Mathematical and Computer Modelling 51 (2010) 1398–1403

Contents lists available at ScienceDirect

Mathematical and Computer Modelling

journal homepage: www.elsevier.com/locate/mcm

Drought forecasting based on the remote sensing data using ARIMA models

Ping Han [a,1], Peng Xin Wang [b,*], Shu Yu Zhang [c], De Hai Zhu [b]

[a] Department of Physics, College of Science, China Agricultural University, West Campus, Beijing, 100193, PR China

[b] Department of Geographic Information Engineering, College of Information and Electrical Engineering, China Agricultural University, East Campus, Beijing, 100083, PR China

[c] Remote Sensing Information Center for Agriculture of Shaanxi Province, Xi'an, 710015, PR China

ARTICLE INFO

Keywords:
Vegetation temperature condition index
Spatio-temporal series
Drought forecasting
Forecasting errors
Simulating errors

ABSTRACT

Regarded as a near real time drought monitoring method, the VTCI index based on remote sensing data is applied to the drought forecasting in the Guanzhong Plain. ARIMA models are used in the VTCI series, and forecast its changes in the future. A new way of modeling for the spatio-temporal series is used in the VTCI series. The time series of 36 pixels are studied firstly for their fitting models. Then the ARIMA model fitting for the whole area is determined. The AR(1) model are chosen to be the best model used in each pixel of the whole area, and the forecast is done with 1–2 steps. The results show that forecasting accuracy is better, 1 step is better than 2 steps. The historical VTCI data are simulated by AR(1) models. Comparing the simulating data with the historical data, the results show that the simulating accuracy is better. Most of the simulating errors are small. All results demonstrate that AR(1) model developed for VTCI series can be used for the drought forecasting in the Guanzhong Plain.

1. Introduction

Prolonged multiyear drought has caused significant damages both in the natural environment as well as in the development of the human society. The annual estimate for the cost of drought in the United States ranges from 6 to 8 billion dollars [1]. In China, the amount of loss caused by drought ranks the first in the list of all natural hazards [2]. How to effectively monitor and forecast the droughts has become the research focus, which can help to take effective strategies and measures in advance to mitigate the damages of droughts.

As the remote sensing technology makes more and more process, with the development of GIS and GPS, the real time monitoring droughts over the large areas can be achieved. The drought monitoring comes into a new world [3]. Some indexes developed from the remote sensing data, such as the Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST), etc., have been used to monitor the agricultural droughts in the relation to the plant growth [4–7]. Wang et al. [8,9] developed the Vegetation Temperature Condition Index (VTCI) based on the information of the LST versus the NDVI scatter plot falling into a triangular shape. It is testified that the VTCI index can be used to effectively monitor the drought of an area in the real time [8,10]. It is found that the VTCI index can better reflect the severity of drought when it is used in Guanzhong Plain, Shaanxi.

* Corresponding author. Tel.: +86 10 62737622.
E-mail address: wangpx@cau.edu.cn (P.X. Wang).

1 Engaged in quantitative remote sensing and application.

0895-7177/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mcm.2009.10.031

The autoregressive integrated moving average (ARIMA) model [11] is one of the most widely used time series models. The popularity of ARIMA model in many areas is due to its flexibility and the systematic searching at each stage (identification, estimation and diagnostic check) for an appropriate model [12]. The ARIMA model approach has several advantages over others, such as moving average, exponential smoothing, neural network, and in particular, its forecasting capability and its richer information on time-related changes [13,14]. ARIMA models have been used to analyze and model hydrologic time series [15,13,16]. In this paper, VTCI is used as a drought index to describe the drought condition of the Guanzhong Plain. The ARIMA models are applied to simulate and forecast VTCI series. A new way of modeling for the spatio-temporal series is used in the VTCI series.

2. Study area and data

The Guanzhong Plain is located in the northwest of China. The average elevation of the plain is approximately 400 m. The mean annual precipitation in the Plain is about 500–700 mm. This area is subjected to water stress and drought condition. The frequency of drought is averagely about once in 7 years [17]. The Plain has a flatness terra and a 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 a reasonable water resources management.

