On the Bayesian network based data mining framework for the choice of appropriate time scale for regional analysis of drought Hazard

Qamar, Sadia; Khalique, Abdul; Grzegorczyk, Marco Andreas

Published in:
Theoretical and Applied Climatology

DOI:
10.1007/s00704-021-03530-2

Theoretical and Applied Climatology (2021) 143:1677–1695
https://doi.org/10.1007/s00704-021-03530-2

ORIGINAL PAPER

On the Bayesian network based data mining framework for the choice of appropriate time scale for regional analysis of drought Hazard

Sadia Qamar[1] · Abdul Khalique[1] · Marco Andreas Grzegorczyk[2]

Received: 15 July 2020 / Accepted: 5 January 2021 / Published online: 16 January 2021

Abstract
Data mining has a significant role in hyrdrologic research. Among several methods of data mining, Bayesian network theory has great importance and wide applications as well. The drought indices are very useful tools for drought monitoring and forecasting. However, the multi-scaling nature of standardized type drought indices creates several problems in data analysis and reanalysis at regional level. This paper presents a novel framework of data mining for hydrological research—the Bayesian Integrated Regional Drought Time Scale (BIRDts). The mechanism of BIRDts gives effective and sufficient time scales by considering dependency/interdependency probabilities from Bayesian network algorithm. The resultant time scales are proposed for further investigation and research related to the hydrological process. Application of the proposed method consists of 46 meteorological stations of Pakistan. In this research, we have employed Standardized Precipitation Temperature Index (SPTI) drought index for 1-, 3-, 6-, 9-, 12-, 24-, and ()month time scales. Outcomes associated with this research show that the proposed method has rationale to aggregate time scales at regional level by configuring marginal posterior probability as weights in the selection process of effective drought time scales.

Keywords Data mining · Drought · Bayesian network · Standardized Precipitation Temperature Index (SPTI) · Time scales

1 Introduction

Due to advancement in technology, the growing sources of information have created large and complex data in several disciplines. One aspect of handling big and complex data is data mining. Data mining is a process that examines large preexisting databases to generate new information (Han et al., 2012). In several fields of research, data mining helps to understand the complex nature of data (Babovic, 2005). In hydrological and geo-science research, data mining has a crucial role. In the past research, several authors have worked on various data mining techniques for drought modeling and monitoring (Tadesse et al., 2004, 2009; Farokhnia et al., 2019).

To handle large and complex data, machine learning and advanced methods of statistics play very important role, for example, the application of Bayesian networks theory in various fields (Lee et al., 2019; Bertone et al., 2018; Moglia et al., 2018; and Liu et al., 2015). Enhancements in computational capabilities, innovations in observation and measurement devices open new and fascinating scenarios for the application of data-driven modeling techniques. Consequently, increased availability of hydrological data of different time, spatial scales, and data mining approaches is helpful for discovery of new knowledge. Under the design of Bayesian network, the exploitation of such a great bulk of new information can contribute greatly to enhance the robustness of hydrological models. Hence, accuracy and efficiency in temporal monitoring and forecasting can be gained in a more comprehensive way.

✉ Sadia Qamar
diyaqau@gmail.com

Abdul Khalique
drkhalique@ncbae.edu.pk

Marco Andreas Grzegorczyk
m.a.grzegorczyk@rug.nl

[1] Department of Statistics, National College of Business Administration and Economics, Lahore, Pakistan

[2] Johann Bernoulli Institute, Groningen University, Groningen, The Netherlands

In hydrology and water science, characterization of drought and its monitoring are the main challenges. Emerging issue in hydrology and related sciences is drought monitoring. Hydrologic droughts are the extension of meteorological and agricultural drought. It is caused due to imbalance between precipitation and surface water (Dracup et al., 1980). Since drought is a natural hazard, it is classified by various climatological and hydrological parameters (Mishra and Singh 2010). Drought occurs due to absence of rainfall or low rainfall in a region for a long time period, and it happens because of various reasons including deforestation, global warming, and many other human activities. The effect of this climatic condition on the environment as well as the living beings is very disastrous. Drought is the natural disaster which occurs virtually in all geographical areas (Hao et al., 2018). Drought is caused by multiple climatic factors that have different characteristics in their spatio-temporal data. Consequently, understanding drought hazards is a more difficult task than other natural hazards (Wilhite et al., 2014; Kiem et al., 2016). In addition, drought has recurrent features whose characteristics differ from one climate region to another. Usually, drought has four main classes named as meteorological drought, hydrological drought, agricultural drought, and socioeconomic drought (Anderson et al., 2011; Wilhite and Glantz 1985).

To mitigate the bad impact of drought, more sound and comprehensive procedures are required for its continuous monitoring, prediction, and temporal and spatial analysis of drought (Ali et al., 2019b, 2018). In the past research, many authors have developed several methods for effective drought mitigation policies (Kogan, 2000; Hayes et al., 2004). Some of them are Palmer Drought Severity Index (PDSI) (Palmer 1965), Standardized Precipitation Index (SPI) (McKee et al., 1993), Effective Drought Index (EDI) (Byun and Wilhite 1999), Aggregate Drought Index (ADI) (Keyantash and Dracup 2004), Reconnaissance Drought Index (RDI) (Tsakiris and Vangelis 2005), the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010), Standardized Precipitation Temperature Index (SPTI) (Ali et al., 2017), etc. A comprehensive list along with data requirements of different available drought indices is available (Svoboda and Fuchs 2016).

One of the major aspects of drought monitoring is the multi-scaling characteristics of drought monitoring. For example, some authors have reported that short-term time scales (i.e., 1-, 3-, or 6-month) are useful for metrological and agricultural drought. On the other hand, higher time scales are providing information related to hydrological drought. However, the multi-scaling nature of drought indices reduces efficiency and accuracy in data analysis and reanalysis (Bazrafshan et al., 2014). Further, the interpretations of various drought events in various time scales are cumbersome for decision-makers and drought practitioners at regional level (Ali et al., 2019).

Therefore, the fundamental objective in hydrological modeling is to resolve the multi-scaling problem of drought indices through some state-of-the-art procedures. In previous research, different authors have provided a range of statistical approaches for multi-scaling problems. For example, Bazrafshan et al. (2014) have provided Multivariate Standardized Precipitation Index (MSPI) by applying principal component analysis (PCA) on various time scales of SPI. In a recent research, Ali et al. (2019) have proposed Boruta algorithm of machine learning for the selection of important time scales in SPTI index.

This research develops a comprehensive data mining framework under Bayesian network theory for resolving the multi-scaling problem of multi-scalar drought indices. The proposed framework is mainly based on three Monte Carlo Markov Chain (MCMC) based simulation runs of Bayesian network theory. The application of the proposed framework is based on the novel multi-scalar drought index called SPTI index. In this research, meteorological stations located in various climatological clusters and regional administrative boundaries of Pakistan have been considered.

2 Material and methods

2.1 Data and study area

In this research, we have acquired time series data of precipitation and temperature. Therefore, the data is collected from Pakistan Meteorological department through Karachi Data Processing Center (KDPC). Detailed descriptions on the source and quality data can be found in Ali et al. (2019a, 2019b).

In the computational section, the research includes 46 meteorological stations of Pakistan (see Fig. 1). Those stations which have poor quality or unavailability issues of time series data are neglected. The division of the selected is further made into five regions (i.e., Punjab, Azad Jamu and Kashmir (AJK), Sindh, Baluchistan and Khyber Pakhtunkhwa (KPK)). The division is based on the provincial/administrative boundaries. For detailed description on the computational procedure, one meteorological station is chosen randomly (without replacement) from every region. So, the selected five meteorological stations from above five characterized regions are Sargodha (32° 03′, 72° 40′), Muzaffarabad (72° 40′, 73° 29′), Hyderabad (25° 23′, 68° 25′), Quetta (30°11’, 66°57’), and Gupis (36° 10′, 73° 24′), respectively.

2.2 Standardized drought indices (SDI)

The enhancement of a drought index is conceptually based upon multiple factors. Around the world, several drought indices have been developed and used over the years. However, SDI is considered more suitable for drought monitoring (Erhardt and Czado, 2015). Various authors have proposed various indicators for obtaining SDI type drought index. For

