Application of the ARIMA Models in Drought Forecasting Using the Standardized Precipitation Index

Ping Han[1], Pengxin Wang[2,*], Miao Tian [2], Shuyu Zhang[3], Junming Liu[2], Dehai Zhu[2]

[1] Department of Physics, College of Science, China Agricultural University, West Campus, Beijing 100193, PR China
hanrose@cau.edu.cn
[2] Department of Geographic Information Engineering, College of Information and Electrical Engineering, China Agricultural University, East Campus, Beijing 100083, PR China
wangpx@cau.edu.cn
[3] Remote Sensing Information Center for Agriculture of Shaanxi Province, Xi'an 710015, PR China
shuyuzhang88@sina.com

Abstract. The standardized precipitation index (SPI) was used to quantify the classification of drought in the Guanzhong Plain, China. The autoregressive integrated moving average (ARIMA) models were developed to fit and forecast the SPI series. Most of the selected ARIMA models are seasonal models (SARIMA). The forecast results show that the forecasting power of the ARIMA models increases with the increase of the time scales, and the ARIMA models are more powerful in short-term forecasting. Further study was made on the correlation coefficient between the actual SPIs and the predicted ones for the forecasting. It is shown that the ARIMA models can be used to forecast 1-month leading values of all SPI series, and 6-month leading values for SPI with time scales of 9, 12 and 24 months. Our study shows that the ARIMA models developed in the Guanzhong Plain can be effectively used in drought forecasting.

Keywords: drought forecasting; standardized precipitation index; autoregressive integrated moving average model

1 Introduction

Drought is a slow-onset and creeping natural hazard that occurs in all regions of the world. Prolonged multiyear drought has caused significant damages in natural environment as well as in human lives. The Estimation for the cost of drought in the

* Corresponding author. Pengxin Wang, Tel.: 86 10 62737622.

United States ranges from $6 to $8 billion annually [1]. In China, the amount of loss caused by drought ranks the first in all natural hazards. With the increase of the population and the severity of drought, an effective mitigation of the impacts caused by drought is imperative.

In 1965, Palmer presented a drought index that incorporated antecedent precipitation, moisture supply, and the pioneering evapotranspiration [2]. McKee et al. developed the SPI as an alternative to the Palmer Index for Colorado [3]. The SPI is considered to have several advantages over the PDSI [3,4,5]. The first advantage is that the SPI is based only on precipitation [3]. Due to this reason, the SPI is also not adversely affected by topography. Second, the SPI is calculated on various timescales, which allows it to describe the various types of droughts: the shorter time scales for meteorological and agricultural droughts, and the longer ones for hydrological drought. Third, because the SPI is normally distributed, the frequencies of drought events at any location for any time scale are consistent. Hayes et al. argued that the SPI detects moisture deficits more rapidly than the PDSI, which has a response time scale of approximately 8 to 12 months [5]. Paulo et al. used several drought indices in Portugal, and found that the SPI showed its superiority for the purpose of drought monitoring [6,7]. Labedzki used SPI to analyse the local meteorological drought and evaluate the drought risk in Bydgoszcz, Poland [8].

The time series forecasting has been widely applied and become an important approach of drought forecasting. One of the most widely used time series model is the autoregressive integrated moving average (ARIMA) model [9]. The wide application of the ARIMA model in many areas is due to its flexibility and systematic search (identification, estimation and diagnostic check) in each stage for an appropriate model [10]. The ARIMA model has several advantages over other approaches, such as moving average, exponential smoothing, neural network, and in particular, its forecasting capability and its richer information on time-related changes [11, 12]. The ARIMA models have also been used to analyze and model hydrologic time series [11,13,14]. Fernandz et al. used SARIMA model to forecast stream-flow in a small watershed in Galicia [15]. Durdu developed linear stochastic models for forecasting droughts in Turkey using SPI series as drought index [16].

The Guanzhong Plain is located in the northwest of China. This area is subjected to water stress and drought conditions. The frequency of drought is on average about once in 7 years [17]. The Plain has flat terra and fertile soil, and is the political and economical center of Shaanxi Province. Drought forecasting for this area can help to mitigate the effects of drought, and is in favor of effective water resource management. In this paper, The SPI is used as a drought index to describe the drought condition of the Guanzhong Plain. The SPI time series of multiple time scales in the Guanzhong Plain are calculated. The ARIMA models are applied to simulate and forecast the SPI series.

2 SPI time series forecasting models

The SPI time series of multiple time scales can be computed (typically 3, 6, 9, 12 and 24 months) according to the McKee's method[4]. The classification of dry and wet spells resulting from the values of the SPI is shown in table 1.

