JOURNAL OF RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY 2020(03)

Available online at http://journals.researchub.org

Identifying Associations between Local Drought and Global Sea Surface Temperature

Mohammad Mohamadkhani[1*], Aydin Shishegaran[2]

[1]Master of science, school of progress engineering, Iran university of science and technology
[2]PhD candidate, water and environmental engineering, school of civil engineering, Iran university of science and technology

ARTICLE INFO

Article history:

Received 05 Jan 2020
Received in revised form 14 May 2020
Accepted 20 July 2020

Keywords:
Local Drought; Climate phenomena; Data mining; Association rules mining; Urmia.

ABSTRACT

It is clear that droughts have a fundamental impact on many different elements of society. To reduce the drought-related losses, it is necessary to give decision makers visibility into relationships of oceanic-atmospheric parameters that cause drought. The main target of this paper is to show the efficiency of data mining methods (especially association rules mining) for Identifying the associations between local droughts and large scale oceanic-atmospheric climate phenomena such as Sea Surface Temperature (SST). In this paper, association rules mining technique was offered to discover affiliation between drought of Urmia synoptic station (located in Iran) and de-trend SSTs of the Black, Mediterranean and Red Seas. To examine the accuracy of the rules, the confidence measures of the rules were calculated and compared for different considering lag times. The computed measures confirm reliable performance of the association rules mining method to monitor local drought so that the confidence between the monthly Standardized Precipitation Index (SPI) values and the de-trend SST of seas is higher than 87 percent.

1. INTRODUCTION

The previous researches have shown that the understanding the relationship between hydro climatic variables and changes in large-scale ocean-atmospheric patterns, can lead to improvements in long-term forecasting of hydro climatic variables such as precipitation and drought. Nicholson and Kim (1997) [10] find a strong connection between the El Nino Southern Oscillation phenomenon and regional precipitation over Africa. It was shown that the El Nino and La Nina cause respectively dry and wet years. Nazemosadat (1998) [8] indicated that there is a significant connection between the variation of Persian Gulf Sea Surface Temperature (SST) and the precipitation of southwestern part of Iran. It was emphasized that when the Persian Gulf sea SSTs are over than normal, winter droughts are expected in the area. Nazemosadat and Cordery (2000) [9] have also investigated the impacts of the El Nino-Southern Oscillation (ENSO) on rainfall variability over several regions of Iran. The results showed a low regression between Southern Oscillation Index (SOI) and precipitation values over whole Iran. Rucong et al. (2001) [17] studied the lagged relation between SST data and summer monsoon in Mid-Eastern China. The study indicated that the SST variations of the Pacific Ocean have a non-negligible association with summer monsoon in the study area. Harshburger et al. (2002) [6] investigated 40 years of precipitation data set from Idaho and its correlation with the variations of SST of the Pacific Ocean. They found that the winter precipitation in the northern mountains has the reverse lagged association with SST data in the Pacific Ocean. Rowell (2003) [16] showed that the anomalies of Mediterranean SSTs have a considerable impact on wet season precipitation over the Sahel in Africa. Iseri et al. (2005) used several climate indices, including the SOI as well as SST, to forecast monthly precipitation of Fukuoka, Japan. Araghinejad et al. (2006) [2] examined the effects of SOI and North Atlantic Oscillation (NAO), on autumn-winter and spring stream flow of the Zayandeh-rud River in Isfahan, Iran and showed that the higher SOI values (Jan-Oct) correspond to drier autumn and winter seasons and higher NAO values correspond to the wetter spring. Revadekar and Kulkarni (2008) [14] studied the impact of the ENSO on India's winter extreme precipitation values. Results indicated that ENSO index can be used to estimate the frequency and intensity of extreme precipitation, 4 to 6 months in advance. Rahimikhoob (2010) [13] studied the relationship between the maximum monthly precipitation of Ilam City in Iran and SSTs of the Persian Gulf and the Red Sea. Results emphasized that the maximum monthly precipitation with one-month time lag can be forecasted using the SSTs of the Persian Gulf and the Red Sea. Rezaebanafsheh et al. (2011) [15] investigated the relation between the winter and autumn precipitation values with prior season anomalies of the Mediterranean SST on several stations in the west of Iran and showed that the cooler condition of the sea in autumn could lead to wetter winter. Meidani and Araghinejad (2014) [7] applied the singular value decomposition (SVD) method for developing temporal expansion series of Mediterranean SSTs according to the precipitation and stream flow variability of southwest of Iran. The results showed the capability of SSTs temporal expansion series as the predictors of precipitation with regard to the well-known climate indices.

