Nonlin. Processes Geophys. Discuss., 1, 841–876, 2014
www.nonlin-processes-geophys-discuss.net/1/841/2014/
doi:10.5194/npgd-1-841-2014

Nonlinear Processes in Geophysics
Discussions
Open Access

An improved ARIMA model for hydrological simulations

H. R. Wang[1], **C. Wang**[1], **X. Lin**[2], **and J. Kang**[2]

[1]College of Water Sciences Key Laboratory for Water and Sediment Sciences Ministry of Education, Beijing Normal University, 19 Xinjiekouwai Street, Beijing, 100875, China
[2]College of Mathematic Sciences, Beijing Normal University, 19 Xinjiekouwai Street, Beijing, 100875, China

Received: 20 March 2014 – Accepted: 1 April 2014 – Published: 30 April 2014

Correspondence to: C. Wang (chengw@knights.ucf.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union & American Geophysical Union.

Abstract

Auto Regressive Integrated Moving Average (ARIMA) model is often used to calculate time series data formed by inter-annual variations of monthly data. However, the influence brought about by inter-monthly variations within each year is ignored. Based on the monthly data classified by clustering analysis, the characteristics of time series data are extracted. An improved ARIMA model is developed accounting for both the inter-annual and inter-monthly variation. The correlation between characteristic quantity and monthly data within each year is constructed by regression analysis first. The model can be used for predicting characteristic quantity followed by the stationary treatment for characteristic quantity time series by difference. A case study is conducted to predict the precipitation in Lanzhou precipitation station, China, using the model, and the results show that the accuracy of the improved model is significantly higher than the seasonal model, with the mean residual achieving 9.41 mm and the forecast accuracy increasing by 21 %.

1 Introduction

Hydrological processes are complicated; they are influenced by not only deterministic, but also stochastic factors (Wang et al., 2006). The deterministic change in a hydrological process is always accompanied by the stochastic change. Generally speaking, determinism includes periodicity, tendency, and abrupt change. In fact, a strict deterministic hydrological process is rare. Stationary time series has been widely used in hydrological data assimilation and prediction to tackle the stochastic factors in hydrological processes. From the point of view of stochastic processes, hydrological data series usually comprises trend term and stationary term. The basic idea of Auto Regressive Integrated Moving Average (ARIMA) model, one of the most commonly used time series model, is to remove the trend term of series by difference elimination, so that a nonstationay series is transformed into a stationary one. Many researchers have used ARIMA model for the analysis of hydrological process without considering the effects of seasonal factors (Jin et al., 1999; Niua et al., 1998; Toth et al., 1999). However, most studies (Ahmad et al., 2001; Lehmann et al., 2001; Qi et al., 2006) neglected stationary test and the influence from inter-monthly variation (IM variation) within a year. In this paper, the seasonal ARIMA model is improved by removing the effect of seasonal factors, and the improved model is tested through a case study. The paper is structured as follows: the ARIMA model is introduced first, followed by the introduction of the issues in the currently existing ARIMA model and our proposed methods to improve it. A case study is conducted and discussion is addressed finally.

2 ARIMA model

A hydrological time series $\{f(t)t = 1,2,\cdots,n\}$ can be divided into stationary and nonstationary time series. Given that there are essentially no strictly deterministic hydrological processes, the analysis of hydrological data by means of nonstationary time series is of importance, among which ARIMA model is one of the choices available.

2.1 ARIMA model

In ARIMA model, ARIMA(p,q) is defined as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q} \tag{1}$$

where the real parameters $\varphi_1, \varphi_2, \cdots$, and φ_p are called autoregressive coefficients, the real parameters $\theta_j (j = 1,2,\ldots,q)$ are moving average coefficients, and u_t is an independent white noise sequence, i.e. $u_t \sim N(0,\sigma^2)$. Usually the mean of y_t is zero; if not, $y'_t = y_t - \mu$ is used in the model.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

NPGD

1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

Lag operator (B) is then introduced, thus

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \quad (2)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q. \quad (3)$$

Then the model can be simplified as

$$\phi(B) y_t = \varphi(B) u_t. \quad (4)$$

If y_t are nonstationary, we can obtain the stationary sequence z_t by means of difference, i.e.

$$z_t = (1 - B)^{\mathrm{d}} y_t = \nabla^{\mathrm{d}} y_t \quad (5)$$

where d is the number of regular differencing. Then the corresponding model (ARIMA(p, d, q)) of y_t can be built (Box et al., 1997).

2.2 Seasonal ARIMA(p, d, q) model

A hydrological time series has obviously seasonal (quasi-periodic) variation (Box et al., 1967). For monthly data, consider whether the autocorrelation coefficients are significantly different from 0 with the lag of 12, 24, and 36, and so on; while for seasonal (quarterly) data, consider whether the autocorrelation coefficients are significantly different from 0 with a lag of 4, 8, 12, and so on.

A seasonal ARIMA model can be built with the following procedure. First, difference is applied to eliminate property of season. To obtain the stationary sequence, a model can be built as follows:

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d (1 - B^s)^D y_t = \theta_q(B)\Theta_Q(B^s) u_t \quad (6)$$

where P is the seasonal autocorrelation coefficient, and Q is the seasonal moving average order.

NPGD
1, 841–876, 2014

2.3 Application of ARIMA model

The procedure of calculating ARIMA model is given by the flowchart in Fig. 1 which involves the following steps:

1. Stationary identification. In this step, whether the hydrological sequence is stationary is identified with scatter diagram, autocorrelation function diagram, and partial correlation function diagram.
2. Stationary treatment. Difference or logarithmic transformation is often used for stationary treatment.
3. Identification of the order of ARIMA model by means of autocorrelation and partial correlation function.
4. Parameter estimation (Chen et al., 2004) using maximum-likelihood method.
5. White noise test for residual sequence.

