A dynamic Bayesian ARMA model for hydrological drought forecasting

Min Li (limintju@126.com)
Yangzhou University
Mingfeng Zhang
Yangzhou University https://orcid.org/0000-0001-6899-4310
Liming Zhu
Yangzhou University
Zhuoyue Peng
Yangzhou University
Runxiang Cao
North-China University of Water Conservancy and Electric Power: North China University of Water Resources and Electric Power

Research Article

Keywords: Hydrological drought forecast, Bayesian theory, ARMA model, Structural change

Posted Date: April 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1566292/v1

A dynamic Bayesian ARMA model for hydrological drought forecasting

Min Li[1], Mingfeng Zhang [1], Liming Zhu [1], Zhuoyue Peng [1], Runxiang Cao [2]

([1]College of Hydraulic Science and Engineering, Yangzhou University, JiangSu, China, [2]College of Water Resources, North China University of Water Resources and Electric Power, Zhengzhou 450046, China)

Abstract A time series analysis is proposed for an adaptive design framework in nonstationary conditions. In this study, a dynamic Bayesian autoregressive moving average model (t-ARMA) is developed for modelling a standardized runoff index (SRI) time series. The t-ARMA model takes into account the possible change points in the SRI series, and the structure of the model changes with time. Bayesian theory is carried out to estimate the t-ARMA model parameters and locations of change points using a strategy of Markov chain Monte Carlo (MCMC) methods. The results demonstrate that using the Bayesian method to estimate the parameters of the ARMA and t-ARMA models could capture the SRI series characteristics, and the performances of the t-ARMA model are compared with those of the traditional autoregressive moving average (ARMA) model, showing that the t-ARMA model capable of taking the change point of the time series into account is more robust than the traditional ARMA model.

Keywords: Hydrological drought forecast; Bayesian theory; ARMA model; Structural change.

1 Introduction

In recent decades, with global climate warming, hydrological cycle processes such as precipitation and evaporation have intensified, resulting in the frequent occurrence of extreme events such as drought and flooding (Leng et al., 2016). Therefore, drought prediction under climate change has important scientific significance and practical application value.

At present, there are many hydrological forecasting models, such as artificial neural networks (ANNs), support vector machines (SVMs), long short-term memory (LSTM), seasonal autoregressive integrated moving average (SARIMA), generalized autoregressive conditional heteroscedasticity (GARCH) models, and adaptive neural-based fuzzy inference systems (ANFISs). (Li et al., 2018; Milenkovic et al., 2015; Koutroumanidis et al., 2009; Turan and Yurdusev 2014; Yarar 2014; Saeid et al., 2017).

Autoregressive moving average (ARMA) models were proposed by Box et al. (1970) and are widely used in the prediction of hydrological series. (McLeod et al. 1977; Stedinger et al. 1985; Burlando et al. 1993; Valipour 2012; Pektaş and Kerem Cigizoglu 2013; Valipour et al. 2013； Liu et al. 2015). Recently, a large number of studies have been performed to improve the accuracy of the models, and some studies have proven that the accuracy of the model can be improved by preprocessing the time series (Hu et al. 2007; Wu and Huang (2009); Shiri and Kisi 2010; Noori et al. 2011; Wu and Chau 2011; Sang et al. 2012).

Wang et al. (2015) decomposed the original time series through EEMD to obtain a finite number of basic pattern components, and each internal mode function was modelled and predicted according to the variation law of each component. The results showed that the EEMD method can effectively improve the prediction accuracy of the model. Shafaei and Kisi used wavelet analysis to decompose lake level time series into low- and high-frequency components and used ARMA, ANFIS and SVR models to predict each component. The results show that the method combined with wavelet analysis is superior to the three simple models.

In addition, some studies attempt to improve the prediction accuracy by integrating ARMA with other models to form a hybrid model (Yurekli et al., 2004; Valipour, 2016; Gairaa et al., 2016).

Sun et al. (2021) applied the combination of ARMA and LSTM to predict railway passenger flow and optimized the weight of the model through the BP neural network algorithm. The results showed that the prediction effect of the ARMA-LSTM combined model was significantly better than that of the other single models.