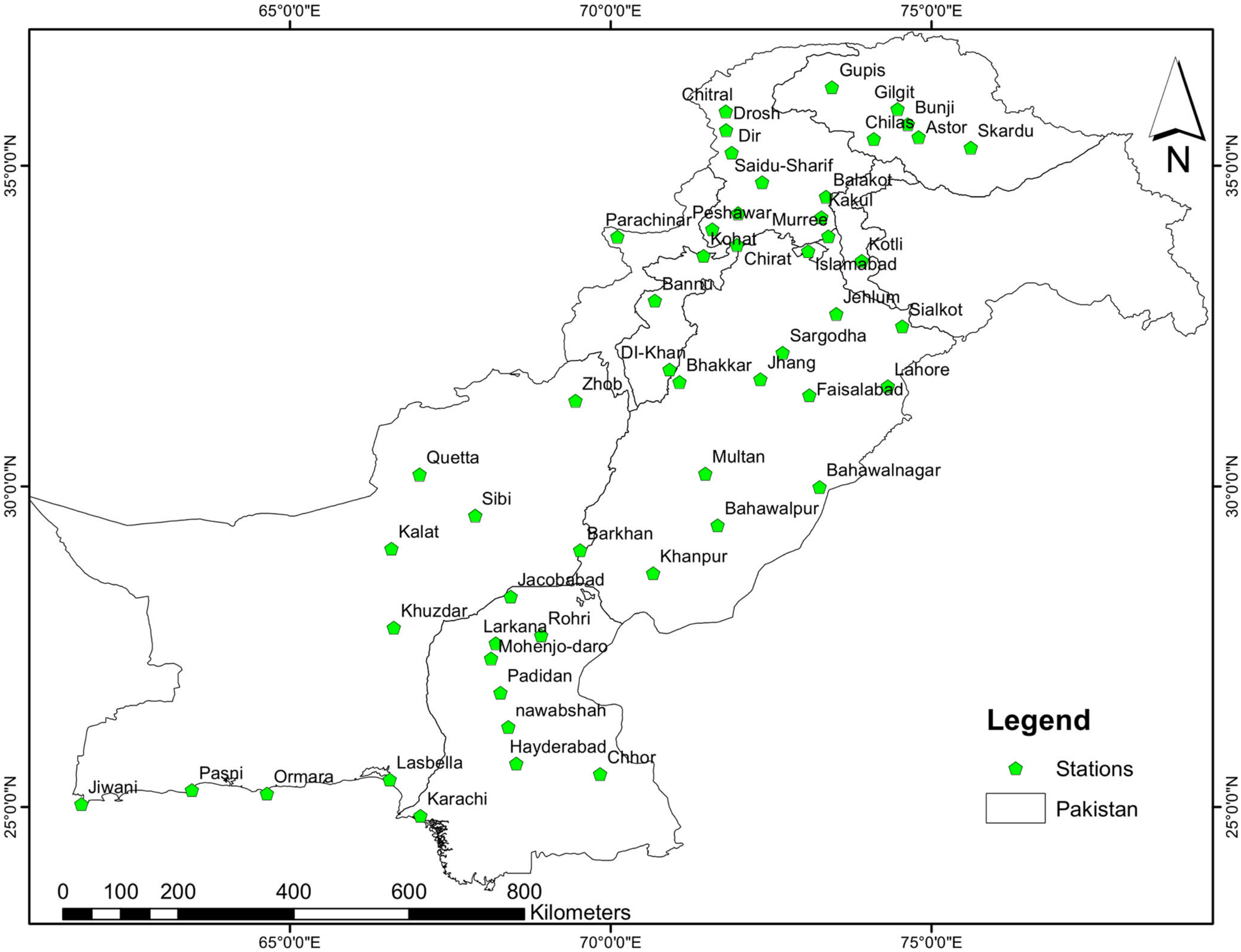


Fig. 1 Geographical distribution of the selected meteorological stations (Ali et al., 2019)

example, McKee et al. (1993) have proposed the Standardized Precipitation Index (SPI) as a meteorological drought monitoring and analysis. SPI can be calculated for 1-, 3-, 6-, 12-, 24-, and 48-month time scales. The SPI is recommended by the World Meteorological Organization (WMO) as a standard drought monitoring index. The SPI is simple and flexible that is why it is distinctive than other indices. McKee et al. (1993) used Gamma distribution to calculate SPI and Guttman (1998, 1999) used Pearson III distribution to calculate SPI. One drawback of SPI is that it uses only a single climatic variable.

Vicente-Serrano et al. (2010) have proposed another SDI index named as the Standardized Precipitation Evaporation Index (SPEI). SPEI is estimated by taking the difference between the Precipitation (P) and the estimated amount of Potential Evapotranspiration (PET). One advantage of SPEI over SPI is that it accounts the effect of temperature by using estimated potential evaporation in the standardization procedure. The main disadvantage of SPEI is that it has an estimation problem in low and high temperature regions (Ali et al., 2017).

In the recent years (Ali et al., 2017), proposed Standardized Precipitation Temperature Index (SPTI) for drought monitoring which utilizes the regional temperature. SPTI drought index can account the role of temperature and precipitation over different time scales.

There are numerous other paradigms and procedures for multivariate data modeling of drought indices. The Multivariate Standardized Drought Index (MSDI) is the combination of SPI and the Standardized Soil Moisture Index (SSI) for drought characterization probabilistically (Hao and AghaKouchak, 2013). Bazrafshan et al. (2014) proposed a new multivariate drought index—the Multivariate Standardized Precipitation Index (MSPI) which is used to resolve multi-scaling issues in univariate SDI method.

Many indices have been developed to monitor drought impact. To quantify and classify droughts' temporal trends, different drought indices play a vital role. Many drought indices are reviewed along with their pros and cons under different conditions in Mishra and Singh (2010, 2011). However, SPTI (Ali et al., 2017) is one of the indices with the key feature to detect and monitor drought conditions with different time scales. SPTI is a multivariate multi-scaler drought index. That is, it uses the De Martonne Aridity Index (DMAI) (De Martonne 1926) which is based on precipitation and temperature data. Unlike SPEI, SPTI has no overestimation and underestimation problem (Ali et al., 2017).

This research is based on the SPTI index, where the time series data of DMAI is standardized under the standardization procedure provided in Abramowitz and Stegun (1965). The detained description on SPTI index can be found in Ali et al. (2017).

2.3 Bayesian network theory

Bayesian network (BN) algorithms are among the most practical approaches to certain types of learning problems, and their results are comparable to the performance of other classifiers (Cheng and Greiner, 1999; Mayfield et al., 2020). BNs can be applied to many areas including estimation, classification, recognition, inference, and prediction (Grzegorczyk, 2010). There are many advantages of BNs as it pools information from multiple drought indices and comes up with a better estimate for drought severity as well as it offers probabilistic estimates for droughts. It decreases the uncertainty of the individual drought indices. Since BNs uniquely defines a joint probability model, inference, drawing conclusions based on observations has solid rules of probability calculus. This implies that there is mathematical consistency and correctness.

Let L be the time series data set on n variables. Here, the problem is to find the relationship among the variables in L. In BNs learning, the posterior probabilities of network graph B are computed by following equation.

$$P(B/L) = \frac{P(L/B)P(B)}{P(L)} \tag{1}$$

Here, the computation of posterior probability depends on the average of over all possible networks of any hypothesis of interest. Consider f be the structural function depicted as an indicator function. Then, the estimated posterior probabilities for certain feature f are estimated as follows:

$$P(f/L) = \sum_{B} P(B/L)f(B) \tag{2}$$

where B represents a model and $f(B)$ is one if there is a feature in B; otherwise, it is zero. After having sample of DAGs (B_1, B_2, B_3, $\ldots B_r$) from posterior distribution, the next step is to compute Marginal Posterior Probability (MPP) of each edge. From the context of sampling graph via MCMC simulations, the following equation gives the estimator of marginal posterior probabilities of edges.

$$\widehat{P}\,(f/L) = \frac{1}{T}\sum_{t=1}^{T} I_f(B_t) \tag{3}$$

where $I_f(.)$ is used as a binary indicator function for all the graph. Here, if the existence of edge relation will be 1 and zero otherwise. From BNs approach, the main aim is to extract the information on feature dependencies from marginal posterior probabilities.

3 The proposed framework—the Bayesian Integrated Regional Drought Time Scale (BIRDts)

This section describes stepwise execution of the proposed framework (see Fig. 2). The BIRDts have four phases. The first phase is related to the choices of selection of regions, stations, and drought indices, while the second phase describes the estimation of drought indices. The third phase consists of the integration and implementation of Bayesian network models. The fourth phase descriptively defines the most important time scale using mode statistics. A brief description on each phase is as follows.

3.1 Phase I. The choices

The first phase theoretically defines the choices of meteorological stations and the type of drought indices. This phase has two steps. The first step consists of the selection of regions and meteorological stations. The second step consists of the selection of drought indices. A brief description on each step is given below.

3.1.1 Selection of regions and stations

The selection of regions in drought monitoring has great significance for accurate and efficient drought mitigation policies. In Pakistan, over the last few decades, a rise is observed in the occurrence and strength of intense climatic events, and about 40% of the people are affected due to different disasters including droughts (Hussain et al., 2010). Pakistan is the country, which is exposed to several natural calamities, including cyclones, floods, drought, intense rainfall, and earthquakes. In this research, 46 meteorological stations are taken from all over Pakistan. However, some regions have a long duration of hot season, while some have cold climate during the year.

3.1.2 Selection of drought indices

The drought condition of a region is monitored by drought indices. In the past few decades, several drought indices have been proposed, but some of those are region specific and have limitations of applicability in other climatic conditions. Here, for drought monitoring of regional temperature, SPTI is the most appropriate to utilize because it monitors drought in a

Fig. 2 Flow chart of the proposed framework

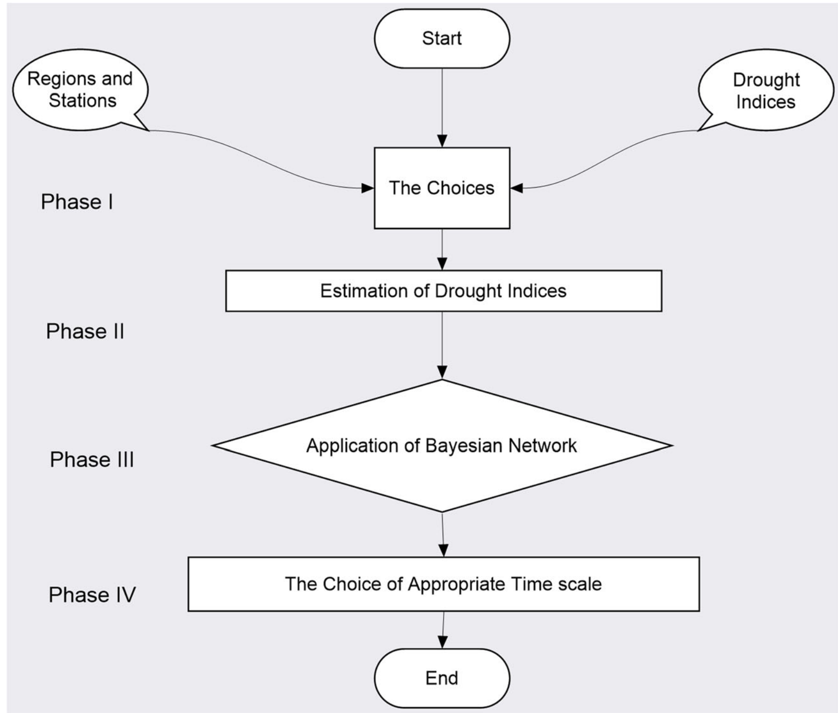


Fig. 3 Observed characteristic of probability function of DMAI-1 series at Gupis Stations. (**a**) Histogram and (**b**) Q-Q plot