If the residual sequence is not a white noise, some useful information cannot be extracted. The method is illustrated as follows.

Null hypothesis: $H_0 : \mathrm{corr}(e_t, e_{t-k}) = 0 \forall kt$

Alternative hypothesis: $H_1 : \mathrm{corr}(e_{t_0}, e_{t_0-k_0}) \neq 0 \exists k_0 t_0$

The autocorrelation of the data series is measured by the autocorrelation coefficient which is defined as

$$r_k = \frac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{n} e_t^2} \quad (k = 1, 2, \cdots, m) \tag{7}$$

where n is the number of cases, m is the maximum lag. If n is very large, m is $\left[\frac{n}{10}\right]$; if n is very small, m becomes $\left[\frac{n}{4}\right]$. When $n \to \infty$, $\sqrt{n} r_k \sim N(0,1)$.

NPGD
1, 841–876, 2014

The test statistics is given by

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}. \tag{8}$$

Given the degree of confidence of $1-\alpha$, if

$$Q < \chi_\alpha^2(m-p-q) \tag{9}$$

then the null hypothesis is accepted.

6. Hydrological forecasting.

The linear least variance is usually applied for rainfall–runoff prediction. In general, based on observation values of n points, the values of future points of $n+L$ can be estimated (Kohn et al., 1986).

3 Improvement of conventional ARIMA model

The seasonal ARIMA model only deals with time series which arranges in turn with a certain time interval or step, e.g. a month. The seasonal ARIMA model is capable of dealing with the data formed by inter-annual variation (IA variation) with a time interval of month; in this case, however, the information of inter-monthly (IM) variation may be lost. In order to obtain the information of IM variation, the conventional seasonal ARIMA model needs to be improved. In this study, twelve seasonal ARIMA models are built, which are referred to as ARIMA model of IM variation, in order to prevent the loss of IM variation information. However, the effects between adjacent months are ignored in the ARIMA model of IM variation. So a new model needs to be built, in which two kinds of temporal variation (both IA variation and IM variation) are simultaneously considered.

Clustering analysis is used for classifying months and characteristic extraction (Sun et al., 2005). The characteristics refer to values of maximum, minimum, and truncated mean. A linear model is built with dependent variables of hydrological data, and with independent variables of maximum, minimum, and truncated mean. In view of annual variation, an ARIMA model is built for the maximum, minimum, and truncated mean of every class, respectively.

The implementation of the improved ARIMA model involves the following procedure in:

1. Perform clustering analysis on monthly data, and group the months with similar hydrological variation.
2. Find the maximum, minimum, and truncated mean.
3. Build linear regression model and determine the associated parameters.
4. Build ARIMA model for the maximum, minimum, and truncated mean of every class and predict the characteristics.
5. Substitute the predicted characteristics into the model of linear regression and predict the monthly precipitation.

The steps of modeling and forecasting are illustrated in a sketch shown in Fig. 2.

4 Case study

In this section, we are presenting an application of the proposed improved ARIMA model to the precipitation forecasting of Lanzhou precipitation station in Lanzhou, China. Lanzhou is located in the upper basin of Yellow River. It has a continental climate of mid-temperate zone, with an average precipitation of 360 mm and mean temperature of 10 °C. In general, rainfall seasons are May, June, September, and October, while drought occurs in spring and winter. The Lanzhou precipitation station is located in 103.70° E, 35.90° N. The data from 1951 to 2000 is used for parameter estimation and

NPGD
1, 841–876, 2014

the precipitation of every month in 2001is then predicted and compared with the observation values. A comparison is also made between the conventional ARIMA model and the improved model.

4.1 Conventional ARMA modeling

4.1.1 Stationary identification and treatment

The precipitation at the Lanzhou precipitation station from 1951 through 2001 and from 1991 through 2001 are plotted as shown Figs. 3 and 4, respectively. The two figures show less precipitation in winter and spring and more in summer and autumn. Fluctuation occurs to the data during high precipitation seasons. Using power transformation with an order of 1/3, fluctuations at high values are removed and the data become stationary, as shown in Fig. 5.

4.1.2 Identification of the order of model

According to autocorrelation and partial correlation functions, as shown in Figs. 6 and 7, seasonal term with a period of 12 exists. With the difference elimination method, the order of the model can be determined from Figs. 8 and 9, and the following model is obtained.

$$(1 - B^{12})y_t = (1 - \theta_1 B)(1 - \theta_2 B^{12})u_t \tag{10}$$

4.1.3 Parameter estimation

The maximum-likelihood method is used for parameter estimation and the results are listed in Table 1. As shown in Table 1, parameter estimation is statistically significant.

NPGD
1, 841–876, 2014

4.1.4 White noise test

A white noise test is performed for the residual sequence. If the test does not pass, the model needs to be improved. As shown in Table 2, with a significance level of 5 %, the test is passed, i.e. the useful information is extracted and the model is reasonable.

4.1.5 Hydrological forecasting

With the linear least variance and precipitation data from 1951 to 2000, the ARIMA model is used for predicting monthly precipitation of 2001, as is shown in Table 3 and Fig. 10. The mean error is 11.93 mm. As shown in Fig. 10, except for the months of May, June, July, September and October, prediction is basically accurate. However, the model needs further improvement to provide better prediction for those months.

4.2 Establishment of ARIMA model of IM variation

As discussed in Sect. 2.2, the data can be classified into 12 groups associated with each month respectively. Stationary identification, stationary treatment, model identification, parameter estimation and residual test are performed for the 12 groups of data. A total of 12 ARIMA models are built and the results of estimated parameters are shown in Table 4.

