is a complex and recurrently occurring hazard (Tsakiris, 2017). Therefore, more comprehensive and rigorous procedures are required for drought monitoring, forecasting, and spatial analysis (Ali et al., 2019b, 2018). In the past three decades, numerous methods and tools have been developed for drought monitoring for various climatological regions. Some of them are Palmer Drought Severity Index (PDSI) (Wayne, 1965),

Keetch–Byram drought index (KBDI) (Keetch and Byram, 1968), Surface Water Supply Index (SWSI) (Dezman et al., 1983), Crop-specific Drought Index (CSDI) (Meyer et al., 1993), and Reconnaissance Drought Index (RDI) (Tsakiris et al., 2007). Svoboda and Fuchs (2016) discussed each of these indices and their data requirement. However, among all drought indices, Standardized Drought Indices (SDI) is one of the most commonly used procedures (Erhardt and Czado, 2018). Example of SDI indices includes Standardized Precipitation Index (SPI) (McKee et al., 1993), Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) and Standardized Precipitation Temperature Index (SPTI) (Ali et al., 2017). These standardized methods

https://doi.org/10.1016/j.jenvman.2020.111296

Received 31 October 2019; Received in revised form 18 August 2020; Accepted 20 August 2020 Available online 6 September 2020 0301-4797/© 2020 Elsevier Ltd. All rights reserved.

^b Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan

^c Johann Bernoulli Institute, Groningen University, 9747AG, Groningen, Netherlands

^d Faculty of Health Studies, University of Bradford, BD7 1DP, Bradford, UK

^e Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

^f University of Sargodha, Sargodha, Pakistan

^g Arriyadh Community College, King Saud University Riyadh, Saudi Arabia

^h KSA Workers University, Nsar, Egypt

ⁱ State Key Laboratory of Hydro-Science and Engineering and Dept. of Hydraulic Engineering, Tsinghua University, Beijing, 100084, China

^{*} Corresponding author. Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.

E-mail addresses: zulfiqarali@stat.qau.edu.pk (Z. Ali), ijaz@qau.edu.pk (I. Hussain), m.a.grzegorczyk@rug.nl (M.A. Grzegorczyk), ghni@tsinghua.edu.cn (G. Ni), m.faisal1@bradford.ac.uk (M. Faisal), diyaqau@gmai.com (S. Qamar), aabdulhamid@ksu.edu.sa (A.M. Shoukry), mosharkawy@ksu.edu.sa (M.A. Wahab Sharkawy), sgani@ksu.edu.sa (S. Gani), faldeek@ksu.edu.sa (F.F. Al-Deek).



Fig. 1. Locations of selected meteorological stations in north region of Pakistan.

provide sufficient evidence for the establishment of effective drought mitigation policies and early warning strategies.

Among other climatic and meteorological variables, the distribution of rainfall and temperature are the main factors of drought hazard (Vicente-Serrano et al., 2010; Easterling et al., 2007; Hu and Willson, 2000). These two climatic variables have great influence and substantial importance in continuous drought monitoring. Most SDI procedures use long time series data of precipitation and temperature (Guttman, 1999; Cancelliere and Salas, 2004; Hiscock, 2009). In this regards, regional drought monitoring involves accurate estimation of SDI indices based on regionally representative time series. However, climate cycle and spatio-temporal features of meteorological stations within a particular region are the main barriers in acquiring regionally representative time series data.

From geostatistical and data mining point of view, accurate estimation and continuous monitoring of drought at the regional level requires a dense meteorological network. In past few decades, several algorithms and methods for the optimal selection of meteorological stations have been the subject of great interest (Arsenault and Brissette, 2014; Lark, 2016; Hong et al., 2016). These algorithms and methods reduce the size of the network and provide more accurate and regionally representative estimates of meteorological variables. However, the implications in numerous meteorological stations (regardless of an optimal number of meteorological stations) at specified region need high cost, time and resources. Especially in developing countries, the high cost of equipment installations and complex sampling design may force to adopt the compromised solution. Further, the adoptions of these techniques raise several questions on the accuracy in measurements. For example, the rainfall is a spatial climatic variable that varies in both spatial and temporal dimensions. Its short distance variability characteristics and being a spatial variable suggest the complex structure of the meteorological network (Scarsoglio et al., 2013; Einfalt et al., 1990).

On the other hand, several studies have suggested advanced statistical and geostatistical methods for monitoring drought and other meteorological variables in various climatological regions (Bayat et al., 2015; Chen et al., 2017; Drogue et al., 2002; Miniscloux et al., 2001). However, these methods are based on temporal data records at a single station, that is, it only covers a single realization at a continuous spatial domain. This deprived the study and findings of the spatial prevalence effect of climate nature and climate change. Further, the increases in uncertainty in the estimation may lead bad impact on climate change policies and reliability of forecasting environmental characteristics. In addition, regional patterns of long-term rainfall and temperature at various seasons play an important role in the continuous drought monitoring. Therefore, drought management must define and monitor drought conditions using more representatively regional data and procedures. Consequently, a regionally representative meteorological observatory or location are required for accurate and precise quantification of drought.

In summary, various factors are involved to diminish the efficiency and accuracy of drought monitoring methods such as seasonal and geospatial climate variation, an inappropriate network of gauging locations, errors in the estimation phase of drought indices, lack of resource, and the creation of complex data at a particular geographical region. In these perspectives, advanced statistical procedures are helpful to find important and regionally representative gauge stations under complex meteorological network setting. This research aims to employ Bayesian network-based probabilistic model for regional drought monitoring. Bayesian network method has been widely used to establish a probabilistic relationship between variables. It is a powerful technique, which produces multivariate Joint Probability Distribution (JPD). These probabilities describe the dependency structure between variables (Pearl, 2014). Some recent applications of Bayesian network theory in various fields are available in Lee et al. (2019), Bertone et al. (2018), Moglia et al. (2018) and Liu et al. (2015).

In this paper, we proposed a new drought indicator - the Seasonally Combinative Regional Drought Indicator (SCRDI). SCRDI integrates Bayesian networking theory and SDI tools coupled with different seasons at varying gauged stations in a specified region. The organization of this article is as follows. Section 2 provides descriptions of material and methods. The proposed regional drought indicator: The Seasonally Combinative Regional Drought index (SCRDI) is presented in section 3. Section 4 and 5 elaborate results and discussion, respectively. Finally, we concluded our findings in section 6. Table 1

Summary statistics of precipitation and	temperature (Time Range,	1971–2017).
---	--------------------------	-------------