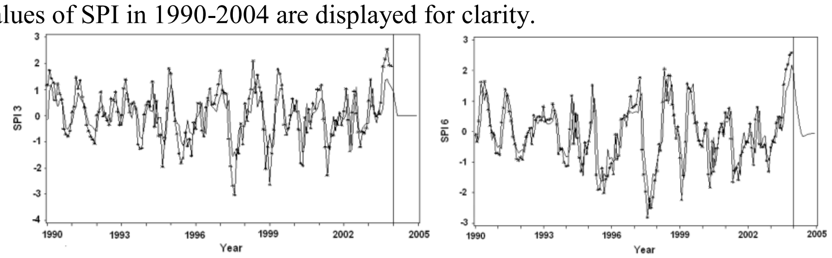
Table 1. Drought classification of SPI.

SPI value	Class
>=2	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
<=-2	Extremely dry

The SPI data set from 1966 to 2003 is used for model development. The data set of 2004 is used for model validation. Based on the steps for developing the ARIMA models [18,19], the ARIMA models fit for SPI3, SPI9, SPI12 and SPI24 are identified respectively. They are SPI3, MA(2); SPI6, ARIMA(1,0,0)(0,0,1)$_6$; SPI9, ARIMA(1,0,0)(0,0,1)$_9$; SPI12, ARIMA(1,0,0)(0,0,1)$_{12}$; and SPI24, ARIMA(0,1,0)(2,1,1)$_{24}$. It is observed that most of the SPI time series have the seasonal features. As the time scale increases, the seasonal feature is more and more distinct, and the series need to be seasonally differenced.
The forecast is done for 12 leading months using the selected models, i.e., forecast the SPI values in 2004. By comparing the predicted data with the original data, the forecasting capability of the models is discussed.

3 Results and discussion

SPI3, SPI6, SPI9, SPI12 and SPI24 are respectively fitted by the selected best models from historical data. The forecast is done with 12-month lead-time. The values of SPI in 2004 are predicted. Fig. 1 shows the plots of fitting (forecasts for 1966-2003) and forecasting (forecasts for 2004) using the chosen ARIMA models, in which only the values of SPI in 1990-2004 are displayed for clarity.



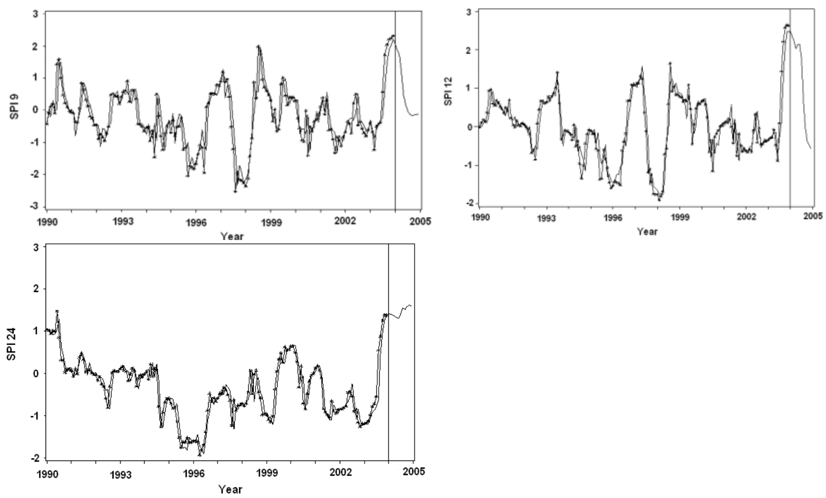


Fig. 1. The SPI series, the fitted series and the forecasts for 12 months ahead (1990-2004). The curve with stars is for the calculated SPI. The one without stars is for fitting and forecasting. The predicted values of SPI for 2004 are on the right side of the reference line.

Observing the fitting figure, how well the chosen model fits the data series is indirectly shown. In Fig. 1, it is obvious that the fitted data follow the original data very closely, especially for higher SPI series (SPI12 and SPI24). This indicates that the chosen best ARIMA models capture the patterns of the SPI series. However, the models that fit the data well do not all have good forecasting capability. The predicting power of the ARIMA models should be valuated. The forecast with 12-month lead-time are analyzed. It is found that the predicted SPI values fit the original ones better and better with the increase of the time scale in Fig. 1. The values of SPI3 predicted for 2004 tend to be constant. However, this situation is changed in SPI6, SPI9, SPI12 and SPI24. The predicted data gradually tends to present the stochastic change of the SPI series. It indicates that the forecasting power of the ARIMA models is improved as the time scale increases.