Banihabib et al. (2018) used the MARMA-NARX hybrid model to predict reservoir inflows to enhance the accuracy of flow forecasting.

Since the estimation of model parameters is critical to the prediction accuracy of hydrological models, some studies have attempted to adopt improved methods to estimate model parameters. For example, Mohammadi et al. (2006) estimated the ARMA model parameters by the goal programming (GP) method.

With the development of computer technology and Markov Monte Carlo (MCMC) simulation technology, some scholars have attempted to adopt a Bayesian method used in hydrological ARMA models (Bezerra et al., 2012; Liang et al., 2016; Molina and Zazo, 2017). Yoshihiro et al. (2009) applied a strategy of Markov chain Monte Carlo (MCMC) methods to estimate the parameters of the ARMA model. Zhou (2011) analysed and predicted the trust net value by using the autoregression moving average model. The results showed that the model based on Bayesian parameter estimation was more accurate than that based on the least square parameter estimation model. Under the Bayesian framework, Suparman (2020) considered the MCMC method to estimate the parameters of the ARMA model, and the results showed that the MCMC algorithm can correctly estimate the parameters of the ARMA model.

Compared with the time series model based on classical statistical theory, the Bayesian method has unique advantages. First, it can be applied to a wider range of statistical fields. Second, it allows rational use of prior information so that specific problems can be analysed more directly and clearly. Third, this method can obtain the probability distribution of predicted values, and the results obtained by this method are more real and useful. Fourth, compared with traditional statistical methods, it can deal with uncertain factors more clearly and reasonably.

According to the Scientific Decade 2013–2022 of IAHS (Montanari et al., 2013), the hydrological cycle is undergoing substantial changes as a result of global climate change. Due to climate change and human activities, many hydrological variables around the world exhibit nonstationary behaviour (Milly et al., 2008, 2015). Nonstationary hydrologic series analysis has become a focus of research in hydrology and an important means of coping with changing environments. At present, there are few time series models that consider changing environments.

This paper proposes and applies an innovative dynamic ARMA model (t-ARMA), in which the parameters of the model are not fixed so as to sufficiently and robustly capture the nonstationarity of hydrological drought under nonstationary conditions. Focusing on hydrological droughts over the period of 1960~2010 in the Luanhe River basin, a strategy of Markov chain Monte Carlo (MCMC) methods was constructed to estimate the parameters of the t-ARMA models to enhance the accuracy of the models. The performances of t-ARMA and traditional ARMA were compared to further test the robustness and reliability of t-ARMA. This research aims to provide advances in the prediction of time hydrological series under nonstationary conditions.

2 Study area and data

The Luanhe River Basin is located in northeastern China, covering an area of 33700km^2, with geographical coordinates of 39°10'N-42°30'N, 115°30'E-119°15'E (as shown in Figure 1). The basin belongs to the subtropical monsoon area, and is less rainy in the spring, more rainy in the summer and prone to drought in the winter. The average annual temperature of the basin is 5~12 ℃, and the difference in the average annual temperature between the north and south of the basin is 11.5 ℃. The distribution of rainfall in the basin is affected by topography, the range of multiyear average precipitation is between 400~700 mm, the annual distribution is uneven, which is mostly concentrated in summer, and the precipitation in summer is between 200~560 mm. The concentration of rainfall is a remarkable feature of the climate in this area. In recent years, due to the increasing influence of global climate change and human activities in the basin, the precipitation and runoff of the Luanhe River basin have tended to decrease obviously, and drought occurs frequently; since 2000, both extreme drought and successive years of drought have increased in frequency. The frequent and persistent drought in the Luanhe River basin has not only caused a large area of disasters and serious losses but also led to a substantial decline in the inflow of Panjiakou Reservoir in the lower reaches of the basin, resulting in multiple water supply crises in downstream cities.

In this paper, the monthly inflow runoff series of the Panjiakou Reservoir downstream of the basin from 1961 to 2010 is used as the research data, and the SRI index time series is calculated to characterize the drought level of the basin. The data were originally provided by the Hydrology and Water Resource Survey Bureau of Hebei Province.