Stations	Variables	Statistics	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
	Precipitation	Average	39.25	47.27	78.13	86.53	69.81	25.12	24.79	24.9	23.61	23.2	17	28.12
		Std.	37.22	34.78	56.4	53.01	61.56	22.65	23.64	22.29	33.17	34.68	25.56	37.28
Astore	Tmin	Average	-7.36	-5.22	-0.9	3.76	7.43	11.22	14.49	14.55	10.36	4.52	-0.45	-4.61
		Std.	1.87	1.85	1.3	1.25	1.38	1.35	1.24	1.19	1.28	1.05	1.09	1.37
	Tmax	Average	2.82	4.47	8.99	15.02	19.78	24.58	27.17	26.89	23.42	17.65	11.35	5.43
		Std.	1.52	1.72	2	1.86	2.24	1.54	1.43	1.5	1.5	1.67	1.66	1.68
	Precipitation	Average	7.39	12.28	16.5	36.6	28.64	15.98	15.38	23.5	11.95	7.94	2.86	6.01
		Std.	10.29	23.12	17.41	62.3	28.71	24.12	17.08	25.11	18.41	14.46	5.94	9.57
Gupis	Tmin	Average	-5.63	-3.23	1.87	7.04	10.9	15.13	17.99	17.12	12.94	6.86	1.16	-3.45
		Std.	1.52	1.72	1.54	1.49	1.66	1.86	1.83	1.91	1.74	1.33	1.12	1.3
	Tmax	Average	4.34	6.79	12.42	18.56	23.29	28.67	31.71	30.77	26.64	20.28	13.77	6.44
		Std.	1.72	1.83	2	1.75	2.66	1.85	1.7	1.98	1.67	1.83	1.64	1.68
	Precipitation	Average	4.41	7.37	13.69	23.44	24.6	9.46	14.94	15.38	9.55	6.75	2.65	4.65
		Std.	6.26	8.44	17.55	24.39	24.49	9.23	14.38	16.01	10.78	14.16	5.17	7.45
Giligit	Tmin	Average	-2.48	0.75	5.79	9.48	12.12	15.15	18.39	17.85	12.97	6.87	0.98	-1.99
		Std.	1.75	1.44	1.28	1.08	1.23	1.48	1.66	1.6	1.64	1.62	1.61	1.96
	Tmax	Average	9.64	12.77	18.28	24.02	28.66	33.94	36.03	35.04	31.55	25.39	18.16	11.7
		Std.	1.65	1.78	1.97	1.84	2.7	1.72	1.78	1.75	1.92	2.02	1.77	1.45
	Precipitation	Average	28.53	27.62	38.1	30.79	28.39	9.04	11.01	13.28	12.21	7.06	5.28	15.6
		Std.	35.06	23.66	35.21	32.13	34.67	9.57	15.87	13.18	21.26	16.18	11.19	19.09
Skardu	Tmin	Average	-8.16	-4.56	1.49	6.46	9.59	13.26	16.34	15.93	11.49	4.34	-1.99	-5.74
		Std.	2.71	2.4	1.04	0.97	1.14	1.39	1.34	1.27	1.41	1.34	1.24	1.74
	Tmax	Average	3.09	6.03	12.33	18.67	23.17	28.42	31.56	31.01	27.12	20.37	13.15	6.47
		Std.	2.32	2.43	2.13	1.85	2.54	1.76	1.67	1.73	1.81	1.63	1.83	1.96
	Precipitation	Average	10.31	16.93	31.22	36.17	29.09	9.44	12.05	14.15	7.75	9.59	5.28	9.46
		Std.	14.59	16.66	30.13	33.21	33.53	12.2	15.37	16.02	17.21	22.25	11.37	16.02
Chilas	Tmin	Average	1.21	3.7	8.52	13.61	18.46	23.83	27.09	26.34	22.44	14.64	7.24	2.48
		Std.	1.29	1.23	1.33	1.49	2.22	1.85	1.41	1.5	1.44	1.38	1.06	1.42
	Tmax	Average	12.06	14.6	20	25.85	31.16	37	39.28	38.47	34.92	28.35	20.7	13.9
		Std.	1.43	1.74	2	1.9	2.64	1.9	1.68	1.64	1.56	1.73	1.42	1.31

2. Material and methods

2.1. Data and study area

The application of this research is based on five meteorological stations named as Astor (35.3570° N, 74.8624° E), Skardu (35.3247° N, 75.5510° E), Gupis (36.2274° N, 73.4421° E), Gilgit (35.8819° N, 74.4643° E), and Chilas (35.4222° N, 74.0946° E). These stations are located in the Northern part of Pakistan (see Fig. 1). Due to a significant change in climate and global warming scenario, the agricultural and industrial sectors of the whole geographical parts of the country are badly suffering from severe drought hazards. In addition, the prevalence of death is reported in many parts of the country. Especially in Tharparker (district of Sindh Province), numerous peoples including children died due to severe condition of drinking water shortage (Siddiqui and Safi, 2019). So, the continuity of drought for several future years will lead in lowering of per capita income that will result in poor health and nutrition especially for pregnant women and newly born children. In this scenario, prediction, continuous monitoring and accurate reporting of drought will strengthen drought mitigation policies. In this context, the selected regions consisting of five meteorological stations have a significant contribution to the whole part of the country and its climatology (Salma et al., 2012).

Table 1 shows the climatology of these stations in different sessions. One can observe that most of the stations have cold climate nearly all month.

2.2. The standardized drought indices (SDI)

During the past three decades, numerous methods developed for drought monitoring (Svoboda and Fuchs, 2016). The SDI method is one of the most commonly used drought monitoring method (Erhardt and Czado, 2018). From a data mining point of view, SDI mainly relies on univariate or multivariate data modeling. For univariate data modeling, several drought indices developed by various authors. For example, McKee et al. (1993) introduced the SPI drought index for the characterization of drought at various time scales. In SPI, monthly time series data of precipitation is standardized by an appropriate probability function. Vicente-Serrano et al. (2010) developed a standardized index-the SPEI. On the same line of SPI, SPEI is obtained by standardizing of water balance equation (see Eq. (1)).

$$DEF_i = P_i - PET_i \tag{1}$$

In the above equation, P_i is the monthly total amount of precipitation, PET_i is the estimated amount of Potential Evapotranspiration (PET). On the same line of McKee et al. (1993) and Vicente-Serrano et al. (2010), Ali et al. (2017) proposed another multi-scaler drought index called SPTI index.

For multivariate data modeling, there are several procedures and frameworks of drought indices. For example, Hao and AghaKouchak (2013) proposed Multivariate Standardized Drought Index (MSDI) by using copula function. Posteriorly, they modified MSDI by including the concept of non-parametric modeling (Hao and AghaKouchak, 2014). To resolve multiscaling issues in univariate SDI method, Bazrafshan et al. (2014) introduced a framework based on the principal component analysis. Consequently, they suggested a new multivariate drought index-the Multivariate Standardized Precipitation Index (MSPI).

In recent years, several drought monitoring studies and applications are based on SDI procedure. Some of them are Vicente-Serrano et al. (2018), Tigkas et al. (2018), Faiz et al. (2018) and Mathbout et al. (2018), Shen et al. (2017), Zarch et al. (2015), Golian et al. (2015).

This research is mainly based on the SPTI drought index. Contrary to SPI, SPTI is based one more than one meteorological variable. In addition, SPTI is an alternate of SPEI drought index particularly design for low-temperature regions (Ali et al., 2017). The indicator used in SPTI (see equation (9)) accounts the direct role of temperature in the selection of appropriate probability distribution. In our case, the selected study area has a cold climate and low temperature in most of the month (see Table 1). Therefore, the selection SPTI index is rationally valid. A brief estimation procedure of SPTI index is as follows:

The first step is to compute De Martonne Aridity Index (DAI) (De Martonne, 1926). (using monthly total precipitation and average



Fig. 2. An example of Bayesian network with five nodes (Grzegorczyk, 2010).

monthly temperature for each selected station separately.

$$DAI_i = \frac{P_i}{10 + T_i} \tag{2}$$

In the above equation, P_i is the monthly total precipitation and, T_i is the mean monthly temperature. In further steps, DAI_i series are used to obtain standardized values. In the computational procedure of SPTI, we have followed the standard guidelines of standardized drought index provided by Stagge et al. (2015). As our main focus is to give procedure of regional drought monitoring, therefore we include the only type of drought which is a meteorological index. To do this, we restrict our research on SPTI-1-time scale drought index. To estimate SPTI values, thirty-two distributions were used to assess their appropriation on the time series data of De Martonne index defined in equation (2) using *propagate* (Spiess, 2014) R package, separately for each station.