Data mining targets to develop semi-automatic approaches for detecting unforeseen and significant relationships from large datasets, which cannot be discovered manually. With this method, it is possible to discover cause-effect relationships, specify which variables have the strongest relationships to the problems

Corresponding author: m_mohamadkhani@vu.iust.ac.ir
DOI: https://doi.org/10.24200/jrset.vol8iss03pp1-4

of interest, and develop models that predict future outcomes. Data mining can be divided into four major methods or research areas [11]: 1) classification, 2) clustering, 3) anomaly detection, and 4) association analysis. The first two methods can be used for grouping of data into classes/clusters. The third method is used for identification of anomalous data. The fourth method is used for identifying the relationships between various variables. This study aims to develop association analysis techniques to discover the interesting and non-trivial associations of hydrologic and climatic variables.

Association rules have been extensively used in various fields of science but the use of this method in hydro-environmental issue, is so limited. As such a few studies, Tadesse et al. (2004) [19] discovered association rules between large-scale ocean-atmospheric phenomena and climatic variables for drought monitoring in Nebraska. Dhanya and Kumar (2009) [4] investigated precipitation time series to determine association rules for floods and droughts in India. Dadaser-Celik et al. (2012) used association rules to explore the relationships between stream flow and climatic variables in the Kızılırmak River Basin in Turkey.

In this paper it is aimed to use association rules mining method to identify complex relationships involving atmospheric and oceanic variables like SSTs that potentially cause droughts over selected stations in Iran.

2. MATERIALS AND METHODS

2.1. Study Area and Data

The study area is Urmia city (37.33°N, 45.04°E) from northwest of Iran (Fig. 1). A 62-year monthly dataset (1955-2016) were used to determine the association between monthly de-trend SSTs of Black, Mediterranean and Red seas and monthly Standardized Precipitation Index (SPI) of study area.

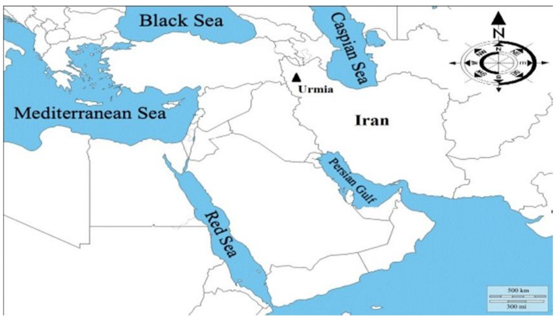

Fig. 1. Location map [11]

The mean, maximum and minimum monthly SST data of the Black Sea, Mediterranean Sea and Red Sea for the study period (1955-2016) are presented in Table 1.

Table 1. Statistics of SST monthly observed data of the Black, Mediterranean and Red Seas

Sea	Sea Surface Temperature (SST) (˚C)			
	Mean	Maximum	Minimum	Std.
Black Sea	12.012	25.836	-1.587	7.431
Mediterranean Sea	19.293	28.125	10.817	4.594
Red Sea	27.970	32.619	17.742	3.428

2.2. Drought indices

Standardized Precipitation Index (SPI) is one of the most practical of drought indices which provide information about drought using climate elements such as precipitation and temperature. The drought indices such as monthly SPI should be used when making decisions about the management of water resources to mitigate the effects of drought [12, 18]. SPI is obtained through normalizing, by subtracting the amount of precipitation in station i (P_i) from the mean amount of precipitation (M) and dividing the difference by standard deviation (SD) and is explained by the following equation:

$$SPI = \frac{P_i - M}{\mathrm{SD}} \qquad (1)$$

Table 2 gives the severity classes according to this method. The drought categories given in this table were used as the determining class labels in forming the association rules mining method.

