Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 154 (2016) 1169 – 1175
Procedia Engineering
www.elsevier.com/locate/procedia

12th International Conference on Hydroinformatics, HIC 2016

Meteorological drought forecasting based on climate signals using artificial neural network - a case study in Khanhhoa province Vietnam

Manh Hung Le[a]*, Gerald Corzo Perez[b], Dimitri Solomatine[b], Luong Bang Nguyen[c]

[a] *Barcelona School of Civil Engineering, Technical University of Catalonia, 08001 Barcelona, Spain*
[b] *UNESCO-IHE Institute for Water Education, P.O. Box 3015, 2601DA Delft, the Netherlands*
[c] *Faculty of Water Resources Engineering, Thuy Loi University, 10000 Hanoi, Vietnam*

Abstract

In Khanhhoa Province (Vietnam) long-lasting droughts often occur, causing negative consequences for this region, so accurate drought forecasting is of paramount importance. Normally, drought index forecasting model uses previously lagged observations of the index itself and rainfall as input variables. Recently, climate signals are being also used as potential predictors. In this study, we use 3-month, 6-month, and 12-month of Standardized Precipitation Evapotranspiration Index (SPEI), with a calculation time during the period from 1977 to 2014. This paper aims at examining the lagged climate signals to predict SPEI at Khanhhoa province, using artificial neural network. Climate signals indices from Indian Ocean and Pacific Ocean surrounding study area were analysed to select five predictors for the model. These were combined with local variables (lagged SPEI and rainfall) and used as input variables in 16 different models for different forecast horizons. The results show that adding climate signals can achieve better prediction. Climate signals can be also used solely as predictors without using local variables – in this case they explain the variation SPEI (longer horizons, e.g.12-month) reaching 61 – 80%. The developed model can benefit developing long-term policies for reservoir and irrigation regulation and plant alternation schemes in the context of drought hazard.

Peer-review under responsibility of the organizing committee of HIC 2016

Keywords: Climate Signals, SPEI, Khanhhoa province, Vietnam, drought forecasting

* Manh Hung Le.
E-mail address: lmh.hydrology@gmail.com

Peer-review under responsibility of the organizing committee of HIC 2016
doi:10.1016/j.proeng.2016.07.528

1. Introduction

According to impact on the socio-economic factor of Vietnam, drought is ranked as the third disaster and is not paid much attention in risk reduction [1]. However, in Khanhhoa province Vietnam, drought phenomenon is the most severe hazard threatening the region development. For example, the exceptionally strong El Nino in 2015 – 2016 is believed affecting an absence of rainfall in Khanhhoa province in a long period, causing 10 000 ha stopping to produce crops [2]. Therefore, an effective method for drought mitigation in the region should be carried out.

Firstly, we have to decide about what type of droughts are to be considered. Among four types of droughts defining by the USA Nation Drought Mitigation Center, meteorological drought normally triggers first, causing the consequences of the rest. Therefore, predicting the onset of meteorological drought can benefit a planner in considering the degree of severity in the whole drought process. Secondly, the role of ocean and atmospheric circulation is undeniable in long-lasting drought. There are several papers involving climate signals which represent the dynamics of ocean and atmospheric circulation in drought prediction scheme. For example, eight climate indices were used to predict annual drought conditions in semi-arid watershed in Iran [3]. In Southern Africa, a close linking ENSO events and drought conditions during the period December – March was included in drought forecasting [4]. By using Sea Surface Temperature (SST) in Nino W and Nino 4 region, Nguyen et. al. [5] initiated a multi-scale drought forecasting based on an oscillation of Pacific Ocean's SST in Cai Basin River, Vietnam. Thirdly, drought forecasting model plays as a key factor in accurate decision for drought mitigation. Traditional multi regression or auto- regression was applied widely in quantifying drought index. However, data driven modelling are paying more attention along with a revolution of computation techniques. Many satisfactory applications by using data driven modelling technique were taken place in various areas around the world [5-7].

This paper focusses on forecasting meteorological index considering climate signals surrounding the study area using a data-driven technique (e.g. an artificial neural network). Section 2 describes the study area. Meteorological index, input selection and drought model are discussed in section 3. Results and conclusion are followed by section 4 and section 5 respectively.