In this article, the specific computational results of SPTI index includes the selection of probability distribution for all the five stations. However, we have skipped the stepwise procedure of standardization of SPTI. One is referred to see the Ali et al. (2019a) and Ali et al. (2018) for a detailed overview and computational step of SPTI index.

After the selection of appropriate probability distributions, their numerical CDFs for all the station is standardized by the transformation function. See standardization procedure in Ali et al. (2017).

2.3. Bayesian networks and posterior probabilities

Bayesian networks are a class of probabilistic graphical models for leaning network topologies (Koller and Friedman, 2009; Pearl, 2014). Bayesian networks have applications in numerous multidisciplinary fields of research. For example, Bayesian networks are used for decision making (Watthayu and Peng, 2004; Kochenderfer, 2015), for prediction (Zhu et al., 2016; Borsuk et al., 2004; Mendes and Mosley, 2008), for anomaly detection (Wong et al., 2003; Sebyala et al., 2002), and reasoning (Neapolitan, 2012; Darwiche, 2009). In a nutshell, Bayesian networks can be used to learn the (un-)conditional dependencies among large numbers of interacting variables. In addition, these models quantify the dependencies between interacting variables in terms of so-called marginal edge posterior probabilities. The underlying concept is the Bayesian model averaging. Consequently, Bayesian network models are a flexible and promising statistical tool for reverse-engineering unknown network structures from data (Uusitalo, 2007). In this paper, we will apply static Bayesian network models to continuous data. We assume that the measurements are continuous and are independent (non-temporal) of domain variables. Here we only provide a brief overview of the Bayesian network methodology, but the readers are referred to a book chapter for a more detailed explanation on Bayesian networks models (Grzegorczyk, 2010).

Bayesian networks use directed acyclic graphs (DAGs) to describe the (un-)conditional independencies among a set of variables $(X_1, X_2, ..., X_n)$. More formally, let X denote a set of n random variables. The variables are then the nodes of the DAG and the directed edges encode the dependency structure among them. From the structure of the DAG (i.e. from its directed edges), it can be seen how the joint probability distribution over the n variables can be factorized into a product of so-called local conditional probability distributions. For instance, Fig. 2 shows a DAG for n = 5 variables (A, ..., E). In Fig. 2 nodes B and C are child nodes of node A, and vice-versa node A is called a parent node of B and C. Similarly, D is a child node of B and C, whereas, node E has parent node D.

In the factorization of the joint probability distribution of the n variables, each variable just depends on its parent nodes:

$$P_G(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | Pa(X_i, G))$$
(3)

where $Pa(X_i, G)$ denotes the set of parent nodes of node X_i in the directed acyclic graph G.

2.3.1. Learning the network structure

Conceptually, the edges of the graph G imply that the (un-)conditional dependencies among the domain variables are such that the factorization in Equation (2) is valid. The goal of Bayesian network analysis is to infer the underlying dependency structure (e.g. to infer the graph that best describes the dependencies among the observed variables).

Let *D* denote the observed data, i.e. a set of *m* independent *n*-dimensional vectors, whose entries are in one-to-one correspondence with the *n* variables. Within a Bayesian modeling framework, the posterior probabilities of a directed acyclic graph *G* is defined as:

$$P(G|D) = \frac{P(D|G) P(G)}{P(D)}$$

$$\tag{4}$$

where P(D|G) is the marginal likelihood, P(G) is the graph prior distribution, and P(D) is a normalization constant. The marginal likelihood is the probability of the data D given the graph G, i.e. it is the likelihood marginalized over all possible parameter instantiations that the graph G could have. Under certain modeling assumptions, the marginal likelihood can be computed analytically so that a closed-form solution is available. The graph prior distribution can be used to bring in pre-knowledge about the dependencies. If there is no genuine pre-knowledge available a uniform distribution over all graphs can be used.

When the data D is sparse, it is often not useful to search for the one single 'best' graph having the highest posterior probability. For sparse data, there might be various graphs that explain the data (approximately) equally well, so that they have (approximately) the same posterior probability in Equation (4). A more robust approach is then to average across all possible graphs, to identify whether there are some 'features' (e.g. particular edges) that all the 'good' graphs share. The marginal posterior probability of a feature is the sum of the posterior probabilities of all those graphs G that possess this feature. Mathematically this can be expressed as follows:

Table 2

BIC	values	of	candidate	distributions	for	all	stations.

Distributions	Astore	Chilas	Giligit	Skardu	Gupis
3P Weibul	-480.83	-379.15	-197.27	-516.14	-418.33
2P Beta	-280.1	-377.99	-160.31	-725.84	-267.36
4P beta	-469.79	-200	-195.95	-735.12	-427.46
Arcsine	-272.85	-39.21	42.22	-289.14	-185.31
Burr	-347.29	194.16	155.82	-459.33	-31.53
Cauchy	-367.96	-97.11	1.5	-750.69	-238.85
Chi	-247.93	-94.6	-91.1	-600.94	-221.81
Chi-Square	-441.42	-58.44	-42.06	-539.16	-169.83
Cosine	-192.31	207.11	160.59	-285.07	5.72
Curvilinear	-375.44	-148.24	-73.94	-403.58	-216.1
Trapezoidal					
Exponential	-401.81	33.01	5.47	-565.41	-221.69
F-	-457.56	7.74	8.62	-584.11	-146.43
Gamma	-477.56	-218.93	-145.18	-612.89	-324.71
Generalized Normal	-476.01	-338.36	-129.57	-783.71	-391.07
Generalized Extreme	-456.48	-266.33	-91.06	-762.49	-363.65
Value					
Gumbel	-389.62	-43.66	26.9	-680.05	-211.28
Inverse Gamma	-396.36	-233.31	-77.1	-627.29	-359.4
Inverse Gaussian	-356.23	-97.89	-33.2	-626.91	-356.78
Inverse Chi-Square	-273.05	41.18	90.63	-611.66	-121.2
Johnson SB	-478.62	-334.56	-128.21	-871.55	-393.17
Johnson SU	-471.37	-333.51	-125.04	-757.85	-386.63
Laplace	-382.24	-62.38	0.2	-641.06	-247.2
Logistic	-365.07	-31.88	39.87	-621.15	-191.2
Log-normal	-478.07	-263.65	-118.63	-625.55	-373.45
Normal	-357.27	$^{-16}$	50.35	-599.08	-175.39
Rayleigh	-373.83	-23.93	43.15	-290.05	-187.19
Scaled/shifted	-369.08	-118.3	-3.4	-761.36	-239.46
Skewed-normal	-384.42	-23.01	40.51	-718.66	-192.54
Trapezoidal	-361.71	13.94	60.23	-315.01	-156.1
Triangular	-366.35	18.2	55.7	-698.34	-160.54
Uniform	-274.09	105.68	107.14	-302.88	-47.07
von Mises	-381	16.07	50.27	-619.38	-176.07

$$P(f \mid D) = \sum_{G} P(G|D)f(G)$$
(5)

where the sum is over all possible directed acyclic graphs G, the posterior probability P(G|D) has been defined in Equation (4), and f is an indicator function for a specific feature (e.g. a particular edge). That is f(G) takes the value 1 if G possesses the feature, and f(G) = 0 otherwise. The higher the marginal feature posterior probability, the more certain we are about the presence of this feature. E.g. if a feature has probability 0.9, then this means that the posterior probabilities of the graphs that support the feature sum up to 0.9, whereas only 0.1 of the probability mass goes to graphs that do not support the feature.

