University of Tabriz

Journal of Civil and Environmental Engineering
Volume 54 (2024), Issue 1 (Spring), 142-151

EXTENDED ABSTRACT

Investigating the Performance of Kstar and GPR Algorithms in Modeling RDI Meteorological Drought Index (Case Study: East of Urmia Lake Basin)

Marzie Sadeghian Aghkandi [a], Hossein Rezaie [a,*], Keivan Khalili [a], Farshad Ahmadi [b]

[a] *Department of Water Science and Engineering, Urmia University, Iran*
[b] *Department of Hydrology and Water Resources, Shahid Chamran University of Ahvaz, Iran*

Received: 13 February 2023**; Review:** 13 April 2023; **Accepted:** 17 May 2023

Keywords:
Kstar algorithm, RDI drought monitoring, Gaussian process regression, east of Lake Urmia.

1. Introduction

Drought is a severe hydrological event that can cause serious problems in human life. In this context, it can have adverse effects on water supply and quality, public health, agricultural productivity, land degradation, desertification, famine, etc. (Madadgar and Moradkhani, 2014; Li et al., 2020). In a general classification, drought events are classified into four different categories, meteorological, hydrological, agricultural and socio-economic droughts (Wilhite and Glantz, 1985; Khadr, 2016). Some of the well-known and common meteorological drought indices in drought monitoring include Palmer Drought Severity Index (PDSI), Drought Identification Index (RDI), Standardized Precipitation Index (SPI) and Standard Precipitation Evapotranspiration Index.

Considering the significant role of evaporation and transpiration in the water balance, it is necessary to consider its effect when studying drought in a particular region. Based on this, RDI, which includes both precipitation and evaporation and transpiration, can be considered as a reliable indicator for drought monitoring (Moeinifar et al., 1400). According to importance of drought as a natural phenomenon in hydrological and meteorological studies, its monitoring and forecasting with a suitable approach can be important. The main role of Drought prediction in risk management, reducing the effects of drought on existing water resources and their optimal use, the possibility of rational decision-making by decision makers to minimize the damages caused by drought, as well as planning and managing resource projects. It has water. (Khadr et al., 2016; Madrigal et al., 2018; Beyaztas and Yaseen, 2019). Among the models worked till date, single Kstar and GPR models are the newest models for drought prediction

2. Methodology

2.1. Data and the study area

It is located in the east of Urmia lake basin with an area of 1084.38 square kilometers between the north latitude of minimum 36 and maximum 38 degrees and east longitude of minimum 46 and maximum 48 degrees in the eastern-most part of the catchment area of Lake Urmia. In this range, the incoming flows are; Surface currents, rainfall and snow and permanent channels. The average annual rainfall in the plains of this area is calculated to be 330 mm. The meteorological information used in this research is the monthly rainfall and potential evapotranspiration data of Ahar, Tabriz, Jolfa and Maragheh synoptic stations in the years 1955 to 2019 to estimate the RDI index.

* Corresponding Author
E-mail addresses: m.sadeghian@urmia.ac.ir (Marzie Sadeghian Aghkandi), h.rezaie@urmia.ac.ir (Hossein Rezaie), k.khalili@urmia.ac.ir (Keivan Khalili), f.ahmadi@scu.ac.ir (Farshad Ahmadi).

2.2. RDI drought recognition index

The RDI index is more popular than other indices due to the combined consideration of precipitation and potential evaporation and transpiration, and in many studies, it is the basis for comparing the severity of drought and estimating its potential damages at the national level. The RDI index with the same structure as SPI has special capabilities for evaluating agricultural and hydrological drought due to the use of potential evaporation and transpiration in its structure (Shkohi, 2013). In the studied area, the performance comparison of Kstar and GPR models on the RDI index has not been done so far. Therefore, according to the capabilities of these two models, in this research, meteorological drought monitoring in the eastern basin of Lake Urmia in a period of 65 years, and the analysis of the results of these two models in the RDI index were considered.