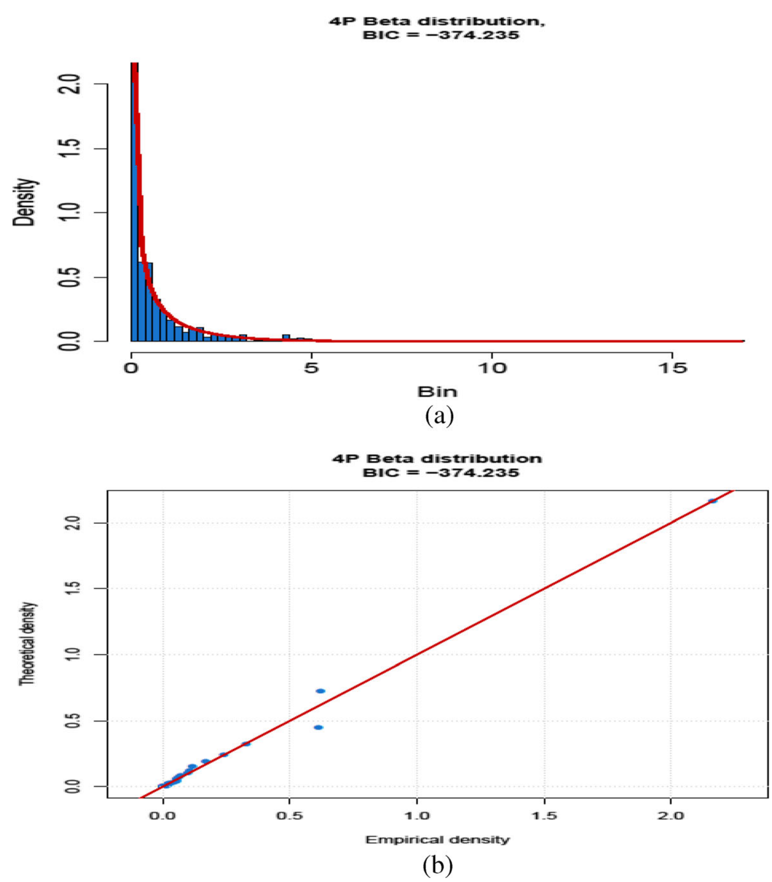


Table 1 BIC values of candidate distributions for SPTI-1

Distribution	Gupis	Muzaffarabad	Sargodha	Quetta	Hyderabad
2P beta	−254.435	−275.8449	−177.91525	−329.725	−168.2607
3P Weibull	−370.546	−547.5354	−296.30557	−563.222	−286.3358
4P beta	−374.235	−576.8157	−290.8399	−548.835	−279.6342
Arcsine	−151.865	−359.6553	−116.55661	−325.434	−217.668
Burr	−21.1118	−349.7142	−15.76597	141.6663	184.7364
Cauchy	−235.773	−461.3411	−125.41078	−369.963	−202.6503
Chi	−188.775	−190.5887	−205.77795	2.050577	137.6462
Chi-square	−146.954	−459.2853	−245.05717	12.75368	141.916
Cosine	13.08168	−396.0813	15.50567	159.4744	191.3188
Curvilinear trapezoidal	−158.547	−546.7706	−90.51514	−317.009	−161.9658
Exponential	−206.773	−552.8117	−99.62765	−49.6274	−165.8195
F-Distribution	−124.805	−358.3821	−224.00279	41.37734	153.7853
Gamma	−294.282	−548.4759	−240.72583	−332.224	−169.2019
Generalized extreme value	−338.986	−523.5672	−204.73996	−509.05	−287.974
Generalized normal	−355.536	−529.4466	−238.4285	−540.256	−283.5128
Gumbel	−205.277	−523.8075	−97.1078	−327.7	−168.5012
Inverse chi-square	−126.691	−374.7858	−53.02042	−287.589	−144.567
Inverse Gamma	−333.927	−422.4105	−191.52094	−375.505	−188.7289
Inverse Gaussian	−334.881	−429.726	−126.77096	−346.197	−173.4299
Johnson SB	−351.094	−540.0478	−237.71476	−549.067	−284.8226
Johnson SU	−351.094	−524.7102	−233.91768	−547.613	−284.823
Laplace	−238.05	−479.9917	−115.02854	−329.333	−169.3704
Logistic	−187.983	−494.2572	−89.62024	−326.341	−167.791
Log-normal	−338.985	−465.1849	−223.37539	−352.932	−177.8327
Normal	−172.804	−495.5055	−81.28676	−322.432	−166.2896
Rayleigh	−182.467	−520.5913	−85.2244	−323.046	−166.425
Scaled/shifted t-distribution	−237.547	−490.7707	−133.94935	−425.646	−247.9503
Skewed-normal	−185.693	−550.6413	−83.25483	−318.297	−162.1895
Trapezoidal	−151.711	−539.5586	−56.9338	−312.119	−157.6891
Triangular	−148.561	−544.2948	−61.44465	−317.009	−161.9658
Uniform	−37.5422	−266.7115	−24.727	72.82902	179.3031
von Mises	−173.353	−398.099	−81.49773	−322.445	−166.291

specified region and deals with multi-scaling. Furthermore, over different time scales, SPTI can be used for monitoring, examining, and detecting drought conditions.

3.2 Phase II. Estimation of drought indices

This phase is straightforward. In this phase, a suitable estimation procedure of drought indices is suggested. In this research, we followed the estimation procedure provided by Stagge et al. (2015). We recommend to include multiparameter extreme values probability distributions for the computation drought indices. To make the list of candidate distributions, one can employ goodness of fit criteria such as chi-squared test and Anderson darling test using *easyfit (*or other software. In addition, the selection of the appropriate distribution can be based on Bayesian Information Criterion (BIC).

3.3 Phase III. Configuration of BNs model

This phase explains the utilization of BNs models on SDI scales data of different meteorological stations. Basically, BNs are directed acyclic graphs (DAGs), in which nodes represent variables in the Bayesian logic. Conditional dependencies are presented by edges. In each node, there is a probability function that takes input from a particular set of values. Moreover, BNs is applied for decision-making processes under conditions of climatic variability and uncertainty, providing logical and holistic reasoning in complex systems. Further, BNs effectively translate the relationship between variables under probabilistic approach (Catenacci and Giupponi 2009). Here, the main purpose is to identify a better scale by obtaining a probabilistic model signifying the uncertainty in the network of various time scales of SPTI index. This phase configures BNs on various time scales of SPTI index.

3.4 Phase IV. The choice of time scale

The main objective of this study is to develop a new framework for the selection of time scale for meteorological stations by incorporating algorithm BIRDts and SPTI time series data. From the results of probabilistic paradigm of BNs, a scale is recommended for all meteorological stations by following three steps:

(a) Let S_1, S_2, S_k be the time scale at a particular meteorological stations. Consider that the marginal posterior distribution at a particular station in a single BNs run has the following mathematical form.

$$\begin{array}{c} \\ S_1 \\ S_2 \\ \vdots \\ S_n \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & \dots & S_n \end{array} \\ \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n1} & \tau_{n2} & \cdots & \tau_{nn} \end{bmatrix} \end{array}$$

By utilizing Eq. (3), the elements in the above matrix are obtained. There are at least three MCMC simulations (means three 7*7 matrices). In the above matrix, the Dependence Probability (DP) between seven scales is shown by the lower side of off diagonal, while independence features are explained by the upper diagonal.

Table 2 BIC values of candidate distributions of all stations at different scales

Scales	Station					
		Gupis	Muzaffarabad	Sargodha	Quetta	Hyderabad
SPTI-1	Distribution	4P beta	4P beta	3P Weibull	3P Weibull	Generalized extreme value
	BIC	−374.235	−576.816	−296.306	−563.222	−287.974
SPTI-3	Distribution	Chi-square	Rayleigh	Gamma	4P beta	3P Weibull
	BIC	−432.182	−645.512	−298.317	−495.494	−416.633
SPTI-6	Distribution	Johnson SU	Chi-square	Rayleigh	3P Weibull	3P Weibull
	BIC	−410.039	−705.226	−424.626	−523.324	−331.421
SPTI-9	Distribution	Johnson SU	Log-normal	Chi-Square	Rayleigh	Exponential
	BIC	−463.209	−925.242	−499.759	−535.565	−366.749
SPTI-12	Distribution	Johnson SU	Gumbel	Rayleigh	Trapezoidal	Gamma
	BIC	−466.565	−827.293	−503.759	−633.020	−377.403
SPTI-24	Distribution	Scaled/shifted t	Gumbel	Cosine	Log-normal	Triangular
	BIC	−629.456	−570.797	−311.136	−530.517	−310.916
SPTI-48	Distribution	Inverse Gaussian	Logistic	Laplace	Logistic	Laplace
	BIC	−532.519	−976.548	−431.542	−659.830	−471.643

(b) The average of each MPP distribution matrix column is taken separately, i.e., $\tau_{.1} = \frac{\sum_{i=1}^{n} \tau_{.i}}{n}$. We called the probabilities as Average Dependency Probabilities (ADP). Therefore, each scale has three ADP (i.e., ADP_1, ADP_2, and ADP_3).

(c) Furthermore, let the grand average be called Average Marginal Posterior Probability (AMPP) (Ali et al., 2020c) for each scale by the following equation:

$$AMPP = \frac{\mathrm{ADP1} + \mathrm{ADP2} + \mathrm{ADP3}}{3} \tag{4}$$

As we have seven columns (i.e., 7 time scales), therefore, there are seven AMPP values for each station. By following this procedure, the best scale is chosen for all meteorological stations based on maximum AMPP values for further data mining.