In terms of improvement of accuracy, the ARIMA model of IM variation needs fewer samples than the conventional ARIMA model. Besides, hydrological sequence often has singular points, which have a great influence upon the ARIMA model. Therefore, the ARIMA model of IM variation is much better.

4.3 The ARIMA model of IM variation based on clustering and regression analysis

1. Clustering analysis for monthly data.

In order to obtain symmetric data, Box–Cox transformation is applied. As the precipitation has values of zero, Box–Cox transformation (Thyer et al., 2002; Meloun et al., 2005; Ip et al., 2004) is corrected as follows.

$$\text{Data after transformation} = \begin{cases} \frac{(\text{original data}+1)^{\alpha}-1}{\alpha} & \alpha \neq 0 \\ \log(\text{original data}) & \alpha = 0 \end{cases}.$$

After Box–Cox transformation, as shown in Fig. 11, the data are much more symmetric, and the samples don't come from the same population. Result of clustering is shown as Fig. 12, indicating that precipitation sequences can be grouped as three classes.

2. Calculating characteristics of every class.

There are large fluctuations at the maximum of every class, as shown in Fig. 13. Instead, fluctuations at mean and minimum are relatively small; sequences of mean and minimum can be regarded as stationary sequences. Therefore, more information may be extracted from the maximum dataset.

3. Linear regression fitting: the coefficients of each class pass the T test with a significance of 5 %, as shown in Table 5, indicating that the linear model is reasonable.

4. Building ARIMA model for Characteristic variables of each class: nine ARIMA models are built, and the results of parameter estimation are listed in Table 6.

5. Prediction and test of accuracy.

The monthly precipitation of 2001 is predicted using the improved ARIMA model and the prediction results are shown in Table 7 and as Fig. 14. Table 7 shows the mean errors of the conventional ARIMA model and the improved model are 11.93 and 9.41 mm, respectively. Except for April, July, September and October, the relative errors of the improved ARIMA model are very small and catches the correct trend overall. The conventional ARIMA model gives accurate prediction for January, February, August and November, but the predictions of other months are far away from the observations.

From Tables 3 and 7 and Fig. 14, the predicted values of July from the two models are almost equal, but both of them have a large difference from the observed values; it is the same case for September. After a closer look at the data, we find that the mean precipitation amount of July is 63.3 mm with a relative prediction error of 4.8 %, and that the precipitation amounts of 38 years are much higher than the value of 2001. Similarly, the mean precipitation of September is 44.99 mm the relative error of the predicted value is 27.8 %, and precipitation of 46 years are lower than the value of 2001. It might be due to the abnormal climate in July and September of 2001 which caused the inaccurate prediction in the two months given that the improved model is mining historical information without the information of predicted month (Liu et al., 2006).

4.4 Comparison with auto-regression (AR) models

Models of AR (24) and AR (36) are used for predicting the monthly precipitation of 2001 for comparative study. The results of each model are presented in Table 8. From Table 8, the mean residual of auto-regression models with an order of 24 and 36, the conventional ARIMA model, and the improved ARIMA model, are 17.05, 17.82, 12.34, and 9.41 mm, respectively, indicating that the improved ARIMA model provides the best prediction. Compared with the conventional ARIMA model, the improved ARIMA model improves the accuracy by 24 %.

5 Results and discussion

Given that both the inter-annual variation and inter-monthly variation of the hydrological data effect the prediction of hydrological time series, it is better to account for both for better prediction. Inter-monthly data may come from different populations as well as nonstationary factors, so the conventional ARIMA model is not effective enough.

NPGD
1, 841–876, 2014

An improved ARIMA model has been built in this paper taking account for both inter-annual and inter-monthly variation of hydrological data. Based on clustering analysis and regression, much more information is extracted from the data series. A case study is conducted for the precipitation of Lanzhou precipitation station with the improved ARIMA model and the comparison with the conventional ARIMA model indicates that the accuracy of the improved ARIMA model is significantly higher than that of the conventional ARIMA model. This improved approach can be applicable to other hydrological processes prediction with time series data, such as runoff, temperature, and so on.

For the improved ARIMA model, some remarking is given as follows:

1. The selection of clustering methods has little effect on prediction. Different clustering methods can be applied. The definition of distance in the hierarchical clustering can be modified (Wang et al., 2005) to obtain better fit.

2. Characteristics value should be constructed by the features of hydrological time series, not limited to the extreme or mean values.

3. Neural network or support vector machine can be used to further improve is the proposed approach.

References

Ahmad, S., Khan, I. H., and Parida, B. P.: Performance of stochastic approaches for forecasting river water quality, Water Res., 35, 4261–4266, 2001.

Box, G. E. P.: Models for forecasting seasonal and nonseasonal time series, in: Spectral Analysis of Time Series, edited by: Harris, B., Wiley, New York, 271–311, 1967.

Box, G. E. P.: Time Series Analysis Forecasting and Control, China Statistics Press, Beijing, China, 1997.

Chen, Y. H. and Li, C. L.: Parameter estimation and forecast in fractional order ARIMA models, Syst. Eng., 22, 87–90, 2004.

Ip, W. C., Wong, H., and Wang, S. G.: A GIC rule for assessing data transformation in regression, Stat. Probabil. Lett., 68, 105–110, 2004.

Jin, J. L., Ding, J., and Wei, Y. M.: Threshold autoregressive model based on genetic algorithm and its application to forecasting the shallow groundwater level, Hydraul. Eng., 27, 51–55, 1999.

Kohn, R.: Estimation, prediction, and interpolation for ARIMA models with missing data, J. Am. Stat. Assoc., 81, 751–761, 1986.

Lehmann, A. and Rode, M.: Long-term behaviour and cross-correlation water quality analysis of the River Elbe, Germany, Water Res., 35, 2153–2160, 2001.

