A SURVEY ON THE RECENT DEVELOPMENTS IN THE AREA OF DROUGHT-FORECASTING

Article *in* International Journal of Pure and Applied Mathematics · May 2018

4 authors, including:

Poornima Selvaraj
SRM Institute of Science and Technology

Pushpalatha Marudappa
SRM Institute of Science and Technology

International Journal of Pure and Applied Mathematics
Volume 118 No. 22 2018, 307-312
ISSN: 1314-3395 (on-line version)
url: http://acadpubl.eu/hub
Special Issue
ijpam.eu

A SURVEY ON THE RECENT DEVELOPMENTS IN THE AREA OF DROUGHT-FORECASTING

[1]Vardaan Kishore, [2]Jaishankar Sujit Shankar, [3]S.Poornima, [4]M.Pushpalatha
[3]Assistant Professor, [4]Professor
[1,2,3,4]Department of Computer Science and Engineering
SRM Institute of Science and Technology
[1]naadvar@gmail.com, [2]sujitshankar1996@yahoo.co.in
[3]poornima.se@ktr.srmuniv.ac.in, [4]pushpalatha.m@ktr.srmuniv.ac.in

Abstract: Over the years, natural calamities like drought have taken a huge toll on human life and resources. With the advent of newer methods of prediction, the adverse effects of such natural calamities can be reduced to an extent by pre-planning and providing sufficient warnings to the people. Many of the previous prediction methods majorly used statistical methods such as ARIMA, but these methods were still lacking the accuracy to provide long-term forecasts, but with advances in the area of Machine Learning especially Artificial Neural Networks and the more bleeding edge versions of it like Deep Neural Networks, there seems to be a method to predict drought in the long term with a good accuracy which can help authorities better prepare and mitigate the losses to a huge extent. This paper compares and contrasts the various methods used in different studies over the years and their results.

Keywords: Drought, Artificial Neural Network, ARIMA, Deep Learning, SPI, SPEI.

1. Introduction

Precipitation is a field that is random in character and as such makes drought prediction a complex task. Drought materializes from a deficiency of rainfall over a period of time. Drought over short time scales (months) characterize meteorological drought whereas long term scales (years) showcase hydrological drought.

The Standardized Precipitation Index (SPI), was proposed [1] in order to help monitor relative wetness and dryness over multiple time scales. Short timescales imply that the SPI is closely related to soil moisture, while at longer timescales it indicates groundwater and reservoir storage. SPI can be used across regions with differing climatic conditions. This results from the quantification of observed precipitation as a selected probability distribution function that is modelled over raw precipitation data. SPI can be used over 1-36 month time scale and can be interpreted as the number of standard deviations by which the observed anomaly deviates from long term mean. Since, SPI is not conducive to climate change associated with evapotranspiration, the Standardized Precipitation-Evapotranspiration Index (SPEI) has also been proposed. Inclusion of SPEI is to ensure that the limited ability of SPI to capture the effect of increased temperatures is overcome. Other indices include Palmer Drought Severity Index (PDSI) and Multivariate Standardized Drought Index (MSDI). (AghaKouchak 2014) SPI cannot be applied directly to standard time series methods due to two hindrances. One is that precipitation is not usually correlated on the same time scale where drought occurs. The other is that any time correlations found in the SPI over long time scales are a result of cumulating precipitation on the selected time scales. [2]. A commonly used methods in Statistics is the Auto Regressive model (AR). Having a long precipitation time series may make it intuitive to apply AR model on it. The AR model would give its prediction as the extract of the seasonal cycle of precipitation. Drought prediction, however may not be feasible since it depends heavily on the departure from the seasonality of precipitation.

2. Literature Survey

Many types of methods have been proposed for drought forecasting over the years with varied results. All these studies can be broadly classified into two major categories.

A. Statistical Approaches

The GAHP – Gamma Highest Probability Method [3], overcomes the problem of extraction of just the seasonal cycle when implementing the AR Model. This limits any knowledge about large deviations that occur. The GAHP approach is focussed on forecasting the precipitation for a future month as the most probable value as described by the probability density function of the precipitation for that month, Since, it is a Gamma distribution, it requires the estimation of the parameters that best fit the frequency histogram of the observed precipitation of the month at hand. The predicted precipitation for the next month is the observed mode of the fitted distribution. The method again rests on assuming that consecutive precipitation events are not correlated, future events will be most probable and precipitation is related only to seasonality. The aforementioned assumptions are empirical and related to the characteristics of precipitation of the region under scrutiny. It is clear that the AR model estimates future precipitation as the mean value in approximation of the precipitation observed for that month. Mean value is not always the most probable value; in matters associated to precipitation, especially since precipitation does not fit the Gaussian Distribution of any month. Future discussions must involve predictions for longer time scales.