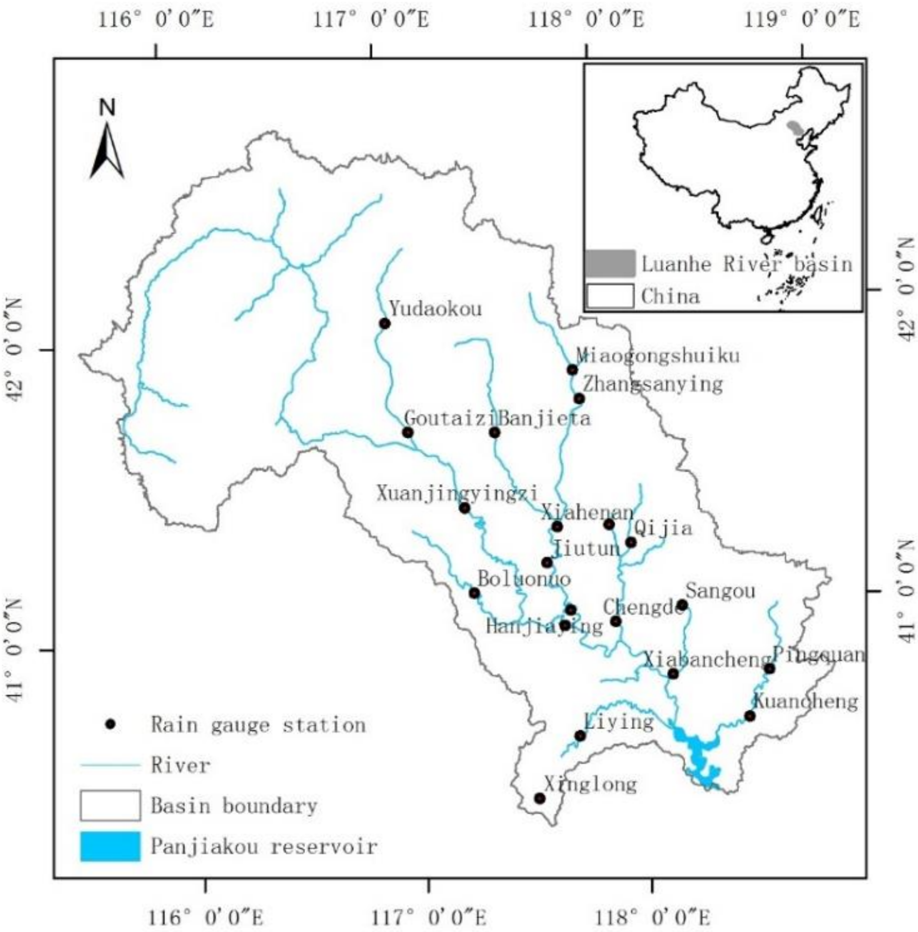

Figure 1 Geographic location map and hydrological station distribution of the Luanhe River Basin

3 Methods

3.1 The ARMA Model

The ARMA (p, q) model of random variable y_t is defined as follows (Said and Dickey, 1984)

$$\phi(B)Y_t=\theta(B)\varepsilon_t \tag{1}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_P B^P \tag{2}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \tag{3}$$

According to Formulas (2) and (3), Equation (1) can be expanded as follows:

$$Y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{4}$$

where $\phi_i(i = 1,2,3,\cdots,p); \theta(i = 1,2,3,\cdots,q)$ are the parameters of the process, p, q represents the order of the ARMA model, y_t follows a normal distribution $N(0, \sigma^2)$ and white noise $\varepsilon_t(t = 1, 2, 3, \cdots, n)$ with mean zero and variance σ^2 is independent (Zeng and Yin, 2016). According to this condition and the stability of the model, only when | ϕ | < 1 can the model have a stable solution.

The precondition of constructing the ARMA (p, q) model is that the target time series y_t must satisfy the stability requirements and the random error term must satisfy the white noise requirements (Hu and Ge, 2021).