Liu, X. C.: Knowledge discovery and data mining in mathematical modeling, China Sci. Technol. Inf., 21, 128–130, 2006.

Meloun, M., Sáňka, M., and Němec, P.: The analysis of soil cores polluted with certain metals using the Box–Cox transformation, Environ. Pollut., 137, 273–280, 2005.

Niua, X. F., Edmiston, H. L., and Bailey, G. O.: Time series models for salinity and other environmental factors in the Apalachicola estuarine system, Estuar. Coast. Shelf S., 46, 549–563, 1998.

Qi, W. and Zhen, M. P.: Winters and ARIMA model analysis of the lake level of salt Lake Zabuye, Tibetan Plateau, J. Lake Sci., 18, 21–28, 2006.

Shumway, R. H. and Stoffer, D. S.: Time Series Analysis and Its Applications, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

Sun, L. H., Zhao, Z. G., and Xu, L.: Study of summer rain pattern in monsoon region of east China and its circulation cause, J. Appl. Meteorol. Sci., 16, 57–62, 2005.

Thyer, M., Kuczera, G., and Wang, Q. J.: Quantifying parameter uncertainty in stochastic models using the Box–Cox transformation, J. Hydrol., 265, 246–257, 2002.

Toth, E. A. and Montanari, B. A.: Real-time flood forecasting via combined use of conceptual and stochastic models, Phys. Chem. Earth B, 24, 793–798, 1999.

Wang, H. R., Ye, L. T., and Liu, C. M.: Problems in wavelet analysis of hydrologic series and some suggestions on improvement, Prog. Nat. Sci., 17, 80–86, 2007.

Wang, Y., Yin, L. Z., and Zhang, Y.: Judging model of fuzzy optimal dividing based on improved objective function clustering method, Math. Pract. Theory, 35, 142–147, 2005.

Table 1. Results of parameter estimation of ARMA model.

Parameter	Estimated value	Standard deviation	T test	Tail probability
θ_1	−0.16379	0.03959	−4.14	< 0.0001
θ_2	0.93434	0.02117	44.14	< 0.0001

NPGD
1, 841–876, 2014

Table 2. Autocorrelation coefficients of the residuals of the ARIMA model.

Order	χ^2 statistic	Degree of freedom	Tail probability	Autocorrelation function					
6	0.770	4	0.943	0.000	−0.007	−0.018	0.021	−0.007	0.020
12	6.910	10	0.734	0.013	0.014	0.012	−0.043	0.086	−0.019
18	13.400	16	0.643	0.092	0.014	0.031	−0.004	0.021	0.020
24	16.810	22	0.774	0.042	0.007	−0.022	−0.026	−0.032	0.039
30	20.650	28	0.840	0.050	−0.031	−0.048	0.003	0.018	0.008
36	28.100	34	0.752	0.045	0.018	0.064	−0.044	0.036	0.044
42	30.900	40	0.849	0.057	−0.015	0.019	0.023	0.006	−0.001
48	52.940	46	0.224	−0.012	0.040	−0.022	0.032	−0.079	−0.156

NPGD
1, 841–876, 2014

Table 3. Predicted values of monthly precipitation in 2001 (mm).

Month	1	2	3	4	5	6	7	8	9	10	11	12
Predicted	0.00	0.00	5.38	11.99	31.26	41.28	64.88	71.82	37.98	20.15	0.00	0.00
Observed	2.80	1.90	0.00	22.20	11.10	33.00	39.50	69.80	82.00	5.20	1.90	0.90
Residual	2.80	1.90	−5.38	10.21	−20.16	−8.28	−25.38	−2.02	44.02	−14.95	1.9	−0.90

NPGD
1, 841–876, 2014

Table 4. Results of estimated parameters from ARIMA model of IM variation.

Month	Model	ML parameter estimation
1	$(1-\alpha B)y_t = (1-\beta B)u_t$	$\alpha = -0.95, \quad \beta = -0.97$
2	$(1-\alpha B^2)y_t = u_t$	$\alpha = -0.49$
3	$y_t = (1-\beta B)u_t$	$\beta = 0.38$
4	$y_t = (1-\beta_1 B-\beta_2 B^2)u_t$	$\beta_1 = 0.27, \quad \beta_2 = -0.22$
5	$y_t = (1-\beta B^2)u_t$	$\beta = -0.3$
6	$y_t = (1-\beta B)u_t$	$\beta = -0.32$
7	$y_t = (1-\beta B^2)u_t$	$\beta = -0.3349$
8	$(1-\alpha B)y_t = (1-\beta B)u_t$	$\alpha = -0.182, \quad \beta = -0.0528$
9	$(1-\alpha B)y_t = (1-\beta B)u_t$	$\alpha = 0.956, \quad \beta = 0.469$
10	$y_t = (1-\beta B)u_t$	$\beta = 0.32$
11	$(1-\alpha B)y_t = (1-\beta B)u_t$	$\alpha = 0.681, \quad \beta = 0.741$
12	$(1-\alpha B)y_t = (1-\beta B)u_t$	$\alpha = 0.650, \quad \beta = 0.766$

NPGD
1, 841–876, 2014

Table 5. Result of coefficient estimating of linear regression.

Class	Month	Constant term	Characteristic variable		
			Maximum	Mean	Minimum
1	1	0.16	0.09	0.39	0.23
	2	0.21	−0.12	1.21	−0.14
	11	−0.54	0.30	1.51	−0.62
	12	0.16	−0.27	0.89	0.53
2	3	1.92	−0.50	0.46	0.53
	4	−0.39	−0.57	2.33	−0.62
	10	−1.53	1.07	0.21	0.09
3	5	2.17	−0.41	0.22	0.98
	6	−0.19	−0.22	1.49	−0.35
	7	−0.22	0.27	1.05	−0.35
	8	−2.11	1.07	0.24	0.05
	9	0.35	−0.72	2.01	−0.33

NPGD
1, 841–876, 2014

Table 6. Results of parameter estimation of ARIMA model for characteristic variables of every class.