2.3. Kstar algorithm

The Kstar algorithm can be defined as a method for cluster analysis whose main goal is to divide n observations into K clusters so that each observation belongs to the cluster with the closest average (Ekmekcioğlu et al, 2020). Kstar algorithm can be described as an example-based learning method that uses entropy theory as a distance measure. This method maximizes the possibility of obtaining information from existing data by providing a consistent approach for managing real valuable features, symbolic features and missing values. In the Kstar algorithm, the distance from one sample to another can be described as the complexity of converting one sample to another (Granata et al, 2019). As mentioned, Kstar regression uses an entropy distance function. Entropic distance is used to obtain samples that are most similar to each other from the data set. Consider.

2.4. Gaussian Process Regression (GPR)

In the theory of statistics and probability, a Gaussian process is a statistical model in which observations occur in a continuous range. In fact, the Gaussian process is a random process that is known as a sequence of random variables, and the random variables are usually calculated as; For example, time is ordered and each finite subset of these variables has a multivariate Gaussian (normal) distribution. In a Gaussian process, every point of the input space is a random variable with a normal distribution. In addition, each finite set of these random variables has a multivariate Gaussian distribution. Gaussian process distribution is the joint distribution of all these random variables (number and infinite). From the point of view of a machine learning algorithm, a Gaussian process measures the similarity between points (the same kernel function) to predict new points from the training data. Gaussian process is a set of random variables, a limited number of which are integrated with Gaussian distributions. Gaussian process is completely determined by its mean function m(x) and its covariance function k(x,x). This process is a natural generalization of the Gaussian distribution, whose mean and covariance are a vector and a matrix, respectively. Gaussian process regression models are based on the assumption that the set observations should carry information about each other. One of the important features of the Gaussian process is the presence of various covariance functions in it, which allows the researcher to make an appropriate selection among them. These models can specify distributions between functions with one or more input variables. When such a function defines the mean response in a regression model with Gaussian errors, matrix calculations can be used for inference. - got stuck; This problem is possible for data sets with more than a thousand samples

3. Results and discussion

3.1. Results

In this study, RDI drought index was calculated and compared for four selected synoptic stations in the east of Urmia lake basin in six, nine and 12 months using meteorological data of precipitation and potential evapotranspiration. The obtained results have high adequacy in showing drought. According to the results obtained from different time scales, the one-year RDI index (12 months) can better describe the drought that occurred. Also, drought monitoring was done in 10-year periods between different calculation scales. The calculation results of RDI (6M) and RDI (9M) are close to each other and have very little difference, while RDI (12M) has obtained different results. According to Ahmadi et al.'s research (1400), they examined the RDI index in the annual, monthly and regional scales to monitor the drought in Durward-Broujerd basin and showed that the annual scale shows the drought better. According to the obtained results, it can be concluded that in the years 1957 to 1964, 1998 to 2003, 2008 to 2010 and 2016 to 2018, mild to very severe drought occurred. In this study, WEKA software was used to develop Kstar and GPR algorithms, and its parameters were determined

by trial and error. In order to predict the monthly drought of the studied stations, 80% of the data (536 months) were considered for training and the remaining 20% for the test (230 months).

In checking the accuracy of the used algorithms, the results of the root mean square error, absolute mean error and Nash-Sutcliff evaluation criteria can be used. The numerical value of RMSE in a one-year time scale with the GPR algorithm is in the range of 0.31 to 0.39 and with the Kstar algorithm in the range of 0.32 to 0.51. Also, in the scales of 6 and 9 months, they modeled RDI drought index from 0.51 to 0.55 with GPR algorithm and from 0.52 to 0.57 with Kstar algorithm. According to the mentioned results, the GPR algorithm can be used with high accuracy in modeling. It can also be said that the one-year time scale shows the drought better and more realistically compared to the monthly scales.