According to the theory mentioned in Wang's paper [8,9], and using the remote sensing data of the Guanzhong Plain, VTCI series from 1999 to 2006 can be acquired. In each year, the period from the first ten days of March to the last ten days of June is studied. Each image reflects the drought of ten days. Each pixel in the image has a VTCI value. The lower value of VTCI, the heavier the drought is. Here, the ARIMA models are used to simulate the VTCI series and forecast their changes in the future. The data set before the last ten days of March in 2006 is used for model development. The data set after that period is used for model validation.

3. VTCI time series forecasting models

3.1. ARIMA models

ARIMA models include three basic types: autoregressive (AR) models, moving average (MA) models, and the combined AR and MA (ARMA) models. AR, MA and ARMA can be used when the data are stationary. A stationary time series can be defined when the data have a constant mean and no trend overtime. The time series generally present ascending or descending trends. Such series are non-stationary. Non-stationary series can be modeled by allowing differencing the data series to stationary. The letter "I" (integrated) in ARIMA indicated that the modeling time series has been transformed into a stationary time series. The time series often contain cyclic features (seasonal effects), to which seasonal differencing is often used to remove the seasonal effects. These kinds of models is known as SARIMA models. Thus, ARIMA models have two general forms: non-seasonal ARIMA (p, d, q) and multiplicative seasonal ARIMA $(p, d, q)(P, D, Q)_S$, where p, d, q are non-seasonal parts of the model, and P, D, Q are seasonal parts of the model [18,19]. In this study, only ARIMA (p, d, q) model is used, which is primarily discussed as followed.

The form of a non-seasonal ARIMA (p, d, q) model is written as:

$$\varphi(B)\nabla^d X_t = \theta(B)a_t \tag{1}$$

where

$$\begin{aligned}\theta(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q \\ \varphi(B) &= 1 - \varphi_1 B - \cdots - \varphi_p B^p\end{aligned} \tag{2}$$

X_t is the observed series, θ_i and φ_i are model parameters, and p and q are orders of the model. B is the backward shift operator ($B^k X_t = X_{t-k}$). Random errors, a_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2 (white noise). ∇^d describes differencing operation to data series to make the data series stationary, and d is the number of differencing.

In ARIMA model approach, the past observations are analyzed to formulate a model describing the inner correlation among them. The time series is then extrapolated into the future according to the model. The development of ARIMA model includes three stages: identification, estimation and diagnostic checking. The identification stage involves the determination of the differencing requirement of making the time series stationary, and the identification of the temporal structure of the model. Stationarity is a necessary condition for building an ARIMA model. If the time series is non-stationarity, d is estimated to reduce the series to stationarity. ACF (Autocorrelation Coefficient Function) and PACF (Partial ACF) of the differenced time series are used to identify the temporal structure of the models (the values of p, q). If the autocorrelation coefficient (r_k) tails off, the partial autocorrelation coefficient (ϕ_k) cuts off and is not significantly different from zero after lag p, it suggests AR $(p, d, 0)$ process. If ϕ_k tails off, r_k cuts off and is not significantly different from zero after lag q, it suggests MA $(0, d, q)$ process. If ϕ_k and r_k both tail off, ARIMA (p, d, q) will be fitted to the series.

In the estimation stage, the model parameters (θ_i, φ_i, Θ_i, Φ_i) are computed by the moment method, the least square method, or the maximum likelihood method. In this study, the least square method is applied. After that the parameters