2.3.2. Model averaging and structure MCMC

As the number of valid directed acyclic graphs (DAGs) grows superexponentially in the number of variables *n*, the edge feature probabilities in Equation (5) cannot be computed analytically. The sum is over too many graphs, and thus computationally not feasible. Henceforth, in practice, the marginal feature posterior distributions are approximated. To this end, a graph sample from the posterior distribution in Equation (4) is generated. This can be done by Markov Chain Monte Carlo (MCMC) simulations. Loosely speaking, the key idea is to design a Markov chain in the space of all valid DAGs, whose stationary distribution is the desired posterior distribution from Equation (5). Simulating this Markov chain yields a trajectory of DAGs. After an initial burn-in period, the stationary distribution (= posterior distribution) is reached, so that the frequency, with which each DAG appears in the trajectory, converges to the posterior probability of this DAG.

Various MCMC sampling algorithms have been developed for Bayesian networks (Madigan et al., 1995; Friedman and Koller, 2000; Grzegorczyk and Husmeier, 2008). By running one of those MCMC sampling algorithms, a sample of DAGs (G_1 , G_2 , G_3 , ..., G_T) from the posterior distribution in Equation (4) is generated. For our study, we

employ improved Metropolis-Hastings sampler (Grzegorczyk and Husmeier, 2008). For the technical details, the reader is referred to (Grzegorczyk and Husmeier, 2008).

2.3.3. Marginal posterior probabilities of edges

After having generated a sample of DAGs $(G_1, G_2, G_3, ..., G_T)$ from the posterior distribution, the next step is to approximate the marginal posterior probability of all features of interest. The most interesting features are the individual edges. For each possible edge, we would like to approximate the marginal edge posterior probability (often referred to as edge score). Compare Equation (5) with the indicator function *f* indicating the presence (or absence) of a particular edge. Given the DAG sample $(G_1, G_2, G_3, ..., G_T)$ from Equation (4) a consistent estimator for the expression in Equation (4) is:

$$\widehat{P}(f|D) = \sum_{t=1}^{T} f(G_t)$$
(6)

where f(G) = 1 if *G* possesses the edge, and f(G) = 0 otherwise.

Computing the marginal edge posterior probabilities for all possible edges defines a ranking of all edges. To obtain a concrete network prediction, one can impose a threshold and extract all edges whose marginal posterior probability is higher than this threshold, whereas edges with probabilities lower or equal to the threshold are assumed to be absent.

To test for convergence (i.e. to test whether the stationary distribution has been reached), we always run three independent MCMC simulations on each data set. A widely applied convergence diagnostic is to plot the edge scores from two independent runs against each other. If all points in those scatter plots are located near the diagonal, it can be concluded that the chains have produced (approximately) the same results, and thus sufficiently converged. In our study, we always observed sufficient convergence, so that we eventually averaged the edge scores of the three independent simulations. This procedure is briefly described in subsection 3.3.

3. The proposed regional drought indicator: The Seasonally Combinative Regional Drought index (SCRDI)

The main objective of this research is to develop a new drought indicator by incorporating most representative information at the regional level. To achieve this, this section is mainly based on the selection of SDI index and Bayesian network procedures. Details of these two methodologies have already been described in section 2. Before the execution of the proposed procedure, we have defined the following three main point which has substantial importance for accurate inferences related to regional settings.

- 1. **Defining Region:** This step determines the selection of region for drought monitoring. In this step, a particular region is identified for regional drought monitoring. However, this research suggests those regions which are the most influential on the larger part of the country/province and have rich climate characteristics. Accordingly, appropriate selection of region will strengthen accurate and efficient drought mitigation policies at the province or country level.
- 2. **Defining Meteorological station:** After the selection of regions, appropriate selection of meteorological stations/monitoring stations is suggested. As we know that long climatic data has a significant role in model building and statistical inferences. Thus, the meteorological stations which have rich drought monitoring history are suggested.
- 3. Seasonality Indication: Within a particular region, climatology among meteorological stations varied in most parts of the world. For example, a real and temporal variability in precipitation and temperature at various locations in a specified region is a renowned topic (Ma et al., 2018; Asfaw et al., 2018; Ongoma and Chen, 2017; Mondal et al., 2015; Gehne et al., 2016). From previous studies, we



(c)

Fig. 3. At Chilas station subfigure (a) represents frequency and fitted Probability distributions on DAI_i time series data, (b) represents Q-Q plot of the selected 3P Weibull distribution, and (c) represents the time series data of SPTI index.

observed that some regions have a long duration of the cold season (Yang et al., 2013), while some have hot climate throughout the year (Wu et al., 2018; Uvo et al., 1998). In this scenario, it is difficult to define a generalized seasonality index. However, to ease in quantification, this study recommends each month as a season. Although various regions have various level of variability within a specific month, it is a simple, comprehensive and most practiced way. Several climatological and environmental studies are based on a monthly-defined seasonal index (Ayugi et al., 2016; Yang et al., 2013; Zhang et al., 2012, 2006).

After defining the above three points, the stepwise execution of the proposed framework consists of four phases. A detailed description of each phase is described in the subsequent subsections.

3.1. Phase 1. the choice of drought indices and their estimation

This phase consists of the selection of drought indicator. In literature, many authors have provided various drought indicators for the

standardized procedure of drought index. Some of them are available in Svoboda and Fuchs (2016). In section 2, we have described a summary of various SDI indicators and their applications in various regions. Parallel to SDI procedures, recent developments also concentrate on the estimation procedure these drought indices such as parametric and non-parametric based estimation Soláková et al. (2014). Therefore, this phase is very important for accurate regional drought monitoring and its analysis.

The major concern of this phase is to select the climatic parameters and the time scale for the estimation of multi-scalar drought indices. Depending on the nature of climatic, soil type and tropical status, various drought indices required various climatic parameters such as temperature, precipitation, solar radiation, and humidity, etc. Therefore, optimized selection of drought indices and their estimation procedure can significantly contribute inaccurate and reliable drought monitoring. In particular, this step requires a deep knowledge of the following issues:

 Table 3

 Marginal posterior probability matrices of three simulation runs in January.

	Simulations Runs	Runs Meteorological Stations				
		Skardu	Gupis	Gilgit	Chilas	Astore
	Skardu	0	0.2566	0.2624	0.7855	0.3113
	Gupis	0.285	0	0.6982	0.3076	0.3223
Run-1	Gilgit	0.2639	0.6914	0	0.438	0.2392
	Chilas	0.6924	0.2592	0.3791	0	0.4012
	Astore	0.3617	0.3502	0.2503	0.5205	0
ADP		0.4008	0.3893	0.3975	0.5129	0.3185
	Skardu	0	0.2802	0.2534	0.7955	0.3086
	Gupis	0.3028	0	0.7198	0.316	0.307
Run-2	Gilgit	0.2787	0.7219	0	0.4621	0.2539
	Chilas	0.693	0.2608	0.387	0	0.4033
	Astore	0.347	0.3349	0.2681	0.5152	0
ADP		0.4054	0.3994	0.4071	0.5222	0.3182
	Skardu	0	0.2744	0.266	0.8013	0.3265
	Gupis	0.3023	0	0.6961	0.3076	0.317
Run-3	Gilgit	0.2813	0.7072	0	0.4784	0.2608
	Chilas	0.663	0.2608	0.3659	0	0.367
	Astore	0.3691	0.336	0.2671	0.4947	0
ADP		0.4039	0.3946	0.3988	0.5205	0.3178
(AMPP)		0.4033	0.3945	0.4011	0.5185	0.3182

 Table 4

 Marginal posterior probability matrices of three simulation runs in February.