2. Study Area

Khanhhoa Province lies between 11°42′50″–12°52′15″N and 108°40′33″–109°27′55″E in South Central Region of Vietnam, with a total area of 5,197 km^2. This study area's climate is affected by tropical monsoon. The annual average rainfall in Khanhhoa province is 1520 mm, is typically lower than a common range of rainfall in Vietnam (1600 – 2500 mm) as well as far away from national annual average (1950 mm). There are two distinct seasons: rainy and dry season. The rainy season lasts in four months, from September to December, accounting for 65 – 75% of the total rainfall. There are no rain in many months in dry season, combining with a high temperature around the year (average 26.7^0 C), leading Khanhhoa province to be a drought prone in Vietnam. Three meteorological stations, namely, Camranh, Khanhvinh, and Nhatrang were chosen as a representative samples of the climate in the study area. The description of three rainfall stations are shown in table 1. Rainfall and temperature data set at those meteorological stations were collected from Vietnam Hydrometeorological Data Centre. The data quality is assured, and reliable enough for calculations. The period used in this paper lasts from January 1977 to December 2014, a duration of 38 years.

Table 1. The description of three rainfall stations in Khanhhoa province

Station Name	Longitude	Latitude	Mean Annual (mm)	Maximum Annual (mm)	Minimum Annual (mm)	Standard Deviation (mm)
Khanhvinh	108^{0}54'E	12^{0}17' N	1629.0	2913.3	679.9	592.3
Camranh	109^{0}10'E	11^{0}57' N	1250.4	2358.0	671.6	433.7
Nhatrang	109^{0}12'E	12^{0}13' N	1411.0	2622.8	802.7	475.0

3. Methodology

3.1. Meteorological drought index

There exists a number of meteorological drought indices. In this paper, we choose a recently rising attention Standardized Precipitation Evapotranspiration Index (SPEI) as simulating drought hazard in the study area because it is multi-scale index and considers precipitation and temperature as input variables [8]. SPEI at three different time scales (i.e. 3-, 6-, and 12- month) was calculated by using SPEI package integrated in R software (https://cran.r-project.org/web/packages/SPEI/)

3.2. Input selection

Vicente-Serrano et. al. [9] stated that drought prediction can be based on lags in the impacts of ENSO events, typically in Nino 3.4 region. In this paper, we consider not only the Nino 3.4 indices but also other climate signals surrounding the study area. We used cross correlation method to evaluate the relationship between the lag time of each input variable and SPEI index. The lag time with best correlation with SPEI were choose as predictors. In this paper, we categorized potential input variables for SPEI prediction into two groups: local variables and climate signals. Local variables include suitable previously lagged observations of SPEI index and rainfall. Climate signals consist of indices in both Pacific Ocean and Indian Ocean. In Indian Ocean, average SST in the box 2°N – 2°S, 70°E – 90°E is chosen as Indian Ocean Index (IND). In the Pacific Ocean, we take into account average SST in Central Tropical Pacific (5°N - 5°S; 160°E - 150°W) or Nino 4 region, average SST in East Central Tropical Pacific or Nino 3.4 region, average SST in a box (15°N – 0°S; 130°E – 150°E) or Nino W region. We also choose a representation of variance in spatial pattern of SST in North Pacific (20°N - 65°N) so-called the Pacific Decadal Oscillation (PDO). Regarding to atmospheric circulation, two indices, namely, the Southern Oscillation Index (SOI), Bivariate ENSO time series (BEST) are selected. The former is developed by the different in Sea Level Pressure (SLP) between Darwin, Australia (12°24'N, 130°54'E) and the island of Tahiti (17°30'N, 149°36'W), while the latter integrates SST and SLP by combining a standardized SOI and a standardized Nino 3.4. SST and SLP are taken from website of the USA National Weather Service Climate Prediction Center (http://www.cpc.noaa.gov/data/indices/).