Class	Characteristic variable	ARIMA model	ML parameter estimating		Standard deviation estimating		Value of P
1	maximum	$(1-B)(1-\alpha B)y_t = u_t$	−0.56		0.13		< 0.0001
	mean	$(1-B)y_t = (1-\beta B)u_t$	0.92		0.07		< 0.0001
	minimum	$(1-B)^2 y_t = (1-\beta B)^2 u_t$	0.84		0.09		< 0.0001
2	maximum	$(1-B)y_t = (1-\beta B)^2 u_t$	−0.30		0.14		0.00311
	mean	$(1-\alpha B^2)(1-B)^2 y_t = u_t$	−0.52		0.12		< 0.0001
	minimum	$(1-\alpha B^2)(1-B)^2 y_t = u_t$	−0.64		0.11		< 0.001
3	maximum	$(1-\alpha B^2)(1-B)^2 y_t = u_t$	−0.45		0.13		0.0006
	mean	$(1-\alpha B)^2(1-B)^2 y_t = (1-\beta B^4)u_t$	−0.82	0.81	0.20	0.16	< 0.0001
	minimum	$(1-\alpha B)^2(1-B)^2 y_t = (1-\beta B^4)u_t$	−0.81	0.80	0.12	0.17	< 0.0001

NPGD
1, 841–876, 2014

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

Table 7. Predicted value of monthly precipitation data for 2001.

2001	1	2	3	4	5	6	7	8	9	10	11	12
Predicted	2.54	1.897	0.099	12.32	12.61	33.58	60.26	72.92	32.50	32.03	1.532	0.898
Observed	2.8	1.900	0.000	22.20	11.10	33.00	39.50	69.80	82.00	5.200	1.900	0.900
Error	0.25	0.003	0.099	9.871	1.515	0.582	20.76	3.12	49.50	26.83	0.368	0.002
Relative error	0.09	0.002	0.099	0.445	0.136	0.018	0.526	0.045	0.604	5.160	0.194	0.003

Table 8. Predicted values of AR (24) and AR (36) models.

24 orders	1	2	3	4	5	6	7	8	9	10	11	12
Predicated	0.27	6.40	4.89	5.81	6.49	77.86	22.55	110.5	65.89	55.45	3.90	0.00
Observed	2.80	1.90	0.00	22.20	11.10	33.00	39.50	69.80	82.00	5.20	1.90	0.90
Residual	−2.53	4.50	4.89	−16.3	−4.61	44.86	−16.9	40.72	−16.11	50.25	2.00	−0.90
36 orders	1	2	3	4	5	6	7	8	9	10	11	12
Predicted	0.57	6.40	5.24	7.25	12.05	79.75	20.09	114.5	63.20	58.78	3.79	0.00
Observed	2.80	1.90	0.00	22.20	11.10	33.00	39.50	69.80	82.00	5.20	1.90	0.90
Residual	−2.23	4.50	5.24	−14.9	0.95	46.75	−19.4	44.73	−18.80	53.58	1.89	−0.90

NPGD
1, 841–876, 2014

Table 9. Comparison of mean errors.

	Improved ARIMA	ARIMA	AR(24)	AR(36)
Mean absolute error	9.41	12.34	17.05	17.82

NPGD

1, 841–876, 2014

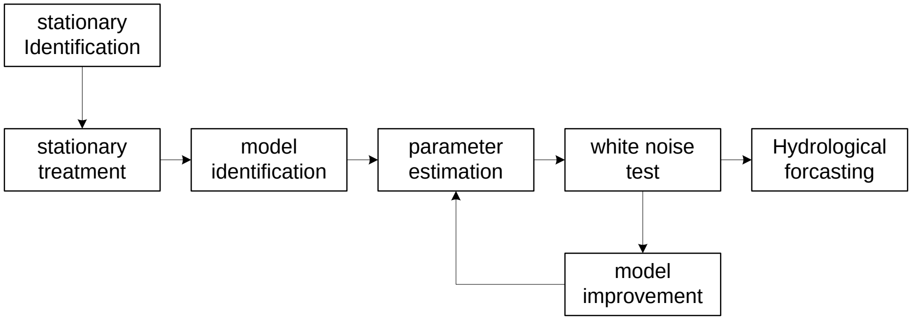

Fig. 1. Procedure of applying ARIMA model.

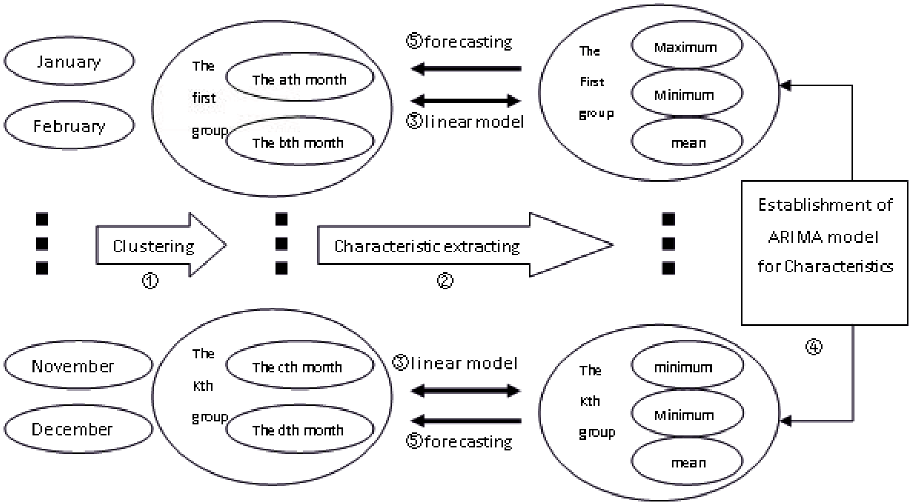

Fig. 2. Prediction steps of ARIMA model of the IM variation based on clustering and regressive analysis.