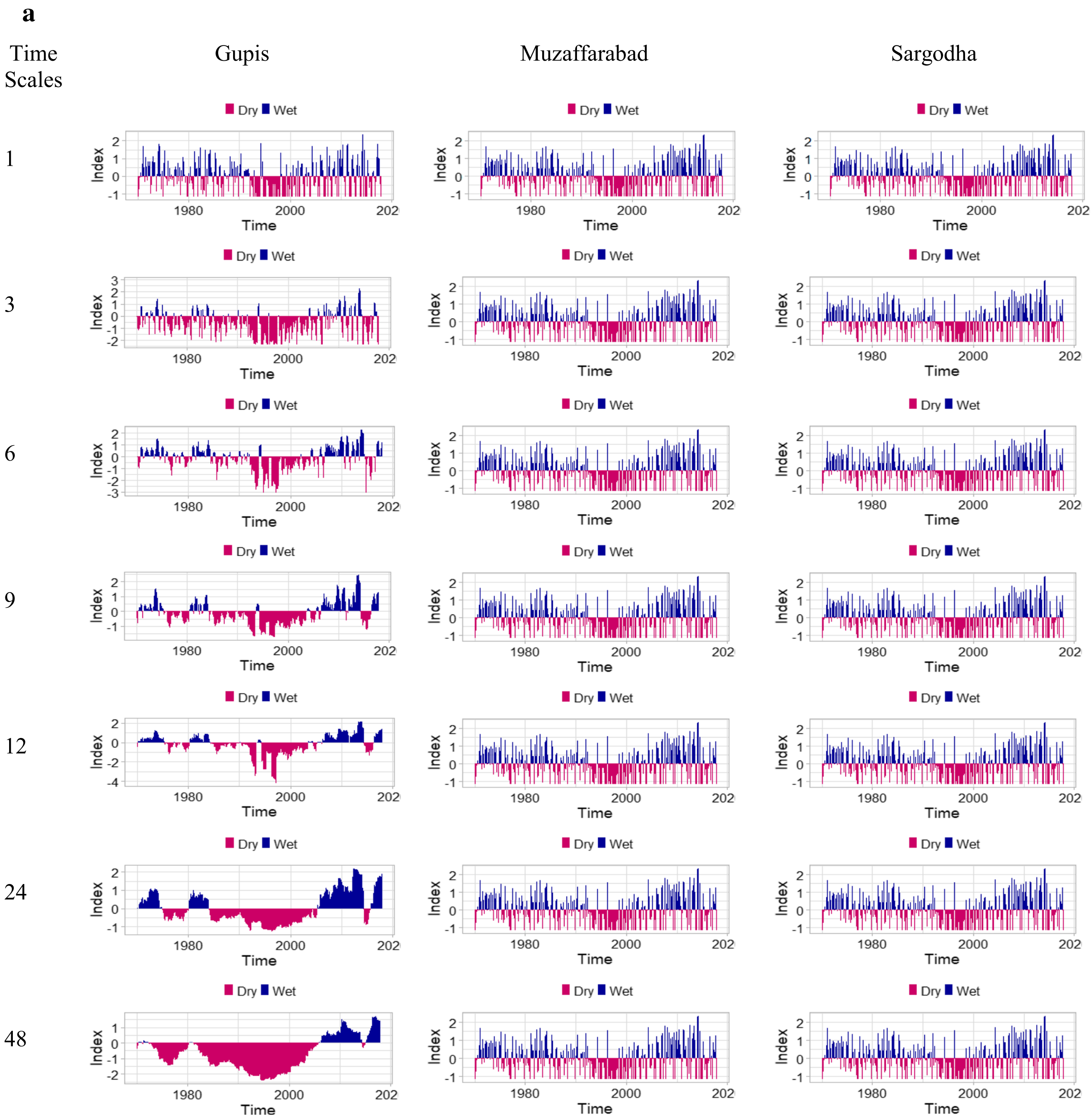


Fig. 4 (**a**) Temporal behavior of SPTI time scales in some selected stations and (**b**) temporal behavior of SPTI time scales in some selected stations

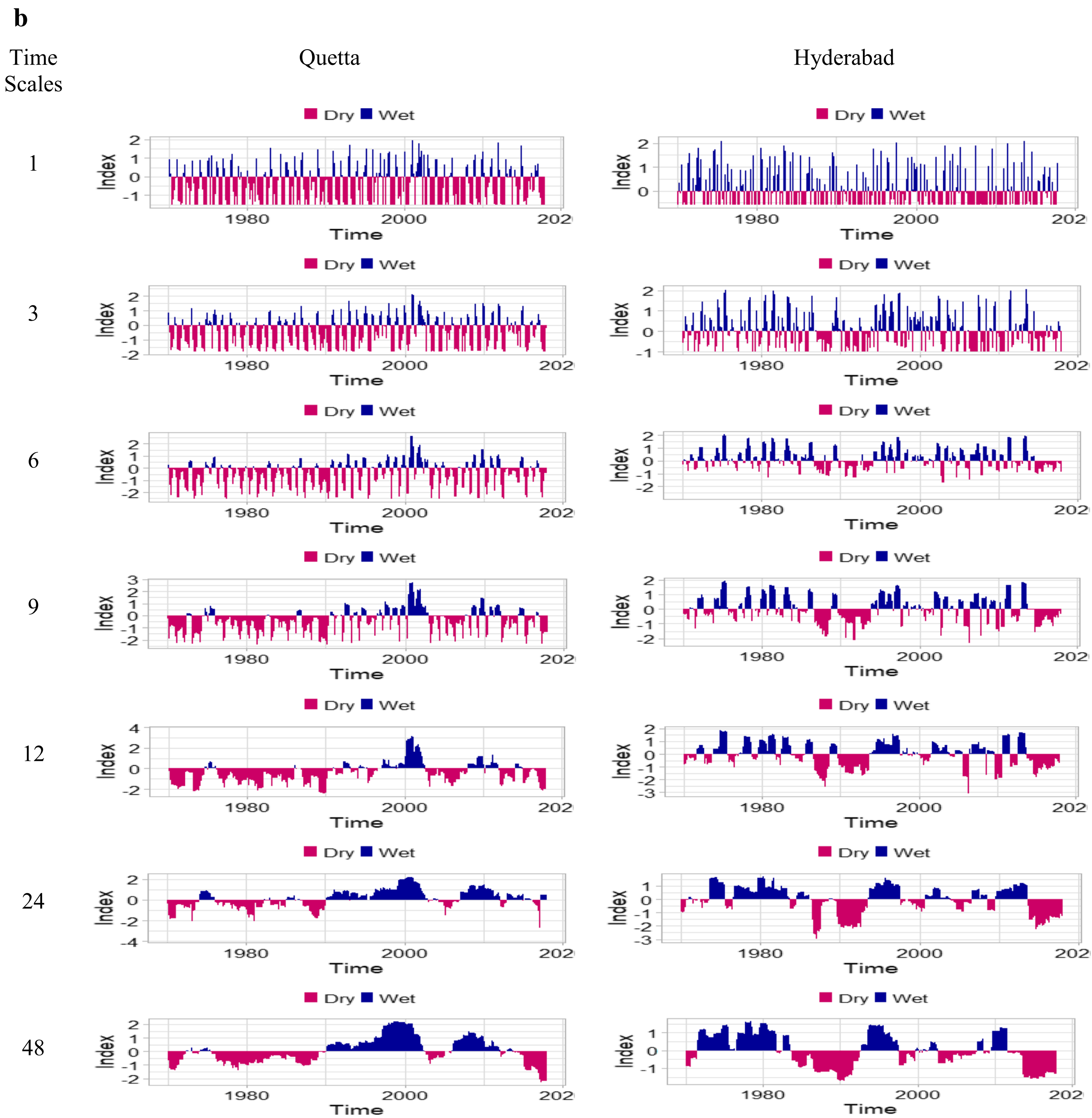


Fig. 4 (continued)

4 Simulation settings and quality measures

4.1 Data setting and simulation study

In the independent run of BNs, the SPTI is configured into seven columns. Each column consists of time series data of individual time scale. This study is based on seven time scales (i.e., SPTI-1, SPTI-3, SPTI-6, SPTI-9, SPTI-12, SPTI-24, and SPTI-48); therefore, we have seven columns. In structure MCMC simulation setting, a total of 200,000 iterations with 100 burn-in steps are configured subjectively (Grzegorczyk, 2010).

4.2 Quality measure and validation

The main objective of this research is the identification of the best time scale of SPTI index based on AMPP. In this article, the dependency of time scale was assessed by observing maximum MPP values. In the previous research, Ali et al. (2020c) have used AMPP of three simulation runs for the identification of the most important time scale.

For the validation of the model, we propose to observe the convergence of the posterior predictive probabilities in the three MCMC simulation runs. Here, posterior predictive checks are a form of internal model validation. In addition, the quality of results is proposed to be assessed using three independent simulation runs. Moreover, the final inference is based on the average of three AMPP values. To observe the convergence, scatter plots between each simulation run can be used. Further, the homogenous behavior of each simulation can be observed using trace plots. In this article, some results associated with quality measure are presented graphically, while most of the results are archived in the author's gallery.