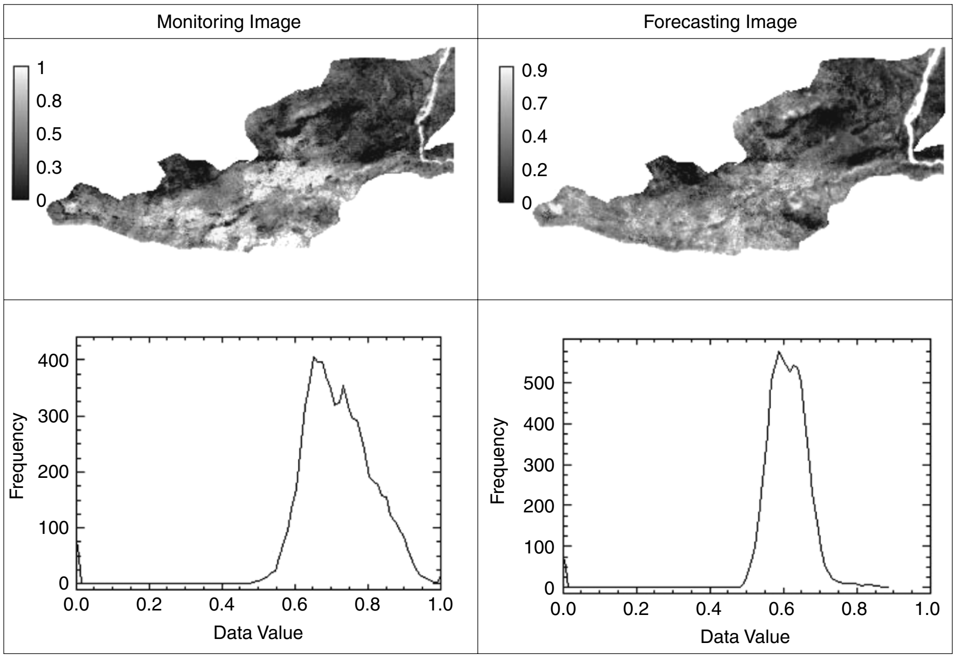

Fig. 1. Monitoring, forecasting images and respective frequency distributing for the first ten days of April in 2006.

should be tested to see whether they are statistically significant or not. It tests whether or not the parameters are zero. The t-statistics and P-value are conducted to test the significance of the parameters. The t-statistics is not very informative by itself, but is used to determine the P-value. If the P-value is less than 0.05, a level corresponding 95% of confidence interval, the parameter is significant.

Finally, the diagnostic checks are used to verify whether the models are adequate (i.e., to determine whether the residuals are white noises). Here, $Q(r)$ statistic suggested by Ljung and Box [20] is used to test the adequacy of a model by choosing a level of significance (5%) and comparing the $Q(r)$ value with the x^2-table critical value. If the $Q(r)$ value is less than the x^2-table critical value, or the probability of $Q(r)$ is greater than 0.05, the model is adequate. If the model is not adequate, going back to the identification stage. Those models passing the adequate checking are the candidate models. Then, the best model based on minimum AIC (Akaike Information Criterion) and SBC (Schwarz Bayesian Criterion) is identified. The idea is that the model should best collocate with the simulating precision and the number of parameters.

3.2. VTCI time series modeling

VTCI series is the spatio-temporal changing series. In comparison with the general forecasting for the series only changing with time, it is very difficult to predict the time series of VTCI images, because that we need not only capturing the time-related variations of one pixel value, but also need getting the spatial changes about the pixel value in the region. To the spatio-temporal series, one forecasting method is that the orthogonal decomposition is applied to make the spatio-temporal series into two parts—time functions and spatial functions. The several dominating time functions are selected, and are predicted by time series models. The product of the forecasting time functions and the spatial functions is regarded as the change of the series in the future. The shortage of this method is that much information on original series is lost due to only using several time functions for predicting. Here, a new method is used. The model is developed to every pixel of the image. Here, some pixels are firstly chosen and studied for their fitting models, and then the best models for all pixels of the whole area can be determined.

Considering of averagely distributed in the area, 36 pixels are chosen, where weather stations are. On the other hand, the meteorological data can be helped for studying the character of the time series of 36 pixels. According to three stages of model development, the selected best models for 36 pixels are identified. It can be found that the AR(1) models are suitable for 36 pixels. We take other pixels as having same models as 36 pixels. So, the AR(1) model are used to each pixel of whole area, and the forecast is done with 1–2 steps. The 1 step forecast is done to get the VTCI image of the first ten days of April in 2006. 2 step forecast is done to get the VTCI image of the middle ten days of April in 2006. By comparing the original data with the predicted ones, the forecasting capability of the models is discussed.

Figs. 1 and 2 are the monitoring, forecasting Images and respective frequency distributing for the first and middle ten days of April in 2006. Fig. 3 is forecasting error images (difference between the monitoring and forecasting images) and respective frequency distributing for the first and middle ten days of April in 2006. Observing these figures, how well the AR(1) model fit the data series can be seen.