3.3. Drought forecasting model

ANN is one of the most widely artificial intelligence techniques, which has also been applied in drought forecasting area. Among a pool of ANN approaches, multilayer perceptron feed forward neural network was used in this study due to its popularity. The architecture of model includes three layers: input, hidden and output layer. Each layer is connected by weights and bias but no weight is assigned between nodes within layers. The multilayer perceptron is worked in such a way minimizing the error between output values of model and target values by updating the weights between each nodes [10]. Levenberg-Marquardt optimization algorithm was chosen as the weight updating method.

To exam the quality of model, verification period was chosen from 2001 – 2006 which covers a drought event in 2004 – 2005. Training data set was also divided into training and cross validation dataset. Cross validation has two purposes: (1) early stopping to avoid the risk of over- fitting that models might reproduce poorly with unseen data (verification data); (2) selecting some best predictions from many ANN runs [11, 12]. The number of hidden nodes affects the model performance (see e.g. Chattopadhyay et. al. [13]). In order to find an optimal number of nodes we varied the number hidden nodes from 1 to 2n + 8, where n is the number of input variables. The learning rate and momentum were assigned with value of 0.1 and 0.9 respectively. Due to the variation of results in each run, each node was run 20 times, then the best of 10 runs was averaged; here we followed an idea of a multi-model which typically result in a lower error. To evaluate the quality of model, this paper used coefficient of determination and root mean square error as performance indices.

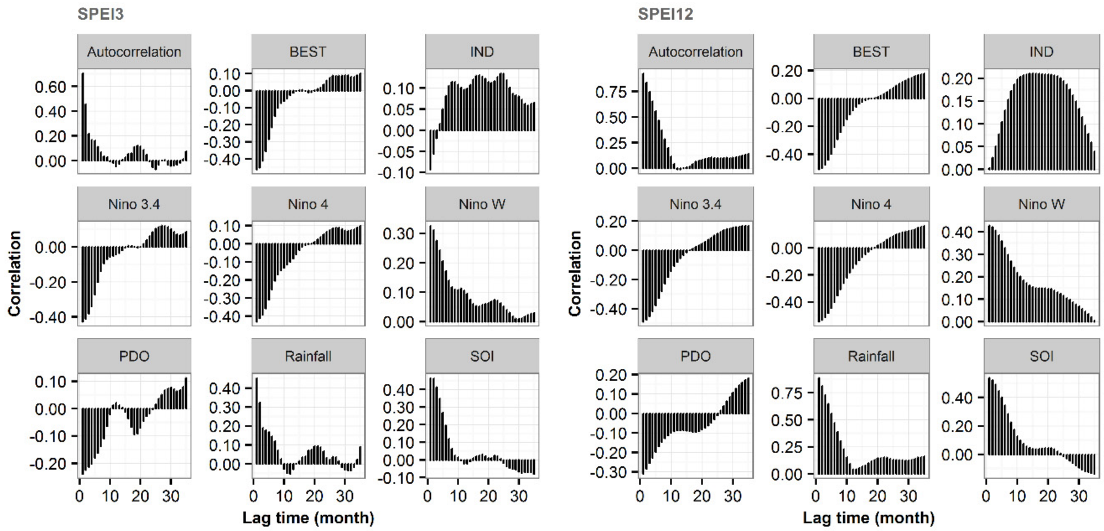

Fig. 1. Auto-correlation and cross-correlation between SPEI3 and SPEI12 with rainfall and climate signals at Khanhvinh station.

4. Results

4.1. Cross correlation analysis and model proposed

Figure 1 shows the relationship between SPEI3 and SPEI12 and theirs potential predictors at Khanhvinh station. It is clear that autocorrelation and rainfall have the highest relationship with SPEI. Among climate signals, good relationships were found in SOI, BEST and SST4, and those relationships increase with an increase in temporal scale of SPEI (SPEI 12). Regarding to SST in Pacific Ocean, except for Nino W region, there are negative correlations between SPEI and Nino 3.4, Nino 4 and PDO. In contrast, SST in Indian Ocean represented by IND shows a positive relationship. Similarly, SPEI also has positive correlation with SOI. However, BEST which is a combination index between Nino 3.4 and SOI shows the negative values in correlation with SPEI. It can be also noticed that IND and PDO shows weak relationships with all SPEI, therefore, they were excluded from model selection.