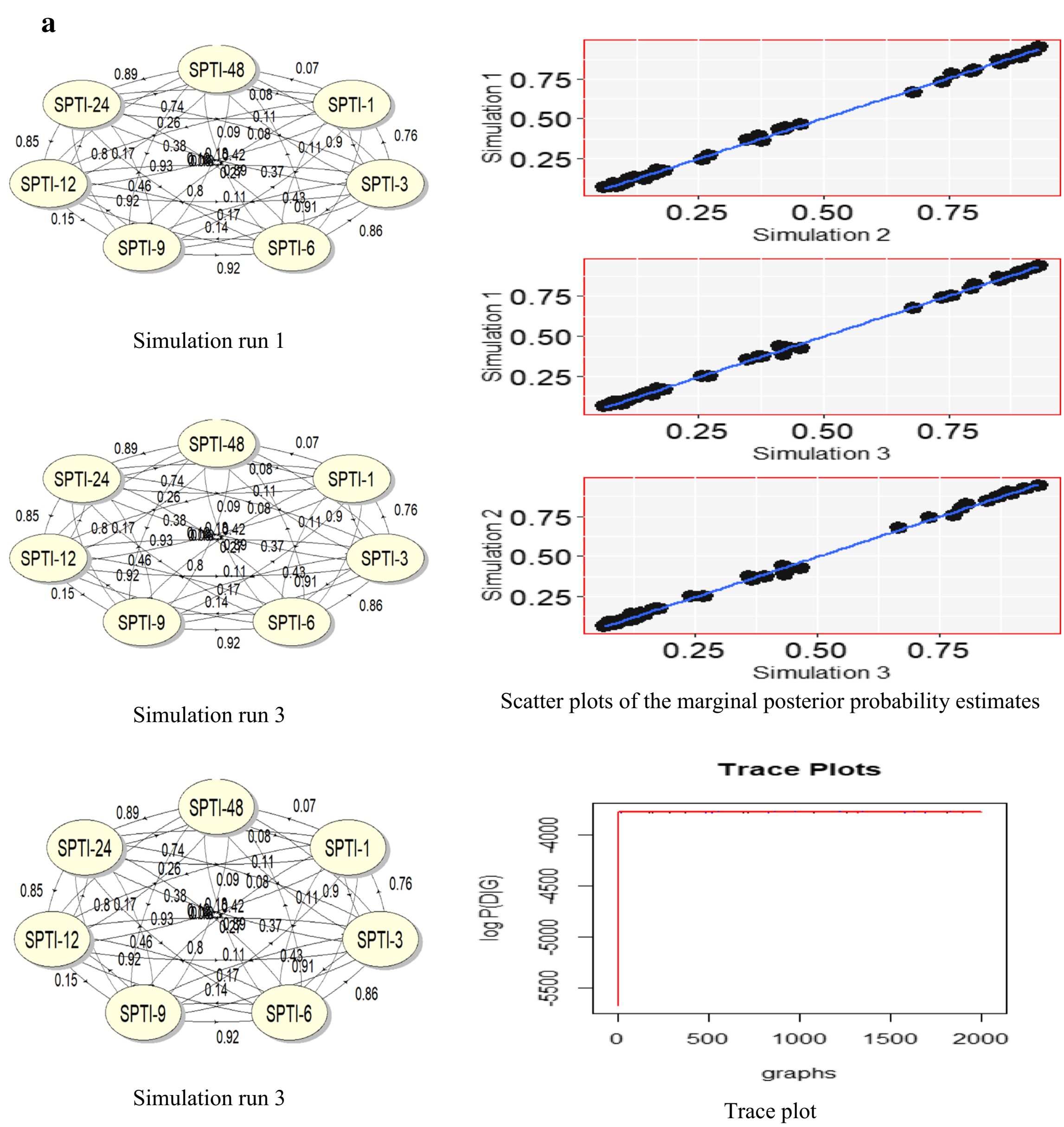


Fig. 5 (**a**) MPP matrix of three simulation runs and their scatter plots at Gupis stations, (**b**) MPP matrix of three simulation runs and their scatter plots at Muzaffarabad stations, (**c**) MPP matrix of three simulation runs and their scatter plots at Sargodha stations, (**d**. MPP matrix of three simulation runs and their scatter plots at Quetta stations, and (**e**) MPP matrix of three simulation runs and their scatter plots at Hyderabad stations

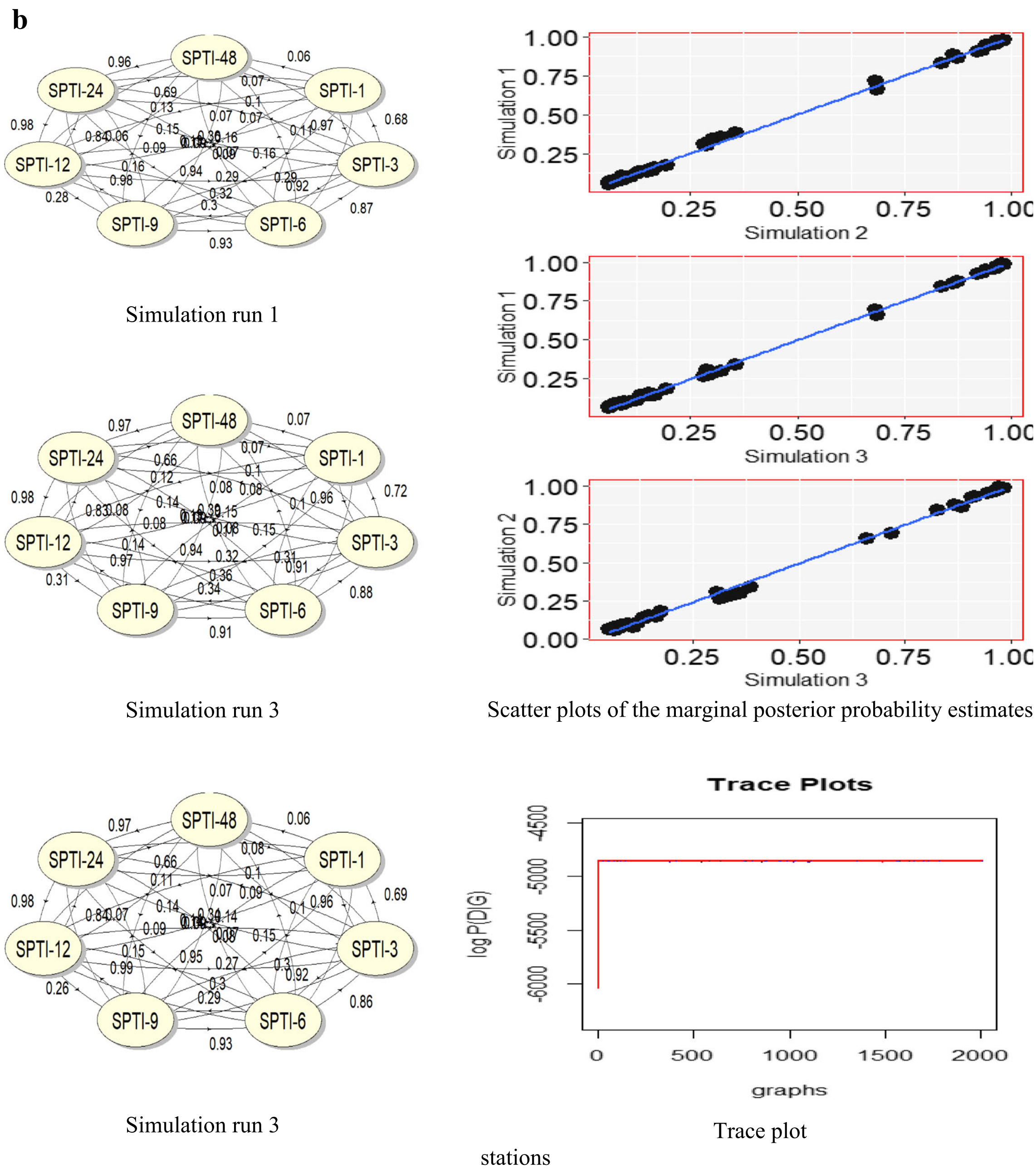


Fig. 5 (continued)

5 Results

5.1 Estimation of drought indices

To assess and evaluate the proposed framework, we computed time series data of SPTI using varying probability distribution concept provided by Stagge et al. (2015). A brief description on the standardization process is as follows:

In this first step, we employed *fitDistr* function of R package propagate (*Spiess*, 2014) to find the appropriate probability function under Bayesian Information Criterion (BIC). For example, Fig. 3 shows the appropriate probability distribution for DMAI-1 series at Gupis Stations. QQ-plot in Fig. 3 shows the scatter plot of theoretical and empirical density. All findings related to probability distribution fitness are done accordingly. We have selected those probability functions that have

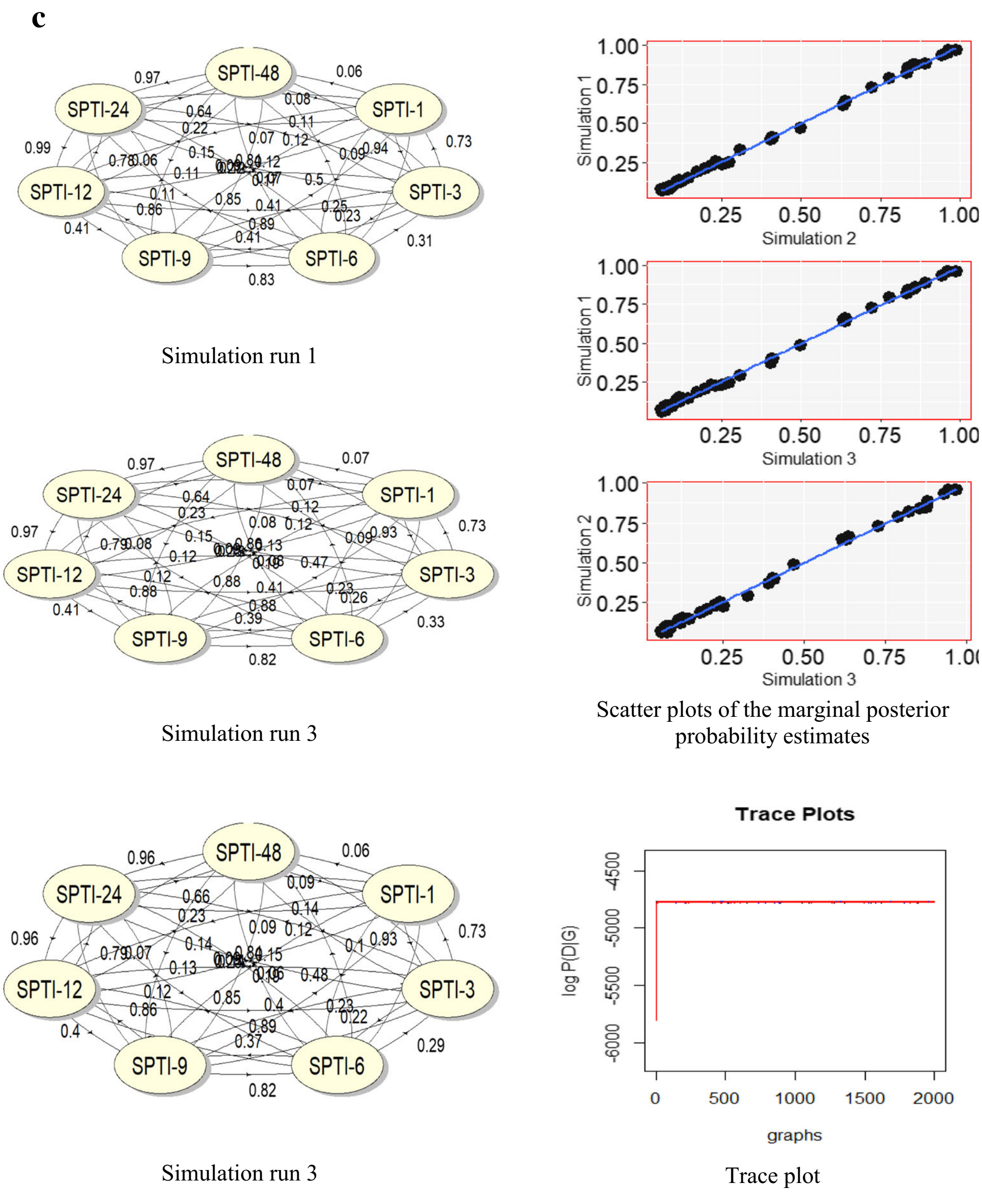


Fig. 5 (continued)

minimum BIC values. Table 1 shows the BIC values of all candidate distributions in the estimation phase of SPTI-1 in five stations. For all the 46 stations, similar practice with all time scales has been done.