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

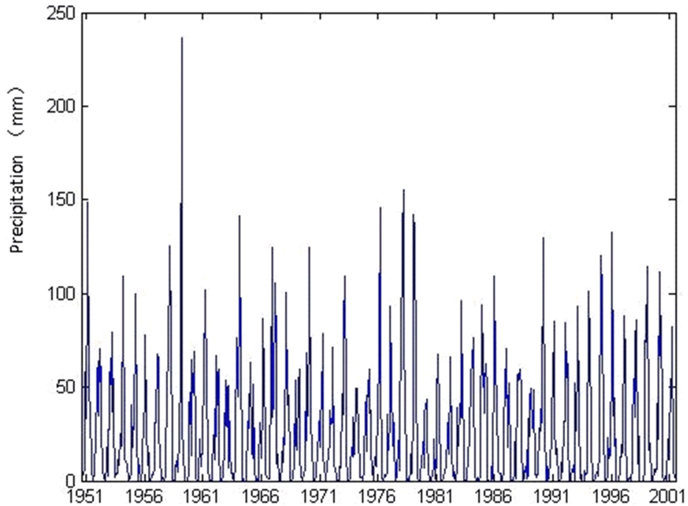

Fig. 3. Line graph of monthly precipitation.

NPGD
1, 841–876, 2014

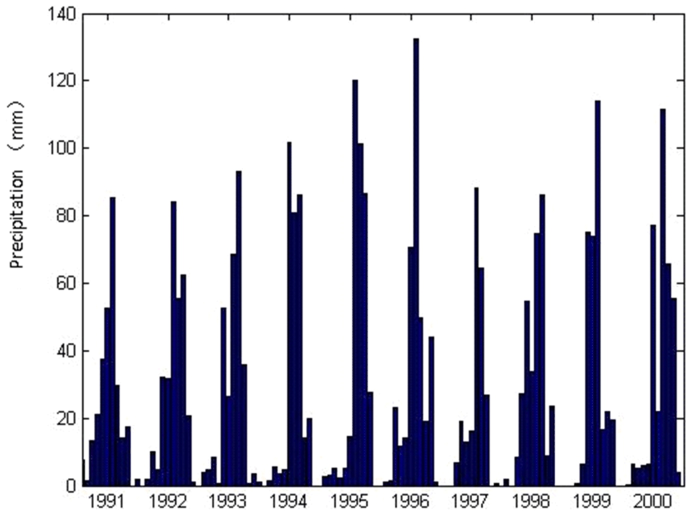

Fig. 4. Columnar section of *f* monthly precipitation.

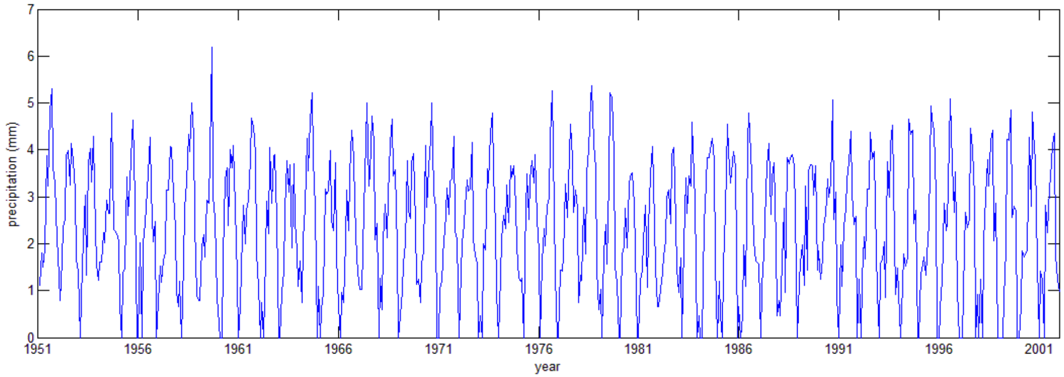

Fig. 5. Line graph of monthly precipitation after power transformation.

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

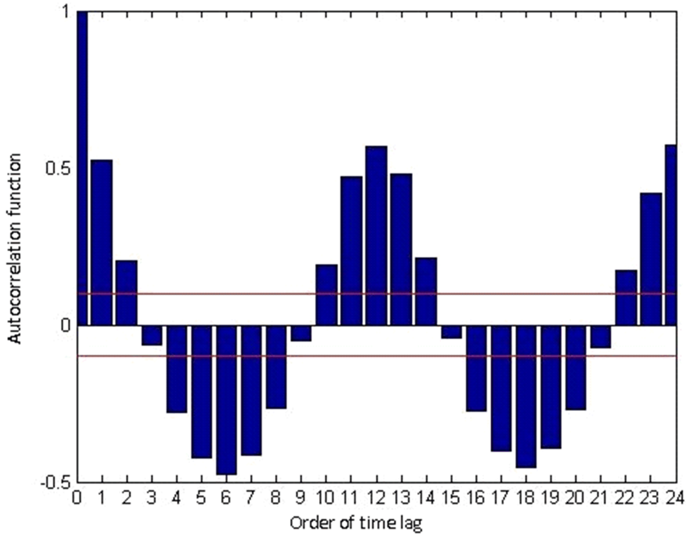

Fig. 6. Autocorrelation function gram of monthly precipitation data.