By comparing the calculated values with the predicted ones, the absolute percentage errors (APE, $\left|(X_i - \hat{X}_i)/X_i\right| \times 100$) are calculated as an analysis on the forecasting power, as shown in Table 2. It is obvious that all SPI series have less APEs on 1-month ahead forecasting. The mean absolute percentage error (MAPE) is 5.8. This demonstrates that the ARIMA models have better result on short-term forecasting. The APEs seem to be less for higher SPI series (SPI12 and SPI24). For 1-month ahead, the APEs of all SPI series decrease with the increase of the time scale. For 1-9 month ahead, the APEs of SPI12 are all less than 6, and the APEs of SPI24 are generally less than 10, with only two exceptions (13.3 and 17.4). These results indicate that if the SPI series have a longer time scale, the ARIMA models will have a better forecasting power.

Table 2. Absolute percentage errors of SPI3, SPI6, SPI9, SPI12 and SPI24 for 12-month lead-time. Only values less than 20 are shown.

SPI series	One month	Two months	Three months	Four months	Five months	Six months	Seven months	Eight months	Nine months	Ten months	Eleven months	Twelve months
SPI3	8.1											
SPI6	8.4	15.1										
SPI9	8.9	14.9										
SPI12	2.2	0.9	5.5	0.9	0.4	5.7	5.4	1.8	5.0	13.2		
SPI24	1.4	4.1	13.3	8.7	3.9	17.4	2.9	7.6	2.5	9.7	8.0	5.3

In order to exactly determine how good the forecasting power of the ARIMA models is, the forecasts are done with 1-month to 8-month lead-time. For example, 2-month lead-time forecast means that during March 2004 the forecast for May 2004 is performed. The correlation coefficients between the actual SPIs and the predicted ones from the ARIMA models are used as the criteria to evaluate the fit. Table 3 provides the coefficients for 1-8 month lead-time. It is observed that with more and more longer lead-time the coefficient decreases. To a given lead-time, the coefficient gradually increases with the increasing of the time scale, which shows that the ARIMA model is an appropriate model for the SPI series, and its power of forecasting is better for higher SPI series (SPI9, SPI12 and SPI24). Evidently, the ARIMA models give good forecast results for 1-month lead-time with the correlation coefficients ranging from 0.873 to 0.980. For high SPI series (SPI9, SPI12 and SPI24), the forecasts have good results up to 6-month lead-time with the correlation coefficients larger than 0.7. Therefore, the selected ARIMA models have a better power of forecasting for the SPI series. The ARIMA models can be used to forecast the change of the series of SPI3 and SPI6 one to two months ahead, and the future change of the series of SPI9, SPI12 and SPI24, six months ahead. The ARIMA model therefore exhibits a strong forecasting capability.

Table 3. Correlation coefficients of different lead-time. Blank cells are for the non-existing correlation coefficients.

SPI series	One-month lead-time	Two-month lead-time	Three-month lead-time	Four-month lead-time	Five-month lead-time	Six-month lead-time	Seven-month lead-time	Eight-month lead-time
SPI3	0.873	0.417	0.360					
SPI6	0.888	0.762	0.673	0.580	0.514	0.248		
SPI9	0.955	0.889	0.830	0.786	0.772	0.803	0.695	0.523
SPI	0.98	0.985	0.985	0.985	0.935	0.74	0.538	0.388

12	9					2		
SPI	0.98	0.885	0.876	0.852	0.821	0.77	0.523	0.475
24	0					5		

4 Conclusions

The Guanzhong Plain is one of the areas subjected to drought. In this study, the ARIMA models are developed to forecast drought using the SPI as the drought index. The results show that the selected ARIMA models are appropriate for the SPI series. Exactly the same as in many other applications[10,14], the ARIMA models demonstrate better capability in short-term forecasting. The absolute percentage errors (APE) between the actual SPIs and the predicted ones for 12-month lead-time are all less on 1-month ahead forecasting. The correlation coefficients between the actual data and the predicted ones are all larger on 1-2 month lead-time forecasting. Moreover, in this study, the ARIMA models also demonstrate a better forecasting power even on 6 month leading values for SPI9, SPI12 and SPI24. The study also shows that the ARIMA models have a good forecasting capability for the SPI series with longer time scales. The APEs are less for higher SPI series (SPI12 and SPI24). To a given lead-time, the correlation coefficient gradually increases with the time scale increasing. This may be because the increase in the length of the time scale effectively reduces the noise of the SPI series. Therefore, the selected best ARIMA models developed from the SPI time series can be used for drought forecasting in the Guanzhong Plain.