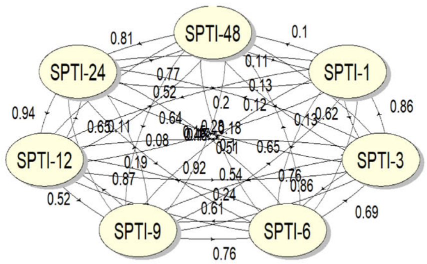
After the selection of the probability functions at varying time scales for all stations, the next step is to standardize their CDF under the appropriate standardization procedure. By following Ali et al. (2017) and Vicente-Serrano et al. (2010), this research employed standardization procedure provided by McKee et al. (1993).

Table 2 shows BIC values of selected distributions in some stations at different scales. We have observed that for the estimation of SPTI-1, 4P beta distribution is best fitted for Gupis and Muzaffarabad stations, while 3P Weibull distribution is more appropriate for Sargodha (BIC = − 296.306) and Quetta (BIC = − 563.222) stations. On the other hand, generalized extreme value distribution with BIC value − 287.974 is selected for Hyderabad region.

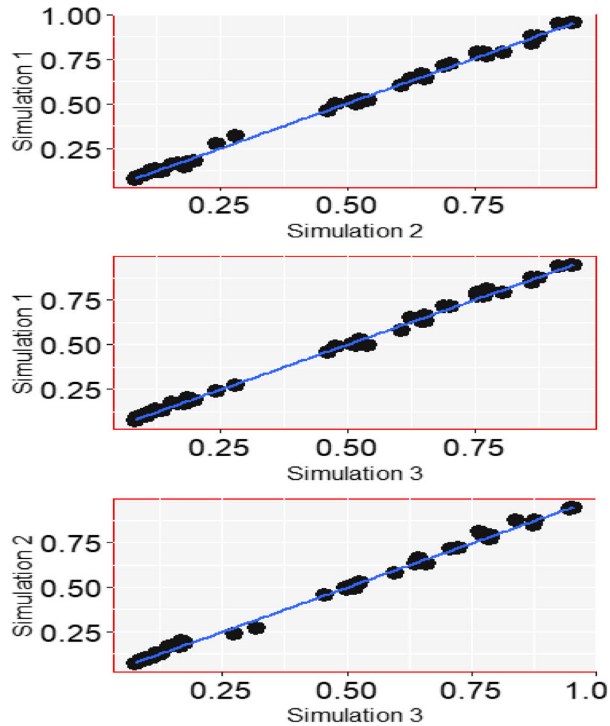
While the Rayleigh distribution is more appropriate for the estimation of SPTI-3 for Muzaffarabad. Further, results reveal that Rayleigh distribution is also best fitted for the estimation of SPTI-6 and SPTI-12 in Sargodha station. On the same lines, 3P Weibull distribution is fitted for the estimation of SPTI-1 in Quetta and Sargodha and for the estimation of SPTI-3 in Quetta stations again. The calculation and inference on the rest of all the stations are carried out on the same line.

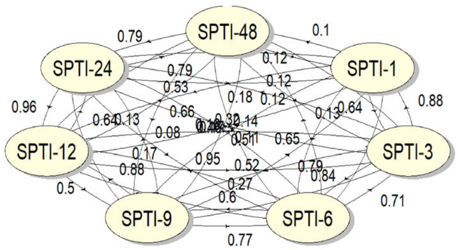
Figure 4 a and b present the temporal behavior of drought indices in their various time scales in some selected stations.

d



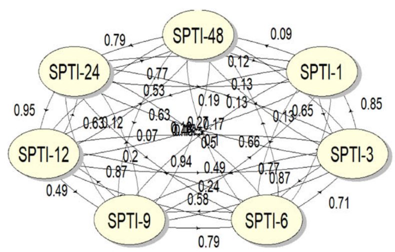
Simulation run 1



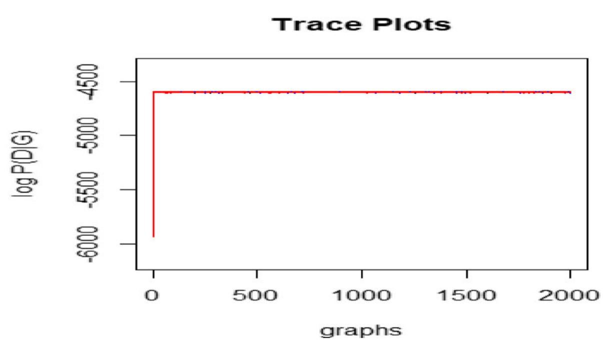


Simulation run 3

Scatter plots of the marginal posterior probability estimates



Simulation run 3



Trace plot

Fig. 5 (continued)

The inferential data on the distribution selection process and standard time series of all the stations are archived in the author's gallery and available on request.

5.2 Bayesian networks implementation and the choice of the scale for metrological station

After the estimation of drought indices, the next step is to configure time series data in BNs and to draw inference according to the proposed framework. In this article, time series data of seven scales of SPTI index at 46 meteorological stations is considered, where separate BNs models are applied on all the selected meteorological stations under structured MCMC simulation. The quality of simulation runs is assessed by scatter plot of three independent runs of MCMC. The scatter plots are based on posterior probabilities that are investigated for validation of Bayesian network.

Figures 5 a–e show MPP matrix of three simulation runs, their scatter plots, and the trace plots at Gupis, Muzaffarabad, Sargodha, Quetta, and Hyderabad, respectively. Outcome associated with the three simulation runs show that there is no significant difference between each marginal posterior probability matrix. Similar findings have been observed in another time scales and stations as well.

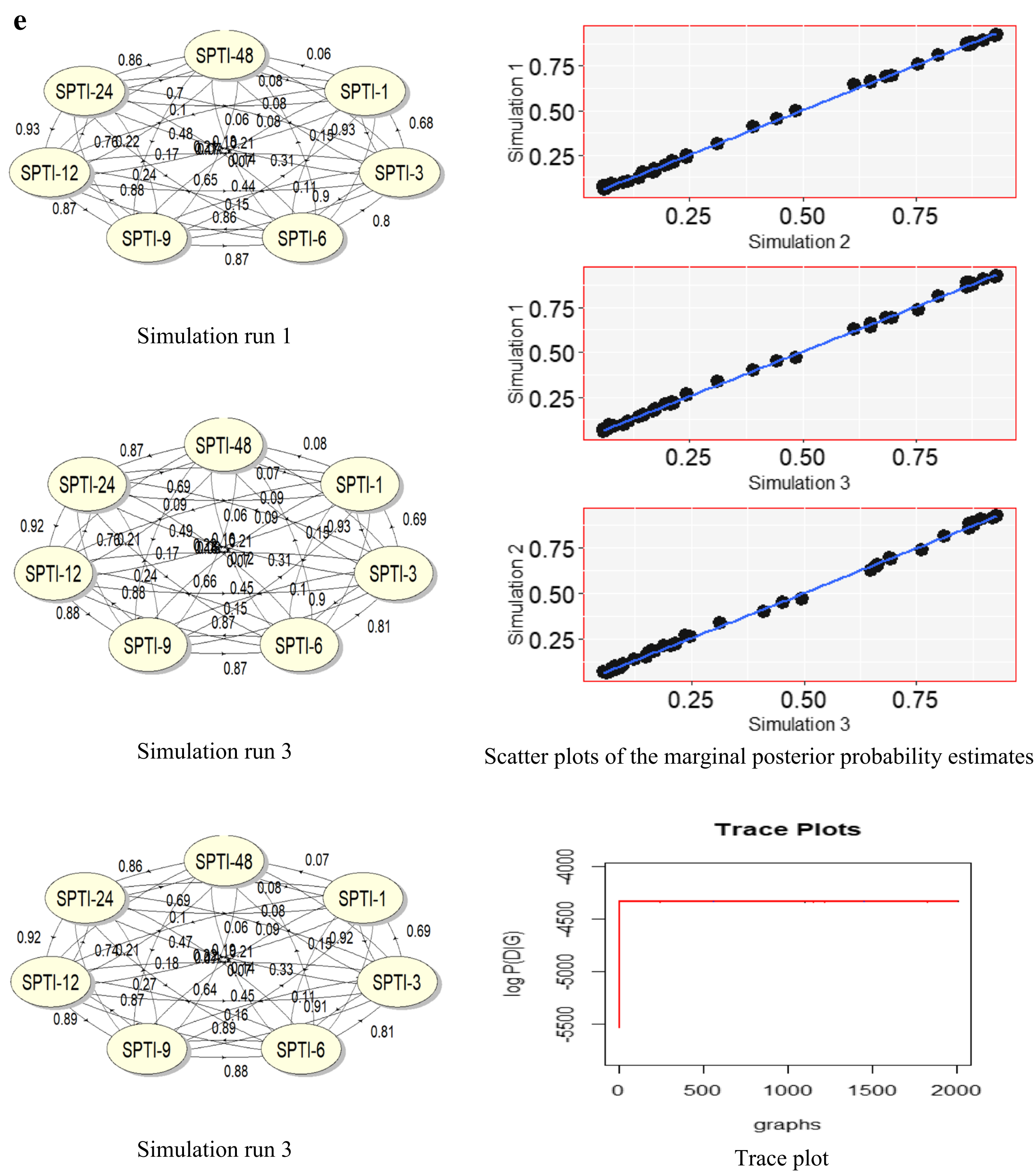


Fig. 5 (continued)

The ADPs and AMPP for Gupis, Muzaffarabad, Sargodha, Quetta, and Hyderabad in each scale are shown in Table 3. Based on the highest values of AMPP, it has been observed that SPTI-9 is the most appropriate time scale in all the five selected stations.

Table 4 shows the ADPs and AMPP values for Gupis, Muzaffarabad, Sargodha, Quetta, and Hyderabad. Results show that SPTI-9 is best representative for Gupis, Muzaffarabad, and Sargodha with AMPP values 0.78648, 0.55795, and 0.511742, respectively, while SPTI-12 is selected for Hyderabad and Quetta region with AMPP values 0.50225 and 0.55433, respectively.

For the whole region, the mode measure of central tendency is used to confer important time scale. Hence, Table 5 provides the list of stations and important time scales in various administrative boundaries of Pakistan. This completes the analysis.