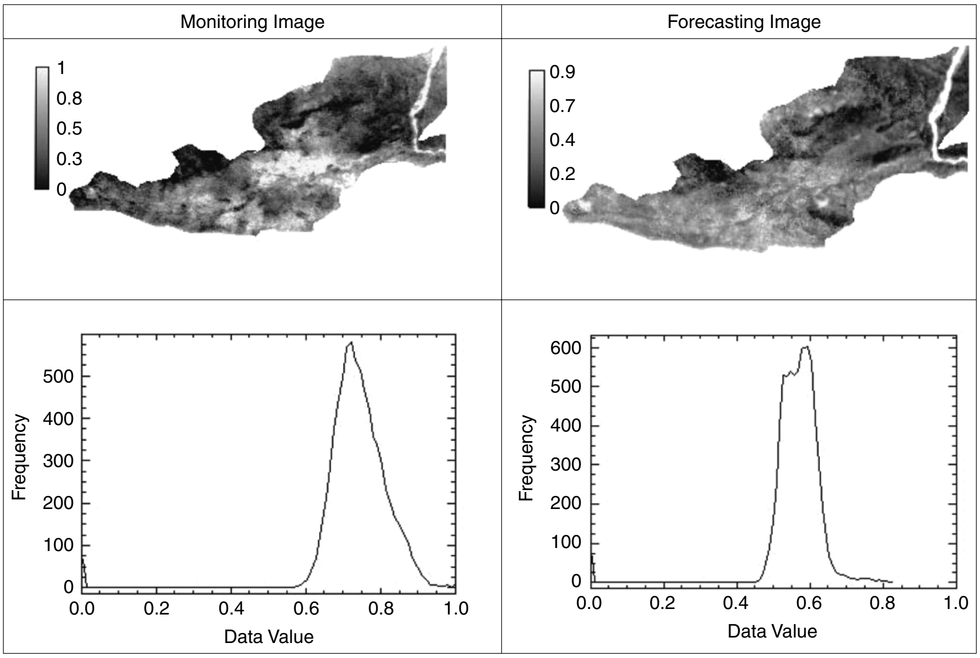

Fig. 2. Monitoring, forecasting images and respective frequency distributing for the middle ten days of April in 2006.

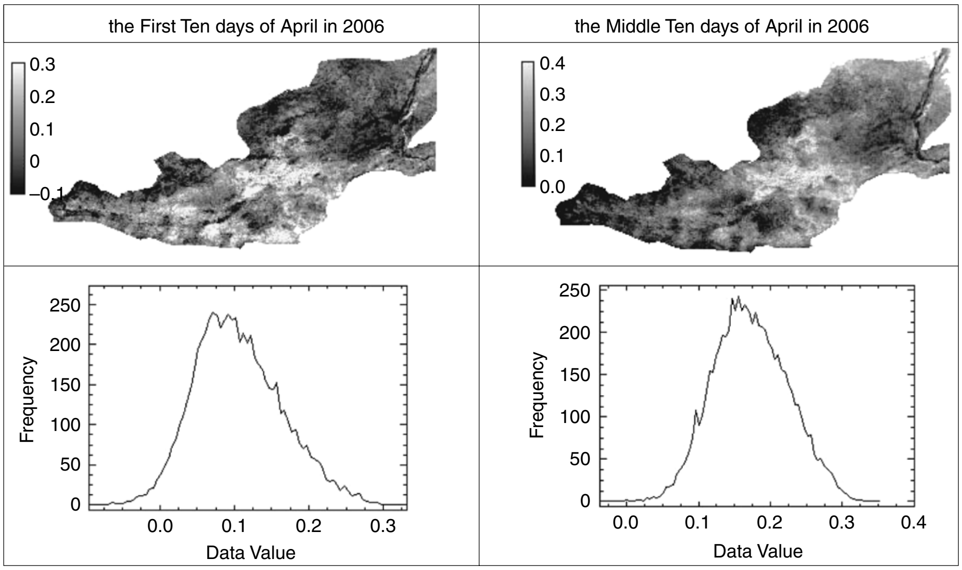

Fig. 3. Forecasting error images and respective frequency distributing for the first and middle ten days of April in 2006.