Table 2. The monthly SPI drought severity categories and determining class labels [19]

Monthly SPI values	Drought severity categories	Symbol
≥ 2	Extremely Wet	EW
1.50 ~ 2	Severely Wet	SW
1 ~ 1.50	Moderately Wet	MW
-1 ~ 1	Normal	N
-1 ~ -1.50	Moderately Dry	MD
-1.50 ~ -2	Severely Dry	SD
≤-2	Extremely Dry	ED

2.3. Association Rule

Data Mining is the discovery of hidden information found in databases and can be considered as a step in the knowledge discovery process (Chen et al. 1996). Data mining functions include clustering, classification, prediction, and link analysis (associations). One of the most important data mining applications is mining of association rules. Association rules, first introduced in 1993 [1], are applied to extract relationships between different items in a database. These relationships are not based on inherent properties of the data themselves (as with functional dependencies), but rather based on co-occurrence of the data items. In the association rule mining, the aim is to extract any rules of the form X→Y that seem to occur in the data with frequency above a given threshold. Here X and Y are events of a certain type, connected by the rule 'if X occurs then Y occurs'. The rules can be extended into the form X_1, X_2, …, X_N→Y, which can be interpreted as 'if X_1, X_2, …, X_N all occur, then B will occur'. Interest measures are used to identify and characterize interesting associations. Support and confidence are two measures that are widely used (Han et al. 2011). Support defines the probability of occurrence of the set of {X, Y} in an itemset (Eq.2). The confidence of the rule (Eq.3) shows how often itemsets containing X also contain Y [5].

$$Support\ (X \to Y) = P(X \cup Y) \qquad (2)$$

$$Confidence\ (X \to Y) = P(X|Y) = P(X \cup Y)/P(X) \qquad (3)$$

A commonly used association discovery algorithm is the Apriori algorithm [1] which aims to discover frequent item sets (patterns) that satisfy the user-defined support threshold (see Fig. 2). The basic principle of the Apriori algorithm can be summarized as "if an item set is frequent, then all of its subsets must also be frequent". Based on this principle, Apriori algorithm discovers size k+1 candidate item sets using size k frequent item sets. Then, the algorithm checks if the candidate size k+1 item sets satisfy the support threshold.

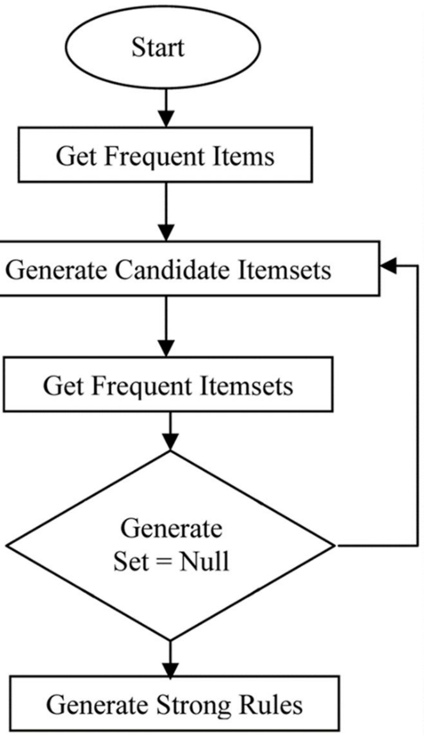

Fig. 2. Flowchart of Apriori algorithm (Sharma and Om, 2014)

In Apriori algorithm, size k subsets are extended by one item in each iteration (a step known as candidate generation), and candidate size k+1 item sets are tested against the data. If it is true, the size k+1 candidate item set is called frequent item set and the algorithm continues to find size k+2 candidate (superset) item sets. This iteration (process) continues until finding all possible superset item sets. The algorithm outputs all size frequent patterns, and finally the directions of the rules are determined by evaluating the confidence measures [3]. In this paper for discovering association rules and modeling, the Weka program by Java code has been used.

3. RESULTS AND DISCUSSION

In association analysis, the data should be in discrete format. In this study, the data was discretized by using their statistical properties (i.e. mean (μ) and standard deviation (σ)). First, the de-trend SST data were arbitrary divided into five groups: "very low (VL)" if they were "smaller than μ-σ", "low (L)" if they were between "μ-σ" and "μ-0.5σ", "medium (M)" if they were between "μ-0.5σ" and "μ+0.5σ", "high (H)" if they were between "μ+0.5σ" and "μ+σ" and "very high (VH)" if they were "higher than μ+σ" (Table 3).