Table 3 ADPs and AMPP distribution at different scales

Stations	Simulation runs	SPTI-1	SPTI-3	SPTI-6	SPTI-9	SPTI-12	SPTI-24	SPTI-48
Gupis	ADP1	0.291798	0.357518	0.444444	0.77401	0.445671	0.419033	0.295216
	ADP2	0.30319	0.357431	0.444707	0.791623	0.436996	0.408079	0.295917
	ADP3	0.295216	0.359534	0.441903	0.793814	0.431826	0.404224	0.296793
	AMPP	0.296734	0.358161	0.443685	*0.78648*	0.438165	0.410445	0.295975
Muzaffarabad	ADP1	0.264458	0.393095	0.415703	0.558447	0.421574	0.331493	0.257711
	ADP2	0.263056	0.404749	0.422801	0.557308	0.434805	0.340168	0.256222
	ADP3	0.263845	0.394059	0.415177	0.558097	0.424115	0.338503	0.258675
	AMPP	0.263787	0.397301	0.417893	*0.55795*	0.426831	0.336722	0.257536
Sargodha	ADP1	0.344462	0.396775	0.407816	0.514809	0.502191	0.340782	0.246758
	ADP2	0.336751	0.3881	0.39476	0.509639	0.492552	0.338328	0.251753
	ADP3	0.341833	0.401945	0.405538	0.51078	0.503943	0.334648	0.249036
	AMPP	0.341015	0.395607	0.402705	*0.511742*	0.499562	0.337919	0.249182
Hyderabad	ADP1	0.277252	0.398177	0.502191	0.427708	0.50368	0.368822	0.349281
	ADP2	0.276902	0.399404	0.499387	0.43428	0.499649	0.371626	0.348668
	ADP3	0.273572	0.393971	0.498598	0.424115	0.503417	0.369699	0.35007
	AMPP	0.275908	0.397184	0.500058	0.428701	*0.50225*	0.370049	0.34934
Quetta	ADP1	0.415177	0.462583	0.540834	0.507273	0.562128	0.38994	0.373116
	ADP2	0.394848	0.465212	0.540221	0.480021	0.550824	0.384858	0.375131
	ADP3	0.418069	0.451455	0.535839	0.515247	0.550035	0.391342	0.373992
	AMPP	0.409364	0.45975	0.538965	0.500847	*0.55433*	0.388714	0.37408

Table 4 AMPP values in all stations

Region	Station	Latitude	Longitude	Scale-1	Scale-3	Scale-6	Scale-9	Scale-12	Scale-24	Scale-48
AJK	Garhi dupatta	34° 13’	73° 37’	0.2684	0.4213	0.4428	*0.5632*	0.4440	0.3747	0.2768
	Kotli	33° 31’	73° 54’	0.2804	0.4730	0.4294	*0.5285*	0.3879	0.3954	0.3268
	Muzaffarabad	34° 22’	73° 29’	0.2638	0.3973	0.4179	*0.5580*	0.4268	0.3367	0.2575
Baluchistan	Zhob	31°-21′	69°-28′	0.3743	0.3716	*0.4550*	0.4119	0.3890	0.3795	0.3662
	Sibbi	29°-33′	67°-53′	0.2527	0.4811	*0.6296*	0.4744	0.5160	0.3764	0.4070
	Quetta	30°-11′	66°-57′	0.4094	0.4597	0.5390	0.5008	*0.5543*	0.3887	0.3741
	Panjgur	26° 58’	64° 06’	0.3741	0.3232	0.4139	0.3817	*0.4301*	0.3271	0.3616
	Nokkundi	29° 44’	62° 94’	0.3587	0.4445	0.4439	0.4477	*0.4891*	0.4037	0.3029
	Lesbella	25° 00’	66° 29’	0.2788	0.2501	0.3252	*0.4305*	0.3695	0.3580	0.3440
	Kalat	29°02’	66°35’	0.5480	0.4537	0.5457	*0.5833*	0.5355	0.3502	0.2615
	Dalbadin	28°54’	64°64’	0.4319	0.4265	0.4996	*0.5192*	0.5096	0.3444	0.2767
	Jiwani	25°38’	61°07’	0.4144	*0.6087*	0.5195	0.4693	0.5805	0.3600	0.2941
	Pasni	25° 16’	63° 29’	0.2901	0.3293	0.3666	0.3670	*0.3753*	0.3301	0.2502
KPK & Northern Area	Astor	35° 20’	74° 54’	0.3898	0.4982	*0.5636*	0.5225	0.5389	0.3462	0.2814
	Chillas	35° 25’	74° 06’	0.3288	0.3929	*0.5291*	0.3740	0.4706	0.3686	0.2476
	Gilgit	35° 55’	74° 20’	0.3362	0.3919	*0.4191*	0.3297	0.4141	0.3143	0.2819
	Gupis	36° 10’	73° 24’	0.2967	0.3582	0.4437	*0.7865*	0.4382	0.4104	0.2960
	Skardu	35° 18’	75° 41’	0.3856	0.4314	0.5328	0.4851	*0.5912*	0.3618	0.3673
	Cherat	33° 49’	71° 33’	0.2517	0.4011	0.3250	*0.4897*	0.4364	0.3189	0.2083
	Chitral	35° 51’	71° 50’	0.4369	0.4331	0.5090	*0.5635*	0.5567	0.3505	0.3181
	Drosh	35° 34’	71° 47’	0.3962	0.4929	0.5491	0.5537	*0.5660*	0.3320	0.3390
	Kohat	33° 89’	71° 29’	0.2612	0.3669	0.4920	0.3969	*0.5848*	0.3656	0.3634
	Parachinar	33° 52’	70° 05’	0.3449	0.3725	*0.4588*	0.3733	0.4498	0.4463	0.3270
	Peshawar	34° 02’	71° 56’	0.3007	0.3264	0.3658	*0.3674*	0.3654	0.3321	0.2512
	Risalpur	34° 51’	71° 76’	0.2488	0.4253	0.4239	0.5285	*0.5385*	0.3630	0.3653
	Balakot	34° 33’	72° 21’	0.2491	0.4552	0.5050	*0.5387*	0.4790	0.4351	0.2734
	Kakul	34° 11’	73° 15’	0.2596	0.4388	0.4420	*0.5488*	0.4595	0.3549	0.2551
	DIK	31° 49’	70° 56’	0.2450	0.4763	0.4902	*0.5059*	0.4220	0.4814	0.3423
Punjab	Murree	33°54’	73°23’	0.2551	*0.4668*	0.4637	0.3775	0.3648	0.4435	0.3200
	Bahawalpur	29° 20’	71° 47’	0.2546	*0.5218*	0.3806	0.4469	0.4474	0.3389	0.2800
	Bahawalngar	29° 20’	73° 51’	0.3825	0.3671	*0.4463*	0.3824	0.3426	0.3696	0.2709
	Faisalabad	31° 26’	73° 08’	0.2866	0.4026	0.3836	0.3482	*0.4076*	0.3389	0.3359
	Lahore PBO	31° 33’	74° 20’	0.2649	0.3519	0.4380	0.2978	*0.4964*	0.3506	0.2864
	Mianwali	32° 58’	71° 53’	0.2461	0.4521	0.4529	*0.5360*	0.4811	0.3510	0.2794
	Multan	30° 12’	71° 26’	0.2472	0.4384	0.3547	0.3664	*0.4387*	0.3264	0.2774
	Sargodha	32° 03’	72° 40’	0.3410	0.3956	0.4027	*0.5117*	0.4996	0.3379	0.2492
	Sialkot	32° 31’	74° 32’	0.2762	0.3952	0.4250	*0.5365*	0.3941	0.3417	0.2591
	Khanpur	28° 39’	70° 41’	0.2466	0.3632	*0.5210*	0.1831	0.4220	0.3597	0.3776
Sindh	Chor	25° 51’	69° 76’	0.3844	0.3622	0.4310	0.3754	*0.4839*	0.3651	0.3579
	Rohri	27° 40’	68° 54’	0.2447	0.4674	0.3837	*0.4716*	0.3718	0.3460	0.2600
	Padidian	26° 51’	68° 08’	0.2432	0.4608	0.3806	*0.5118*	0.3795	0.3552	0.3229
	Nawabshah	26° 15’	68° 22’	0.2514	0.3507	0.4050	0.3829	*0.5008*	0.3593	0.3611
	Jacobabad	28° 18’	68° 28’	0.2461	0.4670	0.4086	*0.5718*	0.3641	0.4708	0.3177
	Hyderabad	25° 23’	68° 25’	0.2759	0.3972	0.5001	0.4287	*0.5022*	0.3700	0.3493
	Badin	24° 38’	68° 54’	0.3292	0.3475	0.4083	*0.4246*	0.3852	0.3519	0.2514
	Karachi	24° 54’	66° 56’	0.2552	0.4462	0.3823	*0.4647*	0.3661	0.3323	0.2461

Table 5 List of stations and important time scales in various administrative boundaries of Pakistan

Scales	Administrative regions				
	Punjab	Balochistan	Sindh	KPK	AJK
SPTI-1	Nil	Nil	Nil	Nil	Nil
SPTI-3	Murree, Bahawalpur	Jiwani	Nil	Nil	Nil
SPTI-6	Bahawalngar, Khanpur	Zhob, Sibbi	Nil	Astor, Chillas, Gilgit, Parachinar	Nil
SPTI-9	Mianwali, Sargodha, Sialkot	Lesbella, Kalat, Dalbadin	Rohri, Padidian, Jacobabad, Badin, Karachi	Gupis, Cherat, Chitral, Peshawar, Balakot, Kakul, DIK	Garhi dupatta, Kotli, Muzaffarabad
SPTI-12	Faisalabad, Lahore, Multan	Quetta, Panjgur, Nokkundi, Pasni,	Chor, Nawabshah, Hyderabad	Skardu, Drosh, Kohat, Risalpur	
SPTI-24	Nil	Nil	Nil	Nil	Nil
SPTI-48	Nil	Nil	Nil	Nil	Nil

6 Discussion

The past studies reveal that the simultaneous inference on drought indices at their various time scales provides a more effective and clearer image of drought. However, large volumes of data on various time scales create chaotic problems for researchers in reanalysis at regional level. Therefore, the selection of a suitable time scale for a region is very important.

In this article, the proposed framework is used to select important time scales at regional level. For this purpose, we first calculate time series data of SPTI index. At this stage, probability distributions are fitted to check the appropriateness of the respective time series of each scale in the estimation procedure. Here, we employed various distribution concepts under the parametric approach (Stagge et al., 2015). This exaggeration in the perceived severity of drought events and bias in drought index values can occur if inappropriate probability distributions are applied. Therefore, the varying probability distribution procedures under the parametric approach are used as suggested by Stagge et al. (2015).