(1) As a whole, the forecasting images of either 1 or 2 step are all similar to the monitoring images. The primary characters of monitoring images are better reflected in forecasting images, which is that drought of the North area is heavier than that of the South area (the Upside is the North, and downside is the South in the image), and the area near the bank of the Yellow River is quite droughty. Comparing monitoring images with forecasting images on the frequency distributing, it can be observed that they are also similar. Maximums of frequency are close. For 1 step forecast, they are 0.65 on monitoring image and 0.58 on forecasting image. For 2 step forecast, they are 0.73 on monitoring image and 0.59 on forecasting image. This result can also be shown on Table 1. Statistical characters for monitoring and forecasting images are close, such as minimum, maximum and standard Error.

(2) On forecasting for the details, 1 step is better than 2 steps. Directly observing Figs. 1 and 2, excepting some parts, the forecasting image better reflect the details of monitoring image on 1 step. In Fig. 3, the forecasting error with maximum of frequency is different. For 1 step, it is 0.09; for 2 steps, it is 0.15. In Table 1, the mean forecasting error is 0.10 for 1 step, and 0.17 for 2 steps. All above shows 1 step forecasting reflects more details of the original image than 2 step forecasting.

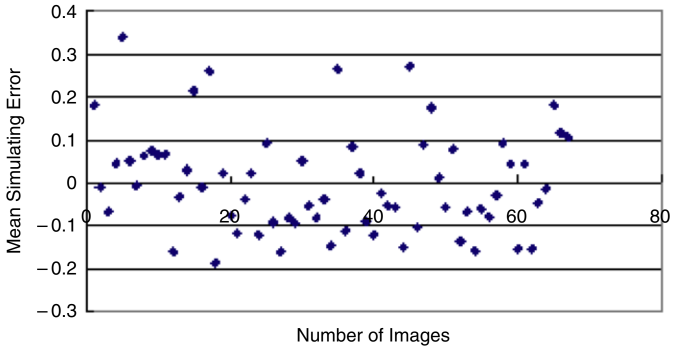

Fig. 4. Scatter plot of mean simulating errors.

Table 1
Statistical results for monitoring, forecasting and forecasting error images.

VTCI image	Minimum	Maximum	Mean	Standard error
2006/4/First ten days (Monitoring)	0	1	0.72	0.89
2006/4/First ten days (Forecasting)	0	0.88	0.61	0.051
2006/4/Middle ten days (Monitoring)	0	1	0.73	0.12
2006/4/Middle ten days (Forecasting)	0	0.82	0.57	0.046
2006/4/First ten days (Forecasting error)	−0.096	0.33	0.10	0.059
2006/4/Middle ten days (Forecasting error)	−0.038	0.35	0.17	0.052

Table 2
Statistics of mean simulating errors.

Mean simulating error	Number of images	Percent (%)
$\|\sigma\| < 0.1$	44	65.7
$0.1 < \|\sigma\| < 0.2$	18	26.9
$\|\sigma\| > 0.2$	5	7.5

Because the structure of models is analyzed from 36 pixels, the AR(1) model is certainly fitting for the series of 36 pixels according to the theory of ARIMA model. However, it does not know whether the AR(1) models fit for the other pixels in the area or not. Thus, the historical data (before the last ten days of March in 2006) is simulated using the AR(1) model. By comparing the simulating data with historical data, the simulating precision can be studied.

Fig. 4 is the scatter plot for mean simulating error. Mean simulating error is the average difference between a monitoring image and a simulating image. It can be found that most of the mean simulating error is in the range of −0.2 to 0.2, only 5 simulating errors are beyond this range. Table 2 is the Statistics for Mean Simulating Errors. It shows that about 65.7% of the number of mean simulating errors is in the range of −0.1 to 0.1, about 92.6% in the range of −0.2 to 0.2, which indicates that AR(1) model fits for the VTCI series.

4. Conclusions

Here, the ARIMA models are developed to simulate the VTCI series, and forecast their changes in the future. The ARIMA models are developed for each pixel according to the steps of modeling. The important step of modeling involves the identification of the structure of the model. The structure of models for each pixel can be analyzed from 36 pixels where the weather stations are situated. It can be found that the AR(1) models are suitable for all VTCI time series of 36 pixels. So, the AR(1) model are used to each pixel of whole area, and the forecast is done with 1–2 steps. The forecasting accuracy of 1 step is better than that of 2 steps. The VTCI series are simulated by the selected best models. Comparing the simulating data with the historical data, the mean errors of simulating are about 65.7% within −0.1 to 0.1, which shows that historical VTCI series are better simulated by AR(1) models. All results as above demonstrate that the AR(1) models are suitable for VTCI series.