[4]cameup with the MDSI which is a multi variateversion of the SPI. It is able to composite multi-index drought information; precipitation and soil moisture. Drought severity and drought occurrence, both are predictable using this model. The standardized normal distribution it models is capable of employing prediction at different time-scales. Hao et.al (2014) hinted at the superiority of MSDi performing better in showing higher probability of droughts being predicted than SPI based models. Ensemble Streamflow Prediction (ESP) was used to predict the MDSI values where it was shown that the persistence based model required higher quality and accuracy of the initial conditions and couldn’t account for fluctuations in or missing data.

B. Neural Network Approaches

ANN have been proven to be suitable for complex time series forecasting. [5]. When the number of hidden layers become more than two, they bring with them non-convex optimization issues,[6] however, showed the resourcefulness of deep neural networks by proposing and unsupervised greedy layer-wise training for deep networks.

In [7] the study was conducted in 6 different provinces of Tehran. An Artificial Neural Network(ANN) was adopted for prediction of future droughts. Of the various indices present, Standard Precipitation Index(SPI) and the Effective Drought Index(EDI) were chosen. The study used 32 years of rainfall data for modelling the ANN and also for prediction purposes. The monthly rainfall data was converted to the chosen indices using the Drought Index Package (DIP) software. The Southern Oscillation Index (SOI) and the North Atlantic Oscillation (NAO) were also considered to be the extra parameters along with the drought indices as input to the ANN. This study used the Multi-Layer Perceptron (MLP) as the training algorithm for the ANN. The Index data was standardized to range between FMIN and FMAX (FMIN=0.1 and FMAX<1) using equation (1), where Xu and Xn represent original and standardized values and the fact_min and fact_max indicate the maximum and the minimum and the maximum values in the original dataset.

$$X_n = FMIN + \frac{(X_u - fact\_min)}{(fact\_\max - fact\_\min)} * (FMAX - FMIN) \quad (1)$$

Various architectures of the ANN were tested with R^2, RMSE, MAE as the error measures for validation. After training all the architectures and comparing the results, the 5-6-1 model with 5 input neurons, 6 hidden neurons and 1 output neuron was selected to be the best suited for the EDI. This model produced an R^2error of 0.84 and 0.79 on training and validation for the EDI index. A similar model with a lot more past information was found to be the best for the SPI giving R^2 values in the range 0.66 -0.80. The study also showed that the extra parameters like the SOI and the NAO were found to have almost no impact on the performance of the network. It was concluded that ANN was best suited to medium-term forecast (6months) and EDI outperformed the SPI because of its sluggish nature with no immediate fluctuation which lead to better results. In another study [8], ANN was again used for forecasting droughts in the region of Pakistan. This study used the Standardized Precipitation Evapotranspiration Index (SPEI) for forecasting. The ANN model was selected because of its superiority in modelling hydrological data. The data used for calculating the SPEI was from 1975 to 2012, the SPEI was calculated for 4 different timescales (1, 3, 6, 12 months) respectively. The SPEI index was calculated by taking the difference between the Precipitation data and the Potential Evapotranspiration, this difference was later fit to various probability distributions like the Gamma, Generalized Extreme Valued Distribution, Log - Logistic Distribution, Generalized Pareto Distribution and standardized to obtain the final SPEI values for the given data. The ANN model was selected to be 30-8-1 in input, hidden and output layers respectively, after experimenting with several other architectures. The momentum for the network was set to 0.5 and it was trained for 10000 epochs and each epoch consisted of a vector of 30 previous values as the input, it was trained on 80% of the data and validated on the rest of the 20%. Error measures like the MAE, Correlation Coefficient (R) and the RMSE were used to evaluate the model. The study gave the following results measured on the Correlation Coefficient, 0.887 to 0.987 for SPEI-1, 0.876 to 0.994 for SPEI-3, 0.876 to 0.994 SPEI-6 and 0.780 to 0.970 for SPEI-12.

Another study [9] used the similar approach of ANN for prediction purpose. This study was based in the Sri Lankan region. The data was collected from 13 meteorological stations over a period of 100 years from 1870 to 1980, covering both dry and wet zones. The SPI index was used by fitting the precipitation data to the frequency distribution and fitting this to a probability distribution. This was performed separately for each month. The dataset was split into training set which consisted of data from 1870 to 1950 and the training set consisted of data from 1951 to 1980. After trying and testing various models, the following model of 30-8-1, with 30 input nodes, 8 hidden nodes, 1 output node was selected as the best.