For decision-making under uncertainty, BNs are increasingly recognized as a valuable tool. BNs provide a transparent, defensible evidence base for mapping and quantifying the important scale.

To avoid the multiple scaling problems, the most appropriate scale is chosen for the whole region as well as for characterized five regions. Overall SPTI-9 can be utilized for the 46 meteorological stations. While region wise according to Table 5 for Punjab SPTI-9 and SPTI-12 both scales can be utilized, SPTI-9 can be used for Sindh, AJK, and KPK. However, SPTI-12 followed by SPTI-9 can be used for Baluchistan region.

7 Summary and conclusion

In drought monitoring and prediction, the choices of important time scales are the most crucial aspect in hydrological data analysis, reanalysis, and complex statistical modeling at a specific region. In this paper, we have proposed a new data mining framework—the BIRDts. The framework of BIRDts identifies the important time scales of SPTI in a certain region. For the validation of the model, we mainly rely on the convergence of the posterior predictive probabilities in the three MCMC simulation runs.

Application of BIRDts is made for 46 meteorological stations located in various climatological clusters and regional administrative boundaries of Pakistan. Under the proposed framework, the posterior predictive checks are a form of internal model validation. Our computational results show that the models are adequate. That is, there is no need to improve the model in current simulation settings. Moreover, the results and inferences of this paper are based on three independent structures MCMC based simulation runs of Bayesian network models.

We inferred that 9-month time scale of SPTI index is the most important in terms of dependency probabilities of Bayesian network theory at Sindh, AJK, KPK, and Punjab, while 12-month time scale is the most important in Baluchistan region. Consequently, the resultant time series in each region are recommended for efficient, regional analyses and reanalysis.

Moreover, we have concluded that (1) the problem of multi-scaling raised by Ali et al. (2020a, b) and Bazrafshan et al. (2014) can be addressed in more adequate and sufficient way using Bayesian network theory and (2) the proposed framework is rather general and can be used for any region and any drought index. In summation, the proposed framework may be considered a data mining device for efficient, regional analyses and reanalysis of hydrological data.

Authors' contribution All authors have equal contribution.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

References

Abramowitz, M., & Stegun, I. A. (1965). Handbook of mathematical functions with formulas, graphs, and mathematical table. In *US Department of Commerce*. National Bureau of standards applied mathematics series 55

Ali Z, Hussain I, Faisal M, Nazir HM, 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 IM, Ismail M, Ahmad M, Ahmad I (2018) A New Weighting Scheme in Weighted Markov Model for Predicting the Probability of Drought Episodes. Adv Meteorol:2018

Ali Z, Hussain I, Faisal M, Elashkar EE, Gani S, Shehzad MA (2019a) Selection of appropriate time scale with Boruta algorithm for regional drought monitoring using multi-scaler drought index. Tellus A Dyn Meteorol Oceanogr 71(1):1604057

Ali, Z., Hussain, I., Faisal, M., Shoukry, A. M., Gani, S., & Ahmad, I. (2019b). A framework to identify homogeneous drought characterization regions. *Theoretical and Applied Climatology*, 1-12

Ali Z, Hussain I, Faisal M, Shad MY, Elashkar EE, Gani S (2020a) An ensemble procedure for pattern recognition of regional drought. Int J Climatol 40(1):94–114

Ali Z, Hussain I, Faisal M, Khan DM, Niaz R, Elashkar EE, Shoukry AM (2020b) Propagation of the multi-scalar aggregative standardized precipitation temperature index and its application. Water Resour Manag 34(2):699–714

Ali Z, Hussain I, Grzegorczyk MA, Ni G, Faisal M, Qamar S et al (2020c) Bayesian network based procedure for regional drought monitoring: the seasonally combinative regional drought Indicator. J Environ Manag 276:111296

Anderson MC, Hain C, Wardlow B, Pimstein A, Mecikalski JR, Kustas WP (2011) Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States. J Clim 24(8):2025–2044

Babovic V (2005) Data mining in hydrology. Hydrol Process 19(7): 1511–1515

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 RA, Zou PX, 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

Byun HR, Wilhite DA (1999) Objective quantification of drought severity and duration. J Clim 12(9):2747–2756

Catenacci M, Giupponi C (2009) Potentials of Bayesian networks to deal with uncertainty in climate change adaptation policies. CMCC Res Paper 70

Cheng, J., & Greiner, R. (1999, July). Comparing Bayesian network classifiers. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (pp. 101-108). Morgan Kaufmann Publishers Inc.

De Martonne E (1926) L'indice d'aridité. Bulletin de l'Association de géographes français 3(9):3–5

Dracup JA, Lee KS, Paulson EG Jr (1980) On the definition of droughts. Water Resour Res 16(2):297–302

Erhardt, T. M., & Czado, C. (2015). Standardized drought indices: a novel uni-and multivariate approach.. *Roy. Stat. Soc.: Series C* (Applied Statistics), 67 (3) (April 2018), pp. 643–664

Farokhnia A, Morid S, Byun HR (2011) Application of global SST and SLP data for drought forecasting on Tehran plain using data mining and ANFIS techniques. Theor Appl Climatol 104(1–2):71–81

Grzegorczyk M (2010) An Introduction to Gaussian Bayesian Networks. In: An introduction to Gaussian Bayesian networks, In *Systems Biology in Drug Discovery and Development* (pp. 121–147). Humana Press, Totowa, NJ

Guttman NB (1998) Comparing the palmer drought index and the standardized precipitation index1. JAWRA J Am Water Resour Assoc 34(1):113–121

Guttman NB (1999) Accepting the standardized precipitation index: a calculation algorithm 1. JAWRA Journal of the American Water Resources Association 35(2):311–322

Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier

Hao Z, AghaKouchak A (2013) Multivariate standardized drought index: a parametric multi-index model. Adv Water Resour 57:12–18

Hao Z, Singh VP, Xia Y (2018) Seasonal drought prediction: advances, challenges, and future prospects. Rev Geophys 56(1):108–141

Hayes MJ, Wilhelmi OV, Knutson CL (2004) Reducing drought risk: bridging theory and practice. Nat Hazards Rev 5(2):106–113

Hussain, A., Zulqarnain, M., & Hussain, J. (2010). Catastrophes in the South Punjab due to climate change and the role of PIDEANS. *Center for Environmental Economics and Climate Change (CEECC), Islamabad*

Keyantash JA, Dracup JA (2004) An aggregate drought index: assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage. Water Resour Res 40(9)

Kiem, A. S., Johnson, F., Westra, S., van Dijk, A., Evans, J. P., O'Donnell, A., . & Jakob, D. (2016). Natural hazards in Australia: droughts. Clim Chang, 139(1), 37–54

Kim S, Parhi P, Jun H, Lee J (2017) Evaluation of drought severity with a Bayesian network analysis of multiple drought indices. J Water Resour Plan Manag 144(1):05017016

Kogan, F. N. (2000). Contribution of remote sensing to drought early warning. Early warning systems for drought preparedness and drought management, 75–87

Lee J, Henning R, Cherniack M (2019) Correction workers' burnout and outcomes: a Bayesian network approach. Int J Environ Res Public Health 16(2):282

Liu Z, Liu Y, Shan H, Cai B, Huang Q (2015) A fault diagnosis methodology for gear pump based on EEMD and Bayesian network. PLoS One 10(5):e0125703

Luo L, Wood EF, Pan M (2007) Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. J Geophys Res-Atmos 112(D10)

Mayfield HJ, Bertone E, Smith C, Sahin O (2020) Use of a structure aware discretisation algorithm for Bayesian networks applied to water quality predictions. Math Comput Simul 175:192–201

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, No. 22, pp. 179-183). Boston, MA: American Meteorological Society

Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391(1–2):202–216

Mishra AK, Singh VP (2011) Drought modeling–a review. J Hydrol 403(1–2):157–175

Mishra S, Saravanan C, Dwivedi VK (2015) Study of time series data mining for the real time hydrological forecasting: a review. Int J Comput Appl 117(23):6–17

Moglia M, Alexander KS, 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

Morid S, Smakhtin V, Moghaddasi M (2006) Comparison of seven meteorological indices for drought monitoring in Iran. Int J Clim 26(7): 971–985

Palmer, W. C. (1965). Meteorological drought, research paper no. 45. *US Weather Bureau, Washington, DC*, *58*

Salma S, Rehman S, Shah MA (2012) Rainfall trends in different climate zones of Pakistan. Pakistan J Meteorol 9(17)

Schittkowski K (2002) EASY-FIT: a software system for data fitting in dynamical systems. Struct Multidiscip Optim 23(2):153–169

Spiess, A. N. (2014). Propagate: propagation of uncertainty. R package version 1.0-http://CRAN.R-project.org/package=propagate (last accessed 27 April 2020)

Stagge JH, Tallaksen LM, Gudmundsson L, Van Loon AF, 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. . Integrated Drought Management Tools and Guidelines Series. World Meteorological Organization, Geneva

Tadesse T, Wilhite DA, Harms SK, Hayes MJ, Goddard S (2004) Drought monitoring using data mining techniques: a case study for Nebraska, USA. Nat Hazards 33(1):137–159

Tadesse, T., Wardlow, B., & Hayes, M. J. (2009). The application of data mining for drought monitoring and prediction. In *Data mining applications for empowering knowledge societies* (pp. 278-289). IGI Global

Tsakiris G, Vangelis H (2004) Towards a drought watch system based on spatial SPI. Water Resour Manag 18(1):1–12

Tsakiris G, Vangelis HJEW (2005) Establishing a drought index incorporating evapotranspiration. Eur Water 9(10):3–11

Vicente-Serrano SM, Beguería S, López-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J Clim 23(7):1696–1718

Wilhite DA, Glantz MH (1985) Understanding: the drought phenomenon: the role of definitions. Water Int 10(3):111–120

Wilhite DA, Sivakumar MV, Pulwarty R (2014) Managing drought risk in a changing climate: the role of national drought policy. Weather Clim Extremes 3:4–13

World Meteorological Organization, (2012): Standardized Precipitation Index User Guide, (WMO-No. 1090), World Meteorological Organization, Geneva, Switzerland