In recent years, the application of more indexes and more models is a study direction for drought prediction. Considering the factors affecting agricultural drought, colligating the results of forecasting for the year of drought disaster, SPI (the standardized precipitation index) series, and VTCI series, and combining medium or long range products of weather forecast, the forecasting system for Guanzhong Plain can be established, which will have high forecasting power. The study of this paper lays the foundation for the system of drought forecasting.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant Nos. 40871159 and 40571111), China National High Technology Research and Development Program (Grant No. 2007AA12Z139), the China National

Support Program of Science and Technology (2006BAD10A01), and the European Commission (Call FP7-ENV-2007-1 212921) as part of the CEO-AEGIS project.

References

[1] S. Schubert, R. Koster, M. Hoerling, Predicting drought on seasonal-to-decadal time scales, Bulletin of the American Meteorological Society 88 (2007) 1625–1630.
[2] L.Ch. Song, ZH.Y. Deng, An.X. Dong, Drought, China Meteorological Press, Beijing, 2003, 22 pp.
[3] V.K. Boken, A.P. Cracknell, R.L. Heathcote, Monitoring and Predicting Agricultural Drought—A Global Study, Oxford University Press, New York, 2005.
[4] F.N. Kogan, Application of vegetation index and brightness temperature for drought detection, Advances in Space Research 15 (1995) 9–100.
[5] T.R. McVicar, D.L.B. Jupp, The current and potential operational use of remote sensing to aid decisions on drought exceptional circumstances in Australia: A review, Agricultural System 57 (1998) 399–468.
[6] A.J. Peters, E.A. Walter-Shea, J. Lei, M.R. Svoboda, Drought monitoring with NDVI-based standardized vegetation index, Photogrammetric Engineering and Remote Sensing 65 (1) (2002) 71–75.
[7] X. Song, G. Saito, M. Kodama, et al., Early detection system of drought in East Asia using NDVI from NOAA/AVHRR data, International Journal of Remote Sensing 25 (16) (2004) 3105–3111.
[8] P.X. Wang, J.Y. Gong, X.W. Li, Vegetation temperature condition index and its application for drought monitoring, Geomatics and Information Science of Wuhan University 26 (2001) 412–418 (in Chinese).
[9] P.X. Wang, Zh.M. Wan, J.Y. Gong, X.W. Li, J.D. Wang, Advances in drought monitoring by using remotely sensed normalized difference vegetation index and land surface temperature products, Advances in Earth Sciences 18 (4) (2003) 527–533 (in Chinese).
[10] Z. Wan, P. Wang, X. Li, Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA, International Journal of Remote Sensing 25 (1) (2004) 61–72.
[11] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1976, 525 pp.
[12] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175.
[13] K. Yurekli, K. Kurunc, F. Ozturk, Application of linear stochastic models to monthly flow data of Kelkit stream, Ecological Modeling 183 (2005) 67–75.
[14] W.B. Hu, S.L. Tong, K. Mengersen, D. Connell, Weather variability and the incidence of cryptosporidiosis: Comparison of time series Poisson regression and SARIMA models, Annals Epidemiology 17 (2007) 679–688.
[15] D.A.K. Fernando, A.W. Jayawardena, Generation and forecasting of monsoon rainfall data, in: Proceedings of the 20th WEDC Conference, Colombo, Sri Lanka, 1994, pp. 310–313.
[16] R. Modarres, Streamflow drought time series forecasting, Stochastic Environmental Research and Risk Assessment 21 (2006) 223–233.
[17] Editing Committee on Brief Actual Records of Historical Natural Disasters for Shaanxi Brief Actual Records of Historical Natural Disasters for Shaanxi, China Meteorological Press, Beijing, 2002, 271 pp (in Chinese).
[18] B.L. Bowerman, R.T. O'Connell, Forecasting and Time Series: An Applied Approach, China Machine Press, Beijing, 2003, 521 pp.
[19] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting and Control, third ed., Post & Telecommuncations Press, Beijing, 2005, 327 pp.
[20] G.E. Ljung, G.E.P. Box, On a measure of lack of fit in time series models, Biometrika 65 (1978) 297–303.