Table 3. De-trend SST groups of the Mediterranean, Red and Black Seas

	VH	H	M	L	VL
Mediterranean sea	> 2.471	(1.237) – (2.471)	(-1.233) – (1.237)	(-2.467) – (-1.233)	< -2.467
Red sea	> 1.972	(0.987) – (1.972)	(-0.983) – (0.987)	(-1.968) – (-0.983)	< -1.968
Black sea	> 4.027	(2.014) – (4.027)	(-2.012) – (2.014)	(-4.025) – (-2.012)	< -4.025

Thereafter, association rules were extracted out among de-trend SSTs and monthly SPI of Urmia synoptic station considering different lags. After generating the rules, the confidence measure of the association rules for Urmia synoptic stations, were calculated and presented in Table 4.

Table 4. Confidence measures for some sample Association Rules between de-trend SST groups of the Black, Mediterranean and Red Seas and monthly SPI groups of Urmia synoptic station at different lags

	Determinant (if … occur)	Result (then … occur)	Confidence
Lag = 0	Black-VL; Mediterranean-L; Red-VL	SPI-MD	0.87
Lag = 1	Black-L; Mediterranean-L; Red-VL	SPI-SD	0.76
Lag = 2	Black-M; Mediterranean-VH; Red-M	SPI-SD	0.68
Lag = 3	Black-H; Mediterranean-VH; Red-H	SPI-SD	0.70
Lag = 4	Black-H; Mediterranean-VH; Red-VH	SPI-MD	0.77
Lag = 5	Black-H; Mediterranean-H; Red-M	SPI-MD	0.73
Lag = 6	Black-H; Mediterranean-H; Red-L	SPI-MD	0.68
Lag = 9	Black-VL; Mediterranean-VL; Red-VL	SPI-SD	0.66
Lag = 12	Black-VL; Mediterranean-L; Red-L	SPI-SD	0.61

According to the presented results in Table 4, de-trend SSTs of the Black, Mediterranean and Red Seas are effective on drought of Urmia synoptic station but the effectiveness is varying at different lag times. The highest confidence which is calculated between the most effective de-trend SST groups of the Black, Mediterranean and Red Seas and the monthly SPI groups of Urmia synoptic station is 87 percent. This means if simultaneously the most effective de-trend SST groups of the Black, Mediterranean and Red Seas are respectively VL, L and VL in that case, the monthly SPI value will be Moderately Dry (MD) with 79% minimum confidence in the same month (lag=0) and if simultaneously the most effective de-trend SST groups of the Black, Mediterranean and Red Seas are respectively L, L and VL in that case, the monthly SPI value will be Severely Dry (SD) after one month with 76% minimum confidence. Also if simultaneously the most effective de-trend SST groups of the Black, Mediterranean and Red Seas are respectively H, VH and VH in that case, the monthly SPI value will be

Moderately Dry (MD) after four months with 77% minimum confidence.

There are four main advantageous of data mining as compared to other technique especially the traditional correlation method. i) Data mining can provide high flexibility in time series analysis to discover the associations between the several parameters that occur together or with time lags. ii) Rules and models that are offered by data mining are easy to understand. iii) Using data mining techniques it will be possible for a researcher to have a wide viewing of the relations between the different parameters. iv) Data mining can be used when large amounts of data including several variables are analyzed as a robust tool for monitoring extreme events such as drought.

4. CONCLUSIONS

The previous researches have shown that the oceanic-atmospheric climate phenomena such as SST are effective predictors of the long-term variation of precipitation in many parts of the world, but in regions nearby the seas, the priority could be on using the sea surface temperatures of the near seas. Since Iran is surrounded by seas on North, South and West, in this study the use of association rule mining techniques was introduced to investigate the associations between the de-trend SSTs of the Black, Mediterranean and Red Sea and the drought of a station in the northwest of Iran. To evaluate the accuracy of rules, the confidence measure was calculated and compared for different conditions. The results showed the ability of SST of around seas to reconstruct the SPI groups in the northwest of Iran and also the ability of data mining tools to forecast the extreme events such as drought. Also, it can be suggested to use other ocean-atmospheric climate phenomena (e.g., SOI and NAO) as predictors to predict drought of stations through the proposed method.