Based on above cross correlation analysis, we use 16 different models for each temporal scale of SPEI. The aim of using different models lies in examining the response of SPEI to climate signals. Input variables in the first two models are local variables which is previous observation of SPEI and rainfall. It should be kept in mind that SPEI has close relevance to rainfall since precipitation is one of input variables for SPEI calculation. Next, we try to use data from the distant areas (thousand kilometres away) of SST or SLP to predict the severity of drought. From model 3 to model 9, we keep self-correlation and rainfall together with climate signals and from model 10 to model 16 without those local features.

Table 2. Model proposed for SPEI forecasting

	Number of	Local variables		Climate Signals				
	input variables	Auto Correlation	Rainfall	SOI	BEST	Nino 4	Nino 3.4	Nino W
M1	2	x						
M2	4	x	x					
M3	7	x	x	x				
M4	7	x	x		x			
M5	6	x	x			x		
M6	6	x	x				x	

M7	13	x	x	x		x	x	x
M8	13	x	x		x	x	x	x
M9	16	x	x	x	x	x	x	x
M10	3			x				
M11	3				x			
M12	2					x		
M13	2						x	
M14	9			x		x	x	x
M15	9				x	x	x	x
M16	12			x	x	x	x	x

4.2. Drought forecasting analysis

Figure 2 shows the performance of different models in drought forecasting at short time scale (forecast horizon) (SPEI3), medium time scale (SPEI6) and long-time scale (SPEI12). Model 1 and model 2 were displayed as blue colours representing input as local variables, while black colours of model 3 to model 9 are models which input variables are different combination between local variables and climate signals. Red colours representing model 10 to model 16 which input variables are climates signals only. Generally, drought prediction in long-term scale is better than that in medium and short-term scales. By including climate signals with local variables as predictors, the performance indicates an improvement for some combinations. For example, getting the highest performance in all 3 stations for SPEI3 is model 8 with a combination of lagged SPEI, rainfall, BEST, Nino 4 and Nino W, while for SPEI6 is model 3 with SOI addition to local features. When it comes to long-time scale SPEI12, model 5 is the most suitable with an addition of Central Tropic Pacific Nino 4 index. However, adding climate signals as input variables sometimes does not improve even reduce the quality of model performance, for example, model 3 for forecasting SPEI3 at Nhatrang station. Therefore, model performance is independent of the time scale and input variables have to be carefully selected in order to achieve best prediction.

Regarding to models with inputs as climate signals only, accumulation of climate signals can predict those SPEI with same time scale better in long-term. For more details, the best models using input as climate signals only explained the variation of SPEI3 from 20 – 38%. This result increased to 34 – 50% in SPEI6, and up to 61 – 81% in SPEI12. However, RMSE measure of climate signals' models still remains high. It varies from 0.70 – 1.15 for all three time scales. We also use Wilcoxon rank sum test for mean equal testing. If p-value is larger than 0.05, it means there is no significant difference between mean of two variables. As can be seen from table 3, except model 12 for SPEI 12 forecasting, there is generally an agreement between mean of SPEI observation and that in SPEI simulation.