	Simulations Runs	Matrix of Stations				
		Skardu	Gupis	Gilgit	Chilas	Astore
	Skardu	0	0.2513	0.2655	0.8175	0.2686
	Gupis	0.2644	0	0.3717	0.3023	0.4453
Run-1	Gilgit	0.2833	0.3843	0	0.2718	0.2933
	Chilas	0.7649	0.2586	0.2239	0	0.3838
	Astore	0.2849	0.4216	0.2681	0.4311	0
ADP		0.3994	0.329	0.2823	0.4557	0.3478
	Skardu	0	0.2571	0.2381	0.8259	0.2765
	Gupis	0.2618	0	0.3554	0.2912	0.4263
Run-2	Gilgit	0.2607	0.3612	0	0.2912	0.2744
	Chilas	0.7749	0.2676	0.2513	0	0.3832
	Astore	0.3038	0.4095	0.2507	0.4327	0
ADP		0.4004	0.3239	0.2739	0.4603	0.3402
	Skardu	0	0.2523	0.2565	0.8201	0.2802
	Gupis	0.2770	0	0.3433	0.296	0.4390
Run-3	Gilgit	0.2891	0.3475	0	0.2670	0.2886
	Chilas	0.7755	0.2728	0.2381	0	0.3880
	Astore	0.2918	0.4074	0.2634	0.42797	0
ADP		0.4084	0.3201	0.2754	0.4528	0.349
(AMPP)		0.4027	0.3243	0.2772	0.4563	0.3456

- The identification of the nature of gauging station and the accessibility of the time series data on the climatic parameters.
- The appropriate selection of multi-scalar drought indicator (i.e. SPI, SPEI, SPTI) that can be accomplished with the available data.

• Type of drought with their corresponding time scale. In this step, the time scale of multi-scalar drought indices is selected. For example, short time scales are recommended for meteorological Guttman (1998), whereas longer time scales are specified for the monitoring of agricultural and hydrological drought Ma'rufah et al. (2017).

3.2. Phase 2. segregation of time series data by seasonality indexing

This phase describes the temporal formation of SDI data. In the previous section, we have described the seasonal index. Accordingly, this phase suggests separate analysis on the time series data set of SDI of a specific region according to the seasonality index. For the sake of generality, this research assumes and defines seasonal index at a monthly level. Let R_1, R_2, \ldots, R_{12} be the time series data indexed by month, where each month is considered as a season. In the consequent

step, each indexed time series for all the stations would be considered as an independent time series data for further practice.

3.3. Phase 3. configuring BN model

This phase describes and configures Bayesian network models on the seasonally separated time series data of SDI data of multiple meteorological stations.

Consider, the network of meteorological stations $(Y_1, Y_2, Y_3, \dots,$ Y_n ; Y) at the specified region. Here, the purpose to obtain a probabilistic model representing the whole uncertainty about the drought in the network at a specific season/month/time. In section 3, we have described how Bayesian networks is a useful and most powerful probabilistic technique for extracting probabilistic information about interacting variables. The main source of probabilistic information is the marginal posterior probabilities of edge nodes/variable. The marginal posterior probabilities describe dependency/independency structure among variable in a quantitative way. In this context, the marginal posterior probabilities of individual seasonal time series data of SDI index at various meteorological stations are suggested to specify regional dependency features of the most dominant meteorological station. Mathematically, let Y_1 , Y_2 , Y_3 , ..., Y_n , be the seasonal (monthly separated) time series vector data of SDI index at meteorological stations, where a sub-index refers to the station number. And each meteorological station $(Y_1, Y_2, Y_3, ..., Y_n)$ is considered as a node/variable. Here, Y_j can take a value or state y_j , where the value or state is the realization of a variable/node/meteorological stations. Mathematically,

$$P(y_1, y_2, \dots, y_n) = P(Y_1 = y_1, \dots, Y_n = y_n)$$
(7)

The structure of the above equation describes the probability of meteorological stations Y_1 in y_1 , meteorological stations Y_2 in y_2 and so on. Accordingly, Equation (6) is suggested to obtain marginal posterior probabilities of edges.

The consequent subsection consists of the uses of these marginal posterior probabilities in our proposed drought method.

3.4. Phase 4. the choice of meteorological stations

In this phase, a meteorological station at a particular season or month is suggested to identify from probability structure of Bayesian networks results.

Suppose that, a particular region/domain Θ has a well-known set of meteorological stations \underline{Y} . In a single run of Bayesian networks, consider that the marginal posterior distribution of Θ at a particular month $R_i \in \underline{R}$ have the following mathematical form.

3.5. $Y_1 Y_2 \dots Y_n$

Y_1	p_{11}	p_{12}		p_{1n}	
2	p_{21}	p_{22}		p_{2n}	
:	1 :	:	۰.	:	
ľ _n	p_{n1}	p_{n2}		p_{nn}	

where the element in the above matrix is mapped from the results of equation (6). In the above matrix, the lower side of the off-diagonal element describes dependence probability between meteorological stations in Θ at a particular month $R_i \in \underline{R}$. Whereas upper diagonal describes independency features accordingly.

By careful implementation of Bayesian networks, we suggest choosing one meteorological station on which most of the station have maximum Average Marginal Posterior Probability Matrix (AMPP). From the above matrix, the mathematical form of AMPP is represented as follows:

$$AMPP = \max(\bar{p}_1 + \bar{p}_2, ..., \bar{p}_n)$$
 (8)



(b) February

Fig. 4. Scatter plots of the marginal posterior probability estimates observed in January and February data.

where, $\bar{p}_{.1} = \frac{\sum_{i=1}^{n} p_{.i}}{n}$ is the average of individual column entities of the marginal posterior distribution matrix. Then, the choice of MS from Y in all the defined seasonal indicators R is made for those that have maximum associated probability AMPP as defined in 8. To maintain accuracy in result, we suggest taking an average of three independent runs of the BN model. That is, there should be at least three MCMC simulations.

For more comprehension, the step-wise summary of the proposed procedure is as follows.

- 1. The first step is to find JPD by configuring Bayesian network model on the time series data of SDI. In each month, the BN model will produce joint probability matrix (see matrix presented in section 3.4).
- 2. The second step consists of the two more replication of step 1 and 2.

3. The third step computes Average Dependence Probability (ADP) for all the stations. This step is due to three runs of the Bayesian network model. The formula of average dependence probability is defined in equation (9). Consequently, we will concentrate on the grand average. Mathematically,

$$\bar{P}_{i} = \frac{\sum_{i=1}^{3} (\bar{p}_{i1} + \bar{p}_{i2} + \bar{p}_{i3})}{3}$$
(9)

4. In the fourth step, monthly time series data of that station which have the highest ADP probabilities are selected for further data mining.

After the estimation of the standardized drought indices for each station, the data is divided into 12 sets. In each set, we have time series variables associating from five selected stations. As the time series range

Journal of Environmental Management 276 (2020) 111296



(a) Graphical representation of marginal posterior probability estimates of first simulation run



(b) Graphical representation of marginal posterior probability estimates of second simulation run





(c) Graphical representation of marginal posterior probability estimates of third simulation run

(d) Scatter plots of the marginal posterior probability estimates

Fig. 5. Observed quality of three structure MCMC simulation runs in March, (a) Graphical representation of marginal posterior probability estimates of first simulation ru, (b) Graphical representation of marginal posterior probability estimates of second simulation run, (c) Graphical representation of marginal posterior probability estimates of third simulation run, (d) Scatter plots of the marginal posterior probability estimates.

from 1971 to 2017, therefore each variable contains 47 observations. For all the 12 data sets, three independent structure MCMC simulations runs are made. In structure MCMC simulation setting, a total 200,000 iterations with 100 step save and burn-in are configured subjectively (Grzegorczyk, 2010).