This model was trained over 500 epochs using the Levenberg-Marquardt optimization weight updating algorithm.. The momentum was set to 0.9 and the weight gradient was set to a lowest of 10^{-6}. The error measure used to evaluate the model were MAE, RMSE, R and percentage error (PE). The error rates for the SPI with smaller windows were high and when the window size was increased the error rate reduced considerably. The results are presented in table 1 on the RMSE error:

Table 1. RMSE values of different time scales of SPI index

SPI	RMSE
SPI -6	0.005
SPI-4	0.028
SPI-3	0.074
SPI-2	0.193
SPI-1	0.444

The reason for the decreasing error rates with increasing time window was that as the time periods increased it the conversion to the normal distribution left out sudden peaks leaving a smooth curve that was easier to predict by the ANN. When the trained model was used for forecasting the future values, it was found that the best results were obtained for 1-month ahead forecasts, and when the time period was increased to 4-5 months, the accuracy became low.

Taking a different approach to drought prediction [10] used ANN to predict drought in Yazd city in Iran. Here the variables like precipitation, minimum temperature, maximum temperature, evaporation, wind speed and wind direction were used directly instead of converting them to a standardized scaled like the SPI or SPEI. The data used was from April 1953 to December 2005. For training the ANN model the data between 1975 to 2001 was used and the data from 2002 to 2007 was used ot test the model. The RMSE and R were chosen as the error measure. The lead time was selected to be 12 months in this study. The two ANN used were the BP Neural Network and the Time Lag Recurrent Network(TLDR), the tangent hyperbolic function was used in the hidden layers and the sigmoid function was used in the output layer. After running correlation experiments on the different variables, it was found that maximum temperature had the highest impact on the precipitation data, thus it was selected to help in prediction purposes. The TLDR gave the best results with a R of 0.95 and RMSE of 0.05, this result was obtained after the model converged after training for over 22000 iterations and the prediction time was 12months ahead.

In another approach[11], this study touched upon the use of state of the art Deep Learning method for short-term forecasting. The study was conducted over 4 regions in the eastern part of China, the data collected was over the period of 1958-2006. This data was then used to calculate the SPI-9, SPI-3, SPI-6, SPI-12. The training set consisted of data from 1958-1999 and the rest was used for testing. The new approach was the use of Restricted Boltzmann Machine which consists of two layers, the input and output, when these RBM's are stacked together to create many hidden layers, they are called as Deep Belief Networks. The Study used the following steps to create and train and train the model:

1. Compute the different time scale SPI series.
2. Normalize the SPI series.
3. Determine the optimum number of input, hidden and output nodes required to attain efficient performance by trial and error method.

After going through the above steps, the model was set to the following architecture 9-5-10-1, with 9 input nodes, 5 nodes in the first hidden layer, 10 nodes in the second hidden layer and 1 node in the output layer. The error measures used to reach the above architecture were MAE, RMSE. To show the superior performance of DBN's the same data was fed to a simple ANN trained using the Back Propagation Algorithm.

The results for one of the stations for SPI-3 and SPI-6 can be summarized in the table 2 below:

Table 2. Comparison of RMSE and MAE between different models used for prediction

Station	Model	SPI-3	SPI-6
Bengbu	DBN	RMSE 0.68	RMSE 0.65
		MAE 0.54	MAE 0.52
	BP Neural Network	RMSE 0.98	RMSE 0.69
		MAE 0.75	MAE 0.58

Thus it was concluded that the Deep Belief Network out-performed the BP Neural Network on all time scales of the SPI, hence proving their superiority in short term forecasts.

C. Other Approaches

1) Atmospheric electricity :[12], came up with the first attempt to use atmospheric electricity for rainfall drought prediction. Instead of using standard drought indices such as the SPI or SPEI, he came up with the atmospheric electrical columnar resistance that involved complex calculations from the data that was obtained from satellites. There is an associated ease of use and practicality attributed to it since its calculations does not involve the intricacies of complex probability The process of convection and aerosols aid in drop formation, therefore, allowing the use of such a parameter, i.e., atmospheric electricity, to predict this non-linear time series activity. The atmospheric electrical columnar resistance (Rc) is the resistance of a column of unit cross-sectional volume extending from the a point on land to the initial layer of the atmosphere over the earth. It was found that the lag anti-correlation between the AIRF (All India Rainfall) series and the filtered Rc series over the Bay of Bengal is of consequence and heavily related to drought prediction. They also examined this over various sub sectional areas all over India to give significant positive results with regards to predicting drought over sub sectional regions.

D. Machine Learning Methods

[13]Explored the Artificial Neural Networks (ANN), Support Vector Regression (SVR) and coupled wavelet – ANNs, which pre-process input data using Wavelet Analysis (WA). Machine learning methods have become more apt and mainstream with respect to accuracy and ease of operation, mainly due to their effectiveness in handling non-linear characteristics of hydrological data. To overcome the issue of non-stationary data, suffered by both the ANN and SVR methods, researchers have begun to us wavelet analysis to pre-process the input hydrological data. The wavelet transform is a mathematical tool that provides a time-frequency representation of a signal in the time domain

[14]. They found that 6 month time scales with 3 month lead times were the most accurate, more than even 1 month lead time predictions. SVR models performed slightly better than ANN in 1 month lead time whereas in 3 month lead time, ANN had slightly better performance. Best results were however yielded from the forecasts of WANN (Wavelet Analysis Neural Network). The results partially presented in table 3 indicate that the use of wavelet analysis as a pre-processing tool provided good forecast results for both ANN and SVR models irrespective of forecast lead time.