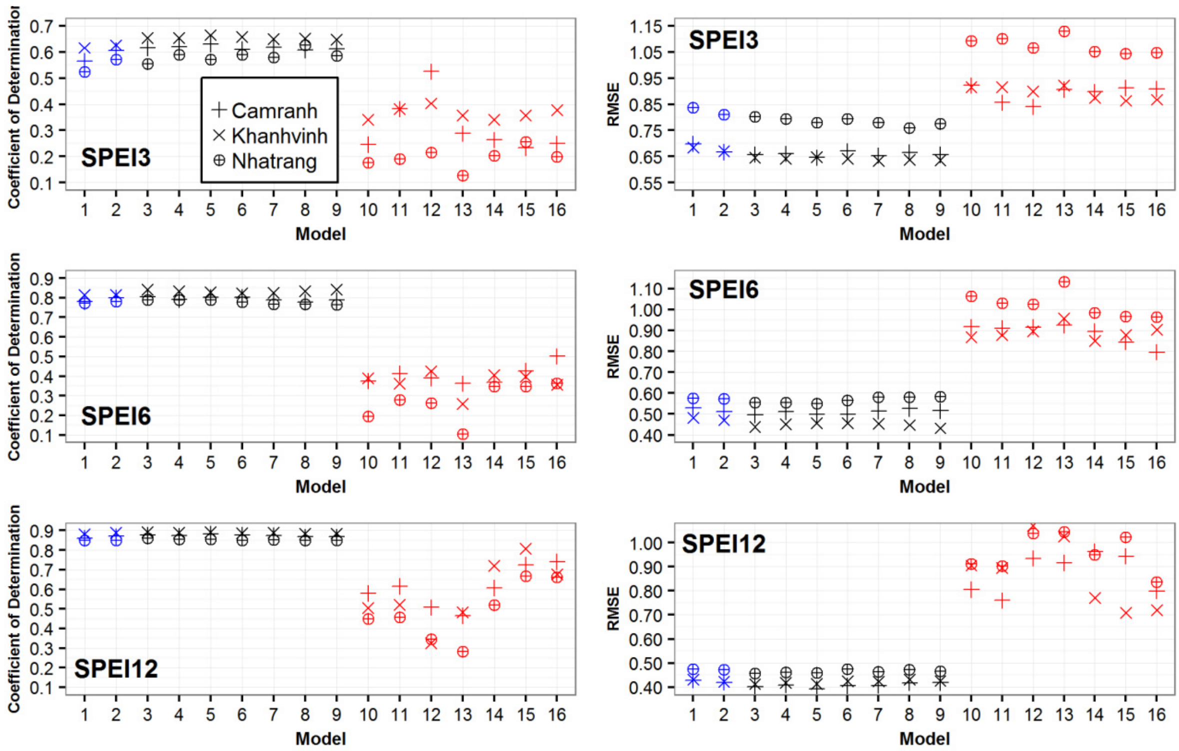

Fig. 2. Comparison 16 models performance in verification period with multi scale SPEI. The verification period is from 2001 to 2006. Blue colors represent models with local variables as input, black colors represent models with local variables and climate signals as input, and red colors represent models with climate signals as input only.

5. Conclusion

The (lagged) climate signals are associated with the change of ocean and atmospheric circulation, leading to rainfall pattern changing across the world. Using those signals could be useful to predict extreme events like droughts. The drought prone Khanhhoa province located in South Central Region of Vietnam, was chosen as the case study to explore how the drought index responds to various varying environmental signals, and to build a data-driven model allowing for predicting the index. The Standardized Precipitation Evapotranspiration Index (SPEI) were calculated with multi time scale: 3-month, 6-month, and 12-month. Five climate signal indices in Pacific Ocean were chosen as input variables, along with previously lagged SPEI and rainfall. The result shows that climate signals can improve the quality of drought forecasting using ANNs, however the input variables have to be carefully selected in order to achieve best prediction. Climate signals themselves can become predictors without using local variables (lagged SPEI and rainfall), and the accuracy increases with long-time scale of drought index.

In further studies, more locations needed to be considered to understand better the response of drought with oscillation of atmosphere and ocean. The presented methodology can be applied to other areas where climate condition is similar to the one in the considered case study.

Table 3. Comparison mean difference between SPEI Observation and SPEI Simulation at Khanhvinh station

Model	SPEI3			SPEI6			SPEI12		
	Mean SPEI Observation	Mean SPEI Simulation	P-value	Mean SPEI Observation	Mean SPEI Simulation	P-value	Mean SPEI Observation	Mean SPEI Simulation	P-value