The evaluation and accuracy of these probabilities are assessed by performing more than three structure MCMC simulations. These simulations are either have different seeded runs or independent and identical. This practice is followed by Grzegorczyk and Husmeier (2008). To do this, scatter plots of the marginal posterior probability of edges

consisting of three pairwise simulations runs are presented in result and discussion section.

4. Results

4.1. Estimation of SPTI index

This section describes and presents some numerical and graphical results related to the estimation of SPTI-1 index for all stations. Table 2 shows the list of probability distributions corresponding with the

Table 5

AMPP values of three simulation runs for all the stations.

Month	Stations	AMPP
Jan	Chilas	0.5185
Feb	Chilas	0.4563
Mar	Skardu	0.4933
Apr	Skardu	0.4572
May	Skardu	0.4966
Jun	Chilas	0.5036
Jul	Chilas	0.4869
Aug	Skardu	0.4407
Sep	Skardu	0.5451
Oct	Skardu	0.5094
Nov	Skardu	0.4897
Dec	Skardu	0.6293

Bayesian Information Criterion (BIC) values. We observed that three parameters (3P) Weibull distribution have the lowest BIC values (-480.82, -379.153, 197.265) for Astor, Chilas and Gilgit respectively. In addition, Fig. 3a shows how 3P probability distribution is well fitted for computing SPTI indicator. Further, Fig. 3b demonstrates the scatter plot of theoretical and empirical density. Accordingly, 4P Weibull with BIC values -427.46 and Johnson Sb distribution with BIC values -871.55 are found to be well-fitted probability distributions for Gupis and Skardu respectively (see Table 2).

After the standardization step, temporal behaviors of SPTI index are observed (see Fig. 3c). From the temporal data, significant discrepancies have been observed. From a very short distanced meteorological stations, such discrepancy has serious data mining problems. That is, it reflects many challenges for future drought forecasting, drought monitoring and data dissemination at a regional scale.

4.2. Implications of Bayesian networks

Tables 3 and 4 comprises on the marginal posterior probabilities of three simulation runs for January and February data sets, respectively. In term of identical convergence of MCMC results, the quality of simulation runs is presented in Fig. 4. These plots are used to evaluate the convergence of marginal posterior probabilities in each simulation. The pairwise comparison shows that the MCMC simulation runs have achieved a sufficient degree of convergence in both of these months. In addition, there are no significant deviations within the marginal posterior probability matrices. Hence, the results of MCMC runs are consistent with each other.

To check the most representative station, the marginal posterior probabilities of each station is averaged according to the proposed setup. We observe that Chilas have maximum values of Average Marginal Posterior Probability Matrix (AMPP). This means that irrespective to other stations, Chilas is the most representative station in January and February. By the rationale of our proposed design, we have concluded that the temporal data of Chilas station is the most representative in January and February under regional settings.

Similar to January and February data sets, structure MCMC practice is made of all other months separately. Fig. 5 exhibits the results of structure MCMC simulation runs. We have presented the estimate marginal posterior probabilities from three independent simulation runs for Jan and Feb (See table). However, to save the volume of pages and ease in understanding, we skipped the presentation of numerical values of marginal posterior probabilities for all the month. The graphical results are archived in the author's gallery.

In the next step, maximum AMPP is identified using equation (8). Table 5 presents the average marginal posterior probability for all the month. We have found that the Chilas is the most representative meteorological station in January, February, June, and July. While Skardu is the most representative in all other months.

5. Discussion

In this article, we first computed the SPTI index under the parametric approach of standardization. In the computation procedure, we have followed Stagge et al. (2015) procedure of varying probability distributions. After the estimation of SPTI index for all the stations, time series data at each station is segregated according to monthly defined season. We have described the segregation process of time series data in methodology. For each data set, Bayesian network models are employed separately under structure MCMC. Our experimental results consist of three independent simulation runs. Where, each simulation run is made on monthly based segregated time series data of SPTI index for all the station combatively. For validation of Bayesian networks, the scatter plots of the posterior probabilities for all the simulation runs are investigated. In addition, marginal posterior probability distributions for all the month are shown in the form of tables and graphs.

After discovering dependency structure using maximum AMPP, the monthly time series values of respective representative stations are desegregated and merged into a single series. We have named this single series of SPTI index as the SCRDI. Fig. 6 shows the temporal plot of SCRDI index. This single time series data of SCRDI have characteristics to describe and characterize drought situation at the regional level. Hence to report regional statistics such as forecasting, model fitting, and dissemination of meteorological event (i.e drought or wet), the proposed index has a strong rationale of its candidacy.



Fig. 6. Temporal representation of regional drought indices: the SCRDI.

The main advantages of the proposed methods are 1) the proposed method provides a sound procedure for composite assessment of regional drought (Chen et al., 2020), 2) it provides a solid way to combined information of drought coming from various stations.

The limitations of the proposed approach is the choice of MCMC simulation setting and BNs. In literature, there are several BNs and simulation setting. However, the optimal selection of the MCMC simulation setting and BNs is required for accurate and comparable results.

6. Conclusions

This paper provides a systematic way to combine SDI time series data of various meteorological stations located in a certain region under Bayesian network theory. Here, a new combinative procedure of defining a regional drought indicator: the SCRDI have been proposed. Practical implementation for establishing SCRDI time series is made for five meteorological stations located in the Northern Area of Pakistan. In this research, we have used seasonally segregated time series data of SPTI-1 index in Bayesian network model under structure MCMC setup. To preserve more accuracy, our results and inferences are based on three independent structure MCMC based simulation runs. Under the rationale of exploring dependency/independency structure among variables by Bayesian network theory, the most representative and influential stations are identified using maximum AMPP of three simulations. Among five meteorological stations, two stations, namely Chilas and Skardu are found to be the most dominant stations. However, the behavior of dominance of these stations varies with months. After these results, the original time series data of SPTI index is desegregated accordingly to the proposed rationale of SCRDI. Here, we conclude that the SCRDI is a regionally representative indicator for characterizing meteorological drought at a regional scale. In addition, being a single regional representative, SCRDI is a good choice for efficient model searching and fitness. That is, SCRDI leads lessen in the calculation, save time and resources in the forecasting of future drought at a regional scale.

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

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group no RG-1437-027.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jenvman.2020.111296.

Astor (35.3570° N, 74.8624° E), Skardu (35.3247° N, 75.5510° E), Gupis (36.2274° N, 73.4421° E), Gilgit (35.8819° N, 74.4643° E), and Chilas (35.4222° N, 74.0946° E).

References

- Ali, Z., Hussain, I., Faisal, M., Nazir, H.M., Abd-el Moemen, M., Hussain, T., Shamsuddin, S., 2017. A novel multi-scalar drought index for monitoring drought: the standardized precipitation temperature index. Water Resour. Manag. 31 (15), 4957–4969
- Ali, Z., Hussain, I., Faisal, M., Almanjahie, I.M., Ismail, M., Ahmad, M., Ahmad, I., 2018. A new weighting scheme in weighted markov model for predicting the probability of drought episodes. Adv. Meteorol. 1–10. https://doi.org/10.1155/2018/8954656, 8954656.