Table 3. Comparison of Various architectures based on RMSE for two different stations

Model Type	Meisso	Hirna
	SPI-3	SPI-3
	RMSE	RMSE
ANN L1	0.106	0.108
ANN L3	0.130	0.150
SVR L1	0.086	0.100
SVR L3	0.110	0.122
WANN L1	0.029	0.023
WANN L3	0.029	0.089

3. Conclusion and Future Work

The retrieved literature used statistical methods like ARIMA on various drought indices like the SPI and SPEI. Most of the latest methods included the use of traditional ANN's for short -term and long-term forecasting.

One of the methods also included the use of RBM's, a new deep learning methodology for prediction and it gave better results compared to the traditional ANN. Other approaches were also taken like the use of Atmospheric electricity for predicting drought and also use of simple machine learning methods like SVR.

Future work may include the use of much advanced deep learning methods like the Long - Short - Term Memory(LSTM) RNN and other variations of it.

References

[1] T. B. Mckee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," Jan. 1993. [Online]. Available: http://mfkp.org/INRMM/article/14027580

[2] G. Box and G. Jenkins, “M.(1970): Time series analysis, forecasting and control,” Holden-D. iv, San Francisco, 1974.

[3] I.Bordi, K. Fraedrich, M. Petitta, and A. Sutera, “Extreme value analysis of wet and dry periods in sicily,” Theoretical and Applied Climatology, vol. 87, no. 1, pp. 61–71, Jan 2007. [Online]. Available: https://doi.org/10.1007/s00704-005-0195-3

[4] Z. Hao and A. AghaKouchak, “A nonparametric multivariate multi-index drought monitoring framework,” Journal of Hydrometeorology, vol. 15, no. 1, pp. 89–101, 2014. [Online]. Available: https://doi.org/10.1175/JHM-D-12-0160.1

[5] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks:: The state of the art,” International Journal of Forecasting, vol. 14, no. 1, pp. 35 – 62, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207097000447

[6] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, 2006. [Online]. Available: http://science.sciencemag.org/content/313/5786/504

[7] S. Morid, V. Smakhtin, and K. Bagherzadeh, “Drought forecasting using artificial neural networks and time series of drought indices,” vol. 27, pp.2103 – 2111, 12 2007.

[8] Z. Ali, I. Hussain, M. Faisal, H. M. Nazir, T. Hussain, M. Y. Shad, A. MohamdShoukry, and S. HussainGani, “Forecasting drought using multilayer perceptron artificial neural network model,” p. 9, 2017. [Online]. Available: https://doi.org/10.1155/2017/5681308

[9] G. Illeperuma and U. Sonnadara, “Forecasting droughts using artificial neural networks,” 07 2009.

[10] M. Dastorani and H. Afkhami, “Application of artificial neural networks on drought prediction in yazd (central iran),” 2011.

[11] J. Chen, Q. Jin, and J. Chao, “Design of deep belief networks for short-term prediction of drought index using data in the huaihe river basin,” p.16, 2012. [Online]. Available: http://dx.doi.org/10.1155/2012/235929

[12] M. N. Kulkarni, “A new tool for predicting drought: An application over india,” Scientific Reports, vol. 5, p. 7680, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1038/srep07680

[13] A. Belayneh and J. Adamowski, “Drought forecasting using new ma-chine learning methods/prognozowaniesuszy z wykorzystaniem automatycznychsamouczacych˛ si˛emetod,” Journal of Water and Land Development, vol. 18, no. 9, pp. 3–12, 2013.

[14] T. Partal and ÃzgÃŒrKiÅi, “Wavelet and neuro-fuzzy conjunction model for precipitation forecasting,” Journal of Hydrology, vol. 342, no. 1, pp. 199 – 212, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022169407003216

[15] S.V.Manikanthan and D.Sugandhi “ Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel ” International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume- 7 ,Issue 1 – MARCH 2014.

[16] T. Padmapriya and V. Saminadan, “Distributed Load Balancing for Multiuser Multi-class Traffic in MIMO LTE-Advanced Networks”, Research Journal of Applied Sciences, Engineering and Technology (RJASET) - Maxwell Scientific Organization, ISSN: 2040-7459; e-ISSN: 2040-7467, vol.12, no.8, pp:813-822, April 2016.