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

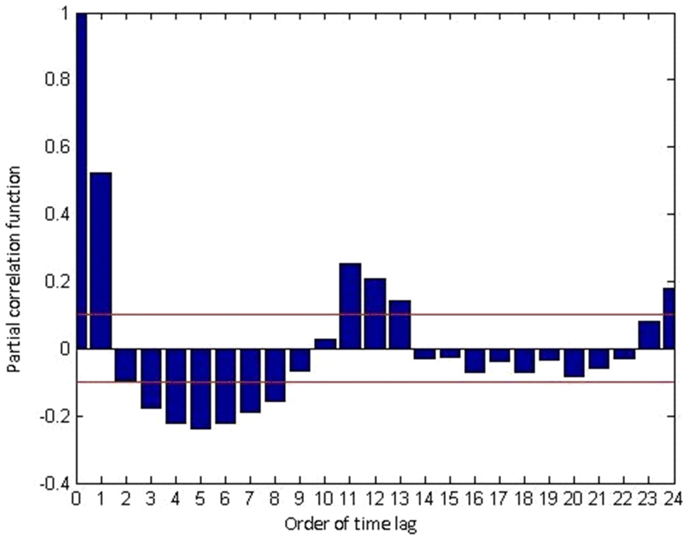

Fig. 7. Partial correlation function gram of monthly precipitation data.

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

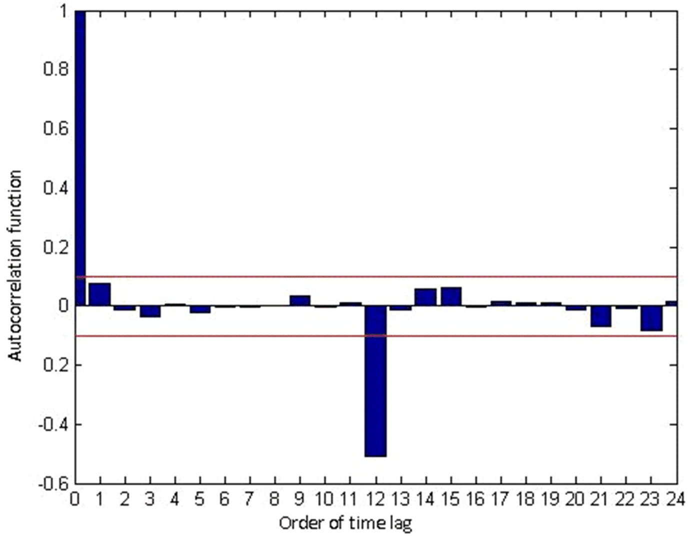

Fig. 8. Autocorrelation function gram of monthly precipitation data after difference.

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

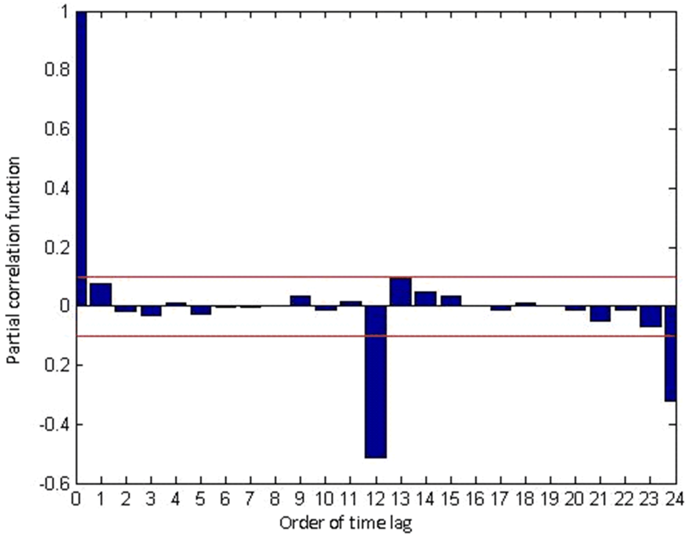

Fig. 9. Partial correlation function gram of data after difference.

NPGD
1, 841–876, 2014

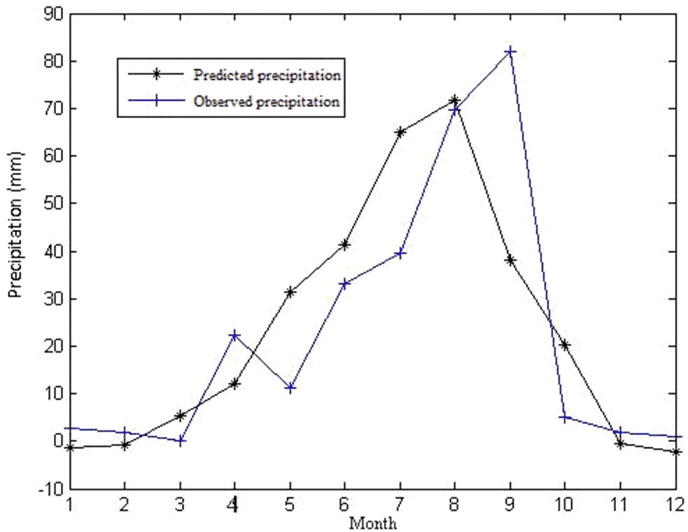

Fig. 10. Comparison of values of predicted and observed precipitation.