- Ali, Z., Hussain, I., Faisal, M., Elashkar, E.E., Gani, S., Shehzad, M.A., 2019a. Selection of appropriate time scale with Boruta algorithm for regional drought monitoring using multi-scaler drought index. Tellus Dyn. Meteorol. Oceanogr. 71 (1), 1–16.
- Ali, Z., Hussain, I., Faisal, M., Shoukry, A.M., Gani, S., Ahmad, I., 2019b. A framework to identify homogeneous drought characterization regions. Theor. Appl. Climatol. 1–12.
- Arsenault, R., Brissette, F., 2014. Determining the optimal spatial distribution of weather station networks for hydrological modeling purposes using RCM datasets: an experimental approach. J. Hydrometeorol. 15 (1), 517–526.
- Asfaw, A., Simane, B., Hassen, A., Bantider, A., 2018. Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: a case study in Woleka sub-basin. Weather and climate extremes 19, 29–41.
- Ayugi, B.O., Wen, W., Chepkemoi, D., 2016. Analysis of spatial and temporal patterns of rainfall variations over Kenya. Studies 6 (11).
- Bayat, B., Nasseri, M., Zahraie, B., 2015. Identification of long-term annual pattern of meteorological drought based on spatiotemporal methods: evaluation of different geostatistical approaches. Nat. Hazards 76 (1), 515–541.
- Bazrafshan, J., Hejabi, S., Rahimi, J., 2014. Drought monitoring using the multivariate standardized precipitation index (MSPI). Water Resour. Manag. 28 (4), 1045–1060.
- Bertone, E., Sahin, O., Stewart, R.A., Zou, P.X., Alam, M., Hampson, K., Blair, E., 2018. Role of financial mechanisms for accelerating the rate of water and energy efficiency retrofits in Australian public buildings: hybrid Bayesian Network and System Dynamics modelling approach. Appl. Energy 210, 409–419.
- Borsuk, M.E., Stow, C.A., Reckhow, K.H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecol. Model. 173 (2–3), 219–239.
- Cancelliere, A., Salas, J.D., 2004. Drought length properties for periodic-stochastic hydrologic data. Water Resour. Res. 40 (2).
- Chen, H., Fan, L., Wu, W., Liu, H.B., 2017. Comparison of spatial interpolation methods for soil moisture and its application for monitoring drought. Environ. Monit. Assess. 189 (10), 525.
- Chen, S., Zhong, W., Pan, S., Xie, Q., Kim, T.W., 2020. Comprehensive drought assessment using a modified composite drought index: a case study in hubei province, China. Water 12 (2), 462.
- Darwiche, A., 2009. Modeling and Reasoning with Bayesian Networks. Cambridge university press.
- De Martonne, E., 1926. Une Nouvelle Fonction Climatologique: L'indice D'aridité. La Meteorologie, pp. 449–458.
- Dezman, L.E., Shafer, B.A., Simpson, H.D., Danielson, J.A., 1983. Development of a surface water supply index- A drought severity indicator for Colorado. 1983. In: International Symposium on Hydrometeorology June 13-17, 1982. American Water Resources Association, Denver, Colorado, pp. 337–341, 3 fig, 3 ref.
- Drogue, G., Humbert, J., Deraisme, J., Mahr, N., Freslon, N., 2002.
 A statistical-topographic model using an omnidirectional parameterization of the relief for mapping orographic rainfall. Int. J. Climatol. 22 (5), 599–613.
 Easterling, D.R., Wallis, T.W., Lawrimore, J.H., Heim Jr., R.R., 2007. Effects of
- temperature and precipitation trends on US drought. Geophys. Res. Lett. 34 (20). Einfalt, T., Denoeux, T., Jacquet, G., 1990. A radar rainfall forecasting method designed
- for hydrological purposes. J. Hydrol. 114 (3–4), 229–244.
- Erhardt, T.M., Czado, C., 2018. Standardized drought indices: a novel univariate and multivariate approach. J. Roy. Stat. Soc.: Series C (Applied Statistics) 67 (3), 643–664.
- Faiz, M.A., Liu, D., Fu, Q., Wrzesiński, D., Baig, F., Nabi, G., et al., 2018. Extreme precipitation and drought monitoring in northeastern China using general circulation models and pan evaporation-based drought indices. Clim. Res. 74 (3), 231–250.
- Friedman, N., Koller, D., 2000. June). Being Bayesian about network structure. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., pp. 201–210
- Gehne, M., Hamill, T.M., Kiladis, G.N., Trenberth, K.E., 2016. Comparison of global precipitation estimates across a range of temporal and spatial scales. J. Clim. 29 (21), 7773–7795.
- Golian, S., Mazdiyasni, O., AghaKouchak, A., 2015. Trends in meteorological and agricultural droughts in Iran. Theor. Appl. Climatol. 119 (3–4), 679–688.
- Grzegorczyk, M., 2010. An introduction to Gaussian Bayesian networks. In: Systems Biology in Drug Discovery and Development. Humana Press, Totowa, NJ, pp. 121–147.
- Grzegorczyk, M., Husmeier, D., 2008. Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. Mach. Learn. 71 (2–3), 265.
- Guttman, N.B., 1998. Comparing the palmer drought index and the standardized precipitation index 1. JAWRA Journal of the American Water Resources Association 34 (1), 113–121.
- Guttman, N.B., 1999. Accepting the standardized precipitation index: a calculation algorithm 1. JAWRA Journal of the American Water Resources Association 35 (2), 311–322.
- Hao, Z., AghaKouchak, A., 2013. Multivariate standardized drought index: a parametric multi-index model. Adv. Water Resour. 57, 12–18.
- Hao, Z., AghaKouchak, A., 2014. A nonparametric multivariate multi-index drought monitoring framework. J. Hydrometeorol. 15 (1), 89–101.
- Hiscock, K.M., 2009. Hydrogeology: Principles and Practice. John Wiley & Sons.
- Hong, N.T., Truc, P.T.T., Liem, N.D., Loi, N.K., 2016. Optimal selection of number and location of meteo-hydrological monitoring networks on vu gia-thu bon river basin using GIS. Int. J. Adv. Sci. Eng. Inf. Technol. 6 (3), 324–328.
- Hu, Q., Willson, G.D., 2000. Effects of temperature anomalies on the palmer drought severity index in the central United States. Int. J. Climatol. 20 (15), 1899–1911.

Z. Ali et al.

Keetch, J.J., Byram, G.M., 1968. A Drought Index for Forest Fire Control. In: Res. Pap. SE-38, vol. 35. US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, NC, p. 38.