1	-0.23	-0.11	0.276	-0.24	-0.16	0.622	-0.15	-0.10	0.995
2	-0.23	-0.15	0.398	-0.24	-0.18	0.772	-0.15	-0.09	0.903
3	-0.23	-0.15	0.437	-0.24	-0.18	0.818	-0.15	-0.11	0.909
4	-0.23	-0.17	0.655	-0.24	-0.19	0.800	-0.15	-0.13	0.769
5	-0.23	-0.14	0.372	-0.24	-0.19	0.944	-0.15	-0.12	0.769
6	-0.23	-0.13	0.387	-0.24	-0.21	0.976	-0.15	-0.09	0.998
7	-0.23	-0.26	0.992	-0.24	-0.22	0.986	-0.15	-0.09	0.989
8	-0.23	-0.22	0.745	-0.24	-0.18	0.769	-0.15	-0.09	0.881
9	-0.23	-0.19	0.662	-0.24	-0.25	0.862	-0.15	-0.12	0.941
10	-0.23	-0.05	0.051	-0.24	-0.10	0.412	-0.15	-0.21	0.256
11	-0.23	-0.13	0.134	-0.24	-0.14	0.416	-0.15	-0.15	0.349
12	-0.23	-0.25	0.481	-0.24	-0.34	0.739	-0.15	-0.37	**0.046**
13	-0.23	-0.09	0.061	-0.24	-0.14	0.355	-0.15	-0.18	0.412
14	-0.23	-0.23	0.602	-0.24	-0.10	0.432	-0.15	-0.30	0.189
15	-0.23	-0.22	0.503	-0.24	-0.15	0.163	-0.15	-0.14	0.671
16	-0.23	-0.13	0.180	-0.24	0.00	0.191	-0.15	-0.05	0.909

Note: We use Wilcoxon rank sum test to exam the difference between mean value of SPEI observation and SPEI simulation. Boldface means there is a significant different between SPEI observation and SPEI simulation.

Acknowledgements

The authors owe special thanks to Vietnam Hydrometeorological Centre Data, Mr. Dang Dinh Duc and Mr. Bui Van Chanh to provide data needed to conduct this research. First author also thanks EU Erasmus Mundus Masters Programme in Flood Risk Management for partial funding of this study.

References

[1].N. Huy and R. Shaw, Chapter 8 Drought Risk Management in Vietnam, in Droughts in Asian Monsoon Region, Emerald Group Publishing Limited, 2011, pp. 141-161.
[2].MARD, Report of drought mitigation on Vietnam Southern Central, Central Highland and Eastern South Regions under El Nino phenomenon (in Vietnamese), 2016
[3].B. Choubin, K.S. Shahram, M. Arash, A. Sajjad, and A. Pedram, Drought forecasting in a semi-arid watershed using climate signals: a neuro-fuzzy modeling approach, J Mt. Sci-Engl., 11 (2014) 1593-1605.
[4].D. Manatsa, T. Mushore and A. Lenouo, Improved predictability of droughts over southern Africa using the standardized precipitation evapotranspiration index and ENSO, Theor. Appl. Climatol., (2015) 1-16.
[5].N.L. Bang, L.Q. Fang, N.T. Anh, and K.Hiramatsu, Adaptive Neuro-Fuzzy Inference System for Drought Forecasting in the Cai River Basin in Vietnam, J Fac. Agr. Kyushu U., 60 (2015). 405-415.
[6].H.B. Abarghouei, M.R. Kousari and M.A.A. Zarch, Prediction of drought in dry lands through feedforward artificial neural network abilities, Arab. J Geosci., 6 (2013) 1417-1433.
[7].A. Belayneh, A., J. Adamowski, B.Khalil and B.Ozga-Zielinski, Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models, J Hydrol., 508 (2013) 418-429.
[8]. S.M. Vicente-Serrano, S. Beguería and J.I. López-Moreno, A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, J Climate, 23 (2010) 1696-1718.
[9].S.M. Vicente-Serrano, S. Beguería and I.A. Juan, A multiscalar global evaluation of the impact of ENSO on droughts, J Geophys Res: Atm., 116 (2011).
[10].S. Haykin, Neural networks, a comprehensive foundation, Prentice-Hall (2004).
[11].K. Chau and C. Wu, A hybrid model coupled with singular spectrum analysis for daily rainfall prediction, J Hydroinform, 12 (2010) 458-473.
[12] A. Elshorbagy, G. Corzo, S. Srinivasulu, and D.P. Solomatine, Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology, Hydrol. Earth Syst. Sci., 14 (2010), 1931–1941.
[13].S.Chattopadhyay and G. Chattopadhyay, Identification of the best hidden layer size for three-layered neural net in predicting monsoon rainfall in India., J Hydroinform, 10 (2008) 181-188.