REFERENCES

[1] Agrawal, R., & Ramakrishnan, S. (1994). Fast algorithms for mining association rules, In Proc. 20th int. *conf. very large data bases, VLDB, 1215,* 487-499.

[2] Araghinejad, S., Burn, D.H., & Karamouz, M. (2006). Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors. *Water Resources Research, 42*(3), W03431.

[3] Dadaser-Celik, F., Celik, M., & Dokuz, A.S. (2012). Associations between stream flow and climatic variables at Kizilirmak river basin in Turkey. *Global NEST Journal, 14*(3), 354-361.

[4] Dhanya, C.T., & Kumar, D. (2009). Data mining for evolution of association rules for droughts and floods in India using climate inputs. *Journal of Geophysical Research, 114*, D02102.

[5] Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers*, San Francisco, California, USA, 740.

[6] Harshburger, B., Ye, H., & Dzialoski, J. (2002). Observational evidence of the influence of Pacific SSTs on winter precipitation and spring stream discharge in Idaho. *Journal of Hydrology, 264*(1–4), 157-169.

[7] Meidani, E., & Araghinejad, S. (2014). Long-lead streamflow forecasting in the Southwest of Iran by sea surface temperature of the Mediterranean Sea. *Journal of Hydrologic Engineering, 19*(8), 05014005.

[8] Nazemosadat, M.J. (1998). The Persian Gulf sea surface temperature as a drought diagnostic for southern parts of Iran. *Drought News Network, 10*, 12-14.

[9] Nazemosadat, M.J., & Cordery, I. (2000), On the relationships between ENSO and autumn rainfall in Iran. *International Journal of Climatology, 20*(1), 47-62.

[10] Nicholson, S.E., & Kim, J. (1997), The relationship of the El Nino-Southern Oscillation to African rainfall. *International Journal of Climatology, 17*(2), 117-135.

[11] Nourani, V., Sattari, M.T., & Molajou, A. (2017). Threshold-Based Hybrid Data Mining Method for Long-Term Maximum Precipitation Forecasting. *Water Resources Management, 31*(9), 2658-2645.

[12] Nourani, V., & Molajou, A. (2017). *Application of a hybrid association rules/decision tree model for drought monitoring*. Global and Planetary Change, DOI: 10.1016/j.gloplacha.2017.10.008.

[13] Rahimikhoob, A. (2010). Forecasting of maximum monthly precipitation of Ilam using data mining techniques. *Iranian Journal of Soil and Water Research, 42*(1), 1-7. (In Persian)

[14] Revadekar, J.V., & Kulkarni, A. (2008). The El Nino-Southern Oscillation and winter precipitation extremes over India. *International Journal of Climatology, 28*(11), 1445-1452.

[15] Rezaebanafsheh, M., Jahanbakhsh, S., Bayati, M., & Zeynali, B. (2011). Forecasting autumn and winter precipitation of west of Iran applying Mediterranean SSTs in summer and autumn. *Physical Geography Research Quarterly, 74*, 47-62.

[16] Rowell, D.P. (2003). The Impact of Mediterranean SSTs on the Sahelian Rainfall Season. *Journal of Climate, 16*(5), 849-862.

[17] Rucong Y., Minghua Z., Yongqiang Y., & Yimin L. (2001). Summer monsoon rainfalls over Mid-Eastern China lagged correlated with global SSTs. *Advances in Atmospheric Sciences, 18*(2), 179–196.

[18] Spinoni, J., Naumann, G., & Vogt, J. (2017). Pan-European seasonal trends and recent changes of drought frequency and severity. *Global and Planetary Change, 148*, 113-130.

[19] Tadesse, T., Wilhite, D.A., Harms, S.K., Hayes, M.J., & Goddard, S. (2004). Drought monitoring using data mining techniques: a case study for Nebraska, USA. *Natural Hazards, 33*(1), 137-159.