- Kochenderfer, M.J., 2015. Decision Making under Uncertainty: Theory and Application. MIT press.
- Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques. MIT press.
- Lark, R.M., 2016. Multi-objective optimization of spatial sampling. Spatial Statistics 18, 412–430.
- Lee, J., Henning, R., Cherniack, M., 2019. Correction workers' burnout and outcomes: a bayesian network approach. Int. J. Environ. Res. Publ. Health 16 (2), 282. Liu, Z., Liu, Y., Shan, H., Cai, B., Huang, Q., 2015. A fault diagnosis methodology for gear
- Lu, Z., Lu, T., Shan, H., Cai, S., ruang, Q., 2015. A fault diagnosis memodology for gear pump based on EEMD and Bayesian network. PloS One 10 (5), e0125703.
 Ma, L., Xia, H., Sun, J., Wang, H., Feng, G., Qin, F., 2018. Spatial-temporal variability of
- Ma, L., Xia, H., Sun, J., Wang, H., Feng, G., Qin, F., 2018. Spanal-temporal variability of hydrothermal climate conditions in the yellow river basin from 1957 to 2015. Atmosphere 9 (11), 433.
- Madigan, D., York, J., Allard, D., 1995. Bayesian graphical models for discrete data. International Statistical Review/Revue Internationale de Statistique 215–232.
- Marquina, A. (Ed.), 2010. Global Warming and Climate Change: Prospects and Policies in Asia and Europe. Springer.
- Mathbout, S., Lopez-Bustins, J.A., Martin-Vide, J., Bech, J., Rodrigo, F.S., 2018. Spatial and temporal analysis of drought variability at several time scales in Syria during 1961–2012. Atmos. Res. 200, 153–168.
- Ma'rufah, U., Hidayat, R., Prasasti, I., 2017. Analysis of relationship between meteorological and agricultural drought using standardized precipitation index and vegetation health index. In: IOP Conference Series: Earth and Environmental Science, vol. 54. IOP Publishing, 012008.
- McKee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference on Applied Climatology, vol. 17. American Meteorological Society, Boston, MA, pp. 179–183.
- Mendes, E., Mosley, N., 2008. Bayesian network models for web effort prediction: a comparative study. IEEE Trans. Software Eng. 34 (6), 723–737.
- Meyer, S.J., Hubbard, K.G., Wilhite, D.A., 1993. A crop-specific drought index for corn: I. Model development and validation. Agron. J. 85 (2), 388–395.
- Miniscloux, F., Creutin, J.D., Anquetin, S., 2001. Geostatistical analysis of orographic rainbands. J. Appl. Meteorol. 40 (11), 1835–1854.
- Moglia, M., Alexander, K.S., Thephavanh, M., Thammavong, P., Sodahak, V., Khounsy, B., Vorlasan, S., Larson, S., Connell, J., Case, P., 2018. A Bayesian Network model to explore practice change by smallholder rice farmers in Lao PDR. Agric. Syst. 164, 84–94.
- Mondal, A., Khare, D., Kundu, S., 2015. Spatial and temporal analysis of rainfall and temperature trend of India. Theor. Appl. Climatol. 122 (1–2), 143–158.
- Neapolitan, R.E., 2012. Probabilistic Reasoning in Expert Systems: Theory and Algorithms. CreateSpace Independent Publishing Platform.
- Ongoma, V., Chen, H., 2017. Temporal and spatial variability of temperature and precipitation over East Africa from 1951 to 2010. Meteorol. Atmos. Phys. 129 (2), 131–144.
- Pearl, J., 2014. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Elsevier.
- Salma, S., Shah, M.A., Rehman, S., 2012. Rainfall trends in different climate zones of Pakistan. Pakistan Journal of Meteorology 9 (17).
- Scarsoglio, S., Laio, F., Ridolfi, L., 2013. Climate dynamics: a network-based approach for the analysis of global precipitation. PloS One 8 (8), e71129.
- Sebyala, A.A., Olukemi, T., Sacks, L., Sacks, D.L., 2002. Active platform security through intrusion detection using naive bayesian network for anomaly detection. In: London Communications Symposium, pp. 1–5.

- Shen, X., Wu, X., Xie, X., Ma, Z., Yang, M., 2017. Spatiotemporal analysis of drought characteristics in Song-Liao river basin in China. Adv. Meteorol. 2017, 3484363 https://doi.org/10.1155/2017/3484363, 2017.
- Siddiqui, S., Safi, M.W.A., 2019. Assessing the socio-economic and environmental impacts of 2014 drought in district Tharparkar, Sindh-Pakistan. International Journal of Economic and Environmental Geology 8–15.
- Soláková, T., De Michele, C., Vezzoli, R., 2014. Comparison between parametric and nonparametric approaches for the calculation of two drought indices: SPI and SSI. J. Hydrol. Eng. 19 (9), 04014010.
- Spiess, A.N., 2014. Propagate: propagation of uncertainty. R package version 1, 0-4.
- Stagge, J.H., Tallaksen, L.M., Gudmundsson, L., Van Loon, A.F., Stahl, K., 2015. Candidate distributions for climatological drought indices (SPI and SPEI). Int. J. Climatol. 35 (13), 4027–4040.
- Svoboda, M., Fuchs, B., 2016. Handbook of drought indicators and indices.
- Tigkas, D., Vangelis, H., Tsakiris, G., 2018. Drought characterisation based on an agriculture-oriented standardised precipitation index. Theor. Appl. Climatol. 1–13.
- Tsakiris, G., 2017. Drought risk assessment and management. Water Resour. Manag. 31 (10), 3083–3095.
 Tsakiris, G., Pangalou, D., Vangelis, H., 2007. Regional drought assessment based on the
- Isakiris, G., Pangalou, D., Vangelis, H., 2007. Regional drought assessment based on the reconnaissance drought index (RDI). Water Resour. Manag. 21 (5), 821–833.
- Turral, H., Burke, J., Faurès, J.M., 2011. Climate Change, Water and Food Security (No. 36). Food and Agriculture Organization of the United Nations (FAO).
- Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecol. Model. 203 (3–4), 312–318.
- Uvo, C.B., Repelli, C.A., Zebiak, S.E., Kushnir, Y., 1998. The relationships between tropical Pacific and Atlantic SST and northeast Brazil monthly precipitation. J. Clim. 11 (4), 551–562.
- Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23 (7), 1696–1718.
- Vicente-Serrano, S.M., Miralles, D.G., Domínguez-Castro, F., Azorin-Molina, C., El Kenawy, A., McVicar, T.R., Peña-Gallardo, M., 2018. Global assessment of the standardized evapotranspiration deficit index (SEDI) for drought analysis and monitoring. J. Clim. 31 (14), 5371–5393.
- Watthayu, W., Peng, Y., 2004. A Bayesian network based framework for multi-criteria decision making. August. In: Proceedings of the 17th International Conference on Multiple Criteria Decision Analysis.
- Wayne, C.P., 1965. Meteorological Drought. US weather bureau research paper, p. 58.
- Wong, W.K., Moore, A.W., Cooper, G.F., Wagner, M.M., 2003. Bayesian network anomaly pattern detection for disease outbreaks. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 808–815.
- Wu, Z., Boke-Olén, N., Fensholt, R., Ardö, J., Eklundh, L., Lehsten, V., 2018. Effect of climate dataset selection on simulations of terrestrial GPP: highest uncertainty for tropical regions. PloS One 13 (6), e0199383.
- Yang, Y.G., Hu, J.F., Xiao, H.L., Zou, S.B., Yin, Z.L., 2013. Spatial and temporal variations of hydrological characteristic on the landscape zone scale in alpine cold region. Huan jing ke xue= Huanjing kexue 34 (10), 3797–3803.
- Zarch, M.A.A., Sivakumar, B., Sharma, A., 2015. Droughts in a warming climate: a global assessment of Standardized precipitation index (SPI) and Reconnaissance drought index (RDI). J. Hydrol. 526, 183–195.
- Zhang, Q., Kang, S., Yan, Y., 2006. Characteristics of spatial and temporal variations of monthly mean surface air temperature over Qinghai-Tibet Plateau. Chin. Geogr. Sci. 16 (4), 351–358.
- Zhang, Z., Xu, C.Y., El-Tahir, M.E.H., Cao, J., Singh, V.P., 2012. Spatial and temporal variation of precipitation in Sudan and their possible causes during 1948–2005. Stoch. Environ. Res. Risk Assess. 26 (3), 429–441.
- Zhu, Z., Peng, B., Xiong, C., Zhang, L., 2016. Short-term traffic flow prediction with linear conditional Gaussian Bayesian network. J. Adv. Transport. 50 (6), 1111–1123.