4. Conclusions

This study was conducted with the aim of modeling RDI drought in three time scales including RDI-6m, RDI-9m and RDI-12m for four stations east of Urmia lake basin. To achieve this goal, two algorithms, Ksatr and GPR, were used to model the drought RDI index and their accuracy was evaluated. Examining the results of calculating the RDI index with the classification presented by Tsakiris et al., (2007) indicated that during the studied years, the eastern part of Lake Urmia was "relatively dry" and "extremely dry" in terms of drought. Dry"

The results showed that the individual Kstar and GPR models have the necessary efficiency and high performance in estimating the monthly RDI index. As shown in the obtained scatter diagrams, it was observed that the mentioned algorithms were able to eliminate the scatter of the data and provide acceptable results. Then, to compare and select a stronger algorithm, the evaluation criteria of correlation coefficient, Nash-Sutcliffe coefficient, root mean square error and absolute mean error were used. The results obtained from the numerical values of the calculation criteria showed that in the one-year time scale (RDI-12m), the correlation coefficient was higher and the error criteria were lower. Also, among the two used algorithms, the GPR algorithm has been able to model the RDI index with a very small difference from Kstar.

5. References

Ahmadi F, Madah MA, "Development of hybrid method of wavelet-Kstar algorithm for forecasting monthly rainfall (case study: Ahvaz synoptic station)", Iran Water and Soil Research, 1400, 52 (2), 409-420 (In Persian).

Beyaztas U, Yaseen ZM, "Drought interval simulation using functional data analysis", Journal of Hydrology, 2019, 579, 124141.

Granata F, Di Nunno F, Gargano R, de Marinis G, "Equivalent discharge coefficient of side weirs in circular channel- a lazy machine learning approach", 2019, Water, 11 (11), 2406.

Khadr M, "Forecasting of meteorological drought using Hidden Markov Model (Case Study: The upper Blue Nile river basin, Ethiopia)", Ain Shams Eng Journa, 2016, 7 (1), 47-56.

Li Q, He P, He Y, Han X, Zeng T, Lu G, Wang H, "Investigation to the relation between meteorological drought and hydrological drought in the upper Shaying River Basin using wavelet analysis", 2020, Atmos. Res. 234, 104743.

Madadgar S, Moradkhani H, "Spatio-temporal drought forecasting within Bayesian networks", Journal of Hydrology, 2014, 512, 134-146.

Madrigal J, Solera A, Suárez-Almiñana S, Paredes-Arquiola J, Andreu J, Sánchez-Quispe ST, "Skillassessment of a seasonal forecast model to predict drought events for water resource systems", Journal of Hydrology, 2018, 564, 574-587.

Mckee TB, Doesken NJ, Kleist J, "The relationship of drought frequency and duration to time scales", AMS 8th Conference, Climatology, 1993, 179-184.

Moinifar S, Asadi MA, Melki-Nejad H, Talebi A, "Determining the appropriate statistical distribution for calculating the RDI index in arid regions (Case Study: central Iran)", Arid Scientific Journal, 1400, 11 (1), 105-121 (In persian).

Mirabbasi Najafabadi R, Ahmadi F, Ashuri M, Nazeri Tehrodi M, "Analysis of droughts in North-Eastern Iran using Joint Deficit Index (JDI)", Ecohydrology, 2016, 4 (2), 573-585 (In persian).

Palmer WC, "Meteorological Drought US Department of Commerce", Weather Bureau, Washington, DC, 1965.

Shokohi A, "Comparison of RDI and SPI indices for station-scale drought analysis based on agricultural drought (case study: Qazvin and Takestan)", Scientific Research Journal of Irrigation and Water Engineering, 1391, 9 (3), 111-122.

Tsakiris G, Pangalou D, Vangelis H, "Regional drought assessment based on the reconnaissance drought index (RDI)", Water Resource, Management, 2007, 21, 821-833.

Vicente-Serrano SM, Beguería S, López-Moreno JI, "A multi scalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index", Journal of Climatology, 2010, 23, 1696-1718.
Wilhite DA, Glantz MH, "Understanding: the drought phenomenon", 1985, the Role of Definitions Water Int. 10, 111-120.