This study contributes to the exploration of a feasible way on drought forecasting in the Plain. The results demonstrate that the developed ARIMA models have a good forecasting power and can be used for drought forecasting. Further study should be focused on improving the precision of the model forecasting, and on studying the types of droughts described by the SPI series with different time scales, which will effectively assist the local authorities to mitigate the impacts caused by drought, and to reasonably use the water resources.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 41071235 and 40871159), the China National Support Program of Science and Technology (Grant No. 2012BAH29B03), and China Agricultural University’s Grant, 2012JC004.

References

1. Schubert S., Koster R., Hoerling M.: Predicting Drought on Seasonal-to-decadal Time Scales. Bulletin of the American Meteorological Society. 88,1625-1630(2007)
2. Palmer W.C.: Meteorological Drought. Research Paper No. 45, 58pp. US Department of Commerce Weather Bureau, Washington, DC (1965)
3. McKee T.B., Doesken N.J., Kleist J.: The Relationship of Drought Frequency and Duration to Time Scales. Eighth Conference on Applied Climatology, pp.179-184. Anaheim, CA, American Meteorological Society(1993)
4. Guttman N.B.: Comparing the Palmer Drought Severity Index and the Standardized Precipitation Index. Journal of the American Water Resources Association. 34, 113-121(1998)
5. Hayes M.J., Svoboda M., Wilhite D.A., Vanyarkho O.: Monitoring the 1996 Drought Using the SPI. Bulletin of the American Meteorological Society. 80, 429-438(1999)
6. Paulo A.A., Pereira L.S., Matias P.G. : Analysis of Local and Regional Droughts in Southern Portugal Using the Theory of Runs and the Standardized Precipitation Index. In: Rossi G., Cancelliere A., Pereira L.S., Oweis T., Shatanawi M., Zairi A. (Eds): Tools for Drought Mitigation in Mediterranean Regions, pp.55-78. Kluwer, Dordrecht (2003)
7. Paulo A.A., Pereira L.S. : Drought Concepts and Characterization, Comparing Drought Indices Applied at Local and Regional Scales. Water International. 31, 37-49 (2006)
8. Labedzki L.: Estimation of Local Drought Frequency in Central Poland Using the Standardized Precipitation Index. Irrigation Drainage. 56, 67-77(2007)
9. Box G.E.P., Jenkins G.M.: Time Series Analysis Forecasting and Control, 525 pp. Holden-Day, San Francisco(1976)
10. Zhang G.P.: Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing. 50,159-175(2003)
11. Yurekli K., Kurunc K., Ozturk F.: Application of Linear Stochastic Models to Monthly

Flow Data of Kelkit Stream. Ecological Modeling. 183,67-75(2005)

12. Hu W.B., Tong S.L., Mengersen K., Connell D.: Weather Variability and the Incidence of Cryptosporidiosis: Comparison of Time Series Poisson Regression and SARIMA Models. Annals Epidemiology. 17, 679-688(2007)
13. Fernando D.A.K., Jayawardena A.W.: Generation and Forecasting of Monsoon Rainfall Data. In Proceedings of the 20th WEDC Conference, pp.310-313. Colombo, Sri Lanka(1994)
14. Modarres R.: Streamflow Drought Time Series Forecasting. Stochastic Environmental Research and Risk Assessment. 21, 223-233(2006)
15. Fernandez C., Vega J.A., Fonturbel T., Jimenez E.: Streamflow Drought Time Series Forecasting: a Case Study in a Small Watershed in North West Spain. Stochastic Environmental Research and Risk Assessment. 23, 1063-1070(2008)
16. Durdu O. F.: Application of Linear Stochastic Models for Drought Forecasting in the Buyuk Menderes River Basin, Western Turkey. Stochastic Environmental Research and Risk Assessment. 24, 1145-1162(2010)
17. Editing Committee on Brief Actual Records of Historical Natural Disasters for Shaanxi: Brief Actual Records of Historical Natural Disasters for Shaanxi (in Chinese), 271pp. China Meteorological Press, Beijing(2002)
18. Bowerman B.L., O’Connell R.T.: Forecasting and Time Series: an Applied Approach, 521pp. China Machine Press, Beijing(2003)
19. Box E.P., Jenkins G.M., Reinsel G.C.: Time Series Analysis: Forecasting and Control, Third Ed., 327pp. Post & Telecommuncations Press, Beijing(2005)