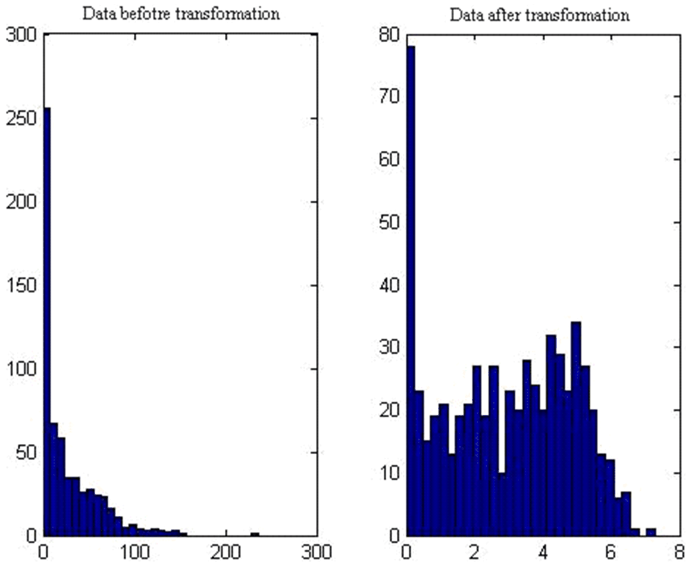

Fig. 11. Histogram of data before and after transformation.

NPGD
1, 841–876, 2014

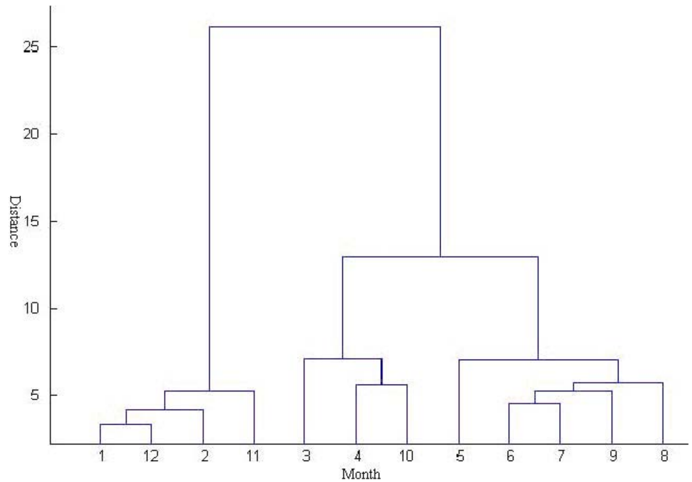

Fig. 12. Result of clustering for monthly precipitation sequence.

NPGD
1, 841–876, 2014

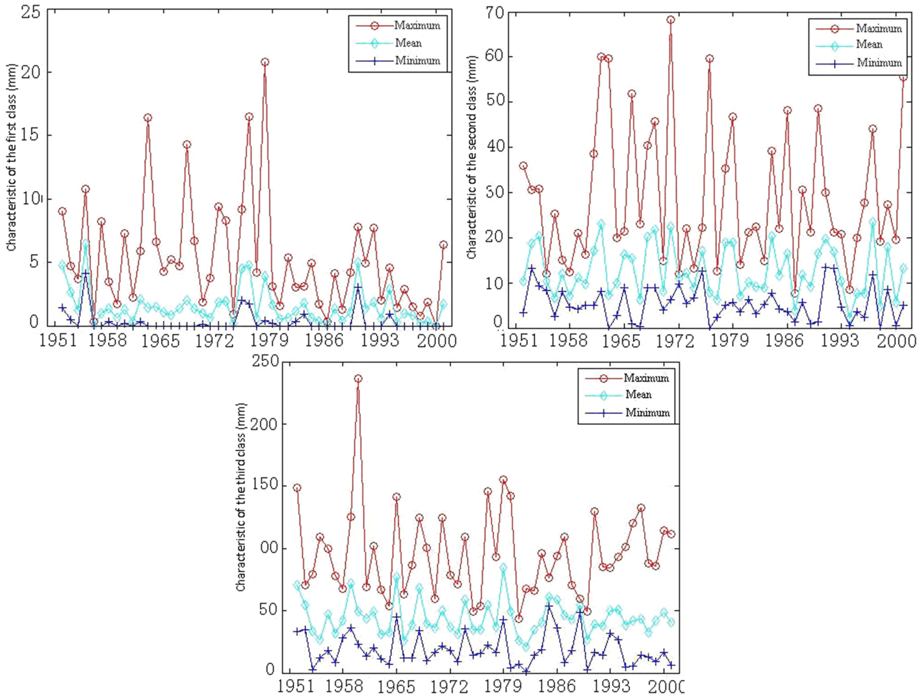

Fig. 13. Annual variation of characteristics Upper: first class; middle: second class; lower: third class.

NPGD

1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀I ▶I
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

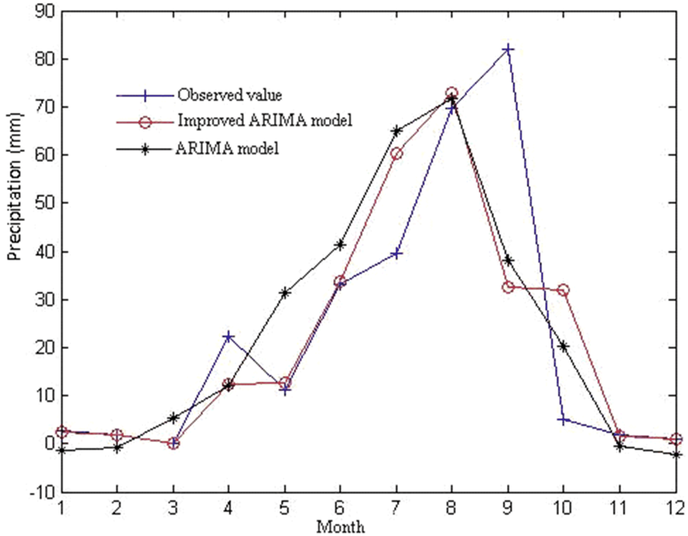

Fig. 14. Comparison between predicted and observed values.

NPGD
1, 841–876, 2014

An improved ARIMA model for hydrological simulations

H. R. Wang et al.