3.2 The t-ARMA Model

The change point is the point where the statistical characteristics of the time series change obviously for one or more variables. If there is a sudden change in a series, the time series with unstable structure is divided into different segments according to the change points, and the structure of the observed series is linearly stable in each segment. The estimated values of the parameters of the model are different in different segments (Fang, 2015)

If there are n change points in the time series, according to chronological order, the change points are recorded as $k_i, i = 1,2, \cdots n,$, namely, $1 \leq k_1 < k_2 < \cdots < k_n <$ T, and $k_0 = 1, k_{n+1} =$ T. The change points cut the whole time series into n+1 segments, and the drought index distribution of each time segment is considered to be the same. The expression of the model is as follows:

$$Y_t = \phi_{1t}y_{t-1} + \phi_{2t}y_{t-2} + \cdots + \phi_{pt}y_{t-p} + \varepsilon_t + \theta_{1t}\varepsilon_{t-1} + \theta_{2t}\varepsilon_{t-2} + \cdots + \theta_{qt}\varepsilon_{t-q} \quad (5)$$

3.3 Bayes Method

3.3.1 Bayesian Theory

Bayesian inference as a research tool has been widely used in recent years to cover various research areas such as market forecasting, disease risk prediction, and hydrological sequence frequency estimation (David et al., 2018). The Bayesian statistical method differs from the traditional method in that it combines the prior information of model parameters to make full use of sample information when inferring model parameters (Roberto et al., 2018). According to the subjective understanding of the distribution of parameters before obtaining samples, the prior distribution of parameters is set.

Combining sample information $f(y|\theta)$ and prior information $\pi(\theta)$ to obtain posterior distribution $\pi(\theta|y)$ is the precondition of Bayesian inference (Nan, 2018). The formula for calculating the probability density of Bayesian formula can be represented as:

$$\pi(\theta|y) = \frac{h(y,\theta)}{m(y)} = \frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta} \tag{6}$$

where $h(y,\theta)$ is the joint probability density function of sample X and parameter θ, and the marginal probability density of Y can be calculated as follows:

$$m(y) = \int h(y,\theta)\, d\theta = \int f(y|\theta)\pi(\theta)d\theta \tag{7}$$

3.3.2 Prior Distribution Function Setting

(1) Prior function setting of ARMA Model

First, according to the ARMA (p, q) model, the prior distribution function can be represented as follows:

$$\sum_{i=0}^{p} \phi_i\, y_{t-i} = \sum_{j=0}^{q} \theta_j\, \varepsilon_{t-j}, \varphi_0 = 1, \theta_0 = 1, 1 \le t \le N \tag{8}$$

Here, $\phi_i(\mathrm{i} = 1, 2, 3, \cdots, \mathrm{p})$ obeys a normal distribution, $\theta_j(\mathrm{j} = 1,2,\cdots,\mathrm{q})$ is the process parameter of the model, and ε_t obeys a normal distribution with variance σ^2 and mean zero. Thus, a parameter vector can be defined as:

$$\beta = [\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q,] \tag{9}$$

The prior function can be expressed as:

$$f(\beta, \sigma^2) = f(\sigma^2) \cdot f(\beta/\sigma^2) \tag{10}$$

$$f(\beta/\sigma^2) \propto (\frac{1}{\sigma^2})^{\frac{K}{2}} exp\left\{-\frac{1}{\sigma^2}(\beta - \mu)^T Q(\beta - \mu)\right\} \tag{11}$$

$$f(\sigma^2) \propto (\sigma^2)^{-(\alpha-1)} exp\left\{-\frac{\gamma}{\sigma^2}\right\} \tag{12}$$

Here, the hyperparameter $\mu \in R^K, \alpha > 0, \gamma > 0$, the parameter σ^2 satisfies the i nverted gamma distribution $\sigma^2 \sim IG(\alpha, \gamma)$, and Q is a positive definite matrix of order K (Zhen and Zhu, 2005).

(2) Prior function setting of t-ARMA Model

The t-ARMA model with change points can be represented as follows:

$$Y_t = \sum_{i=1}^{p} \phi_{it} y_{t-i} + \sum_{j=1}^{q} (\varepsilon_t + \theta_{qt} \varepsilon_{t-q}) \tag{13}$$

The vector is defined as $\mathrm{B} = [\phi_{1t}, \cdots, \phi_{pt}, \theta_{1t}, \cdots, \theta_{qt},]$; the vector of change points is $\mathrm{K} = [k_1, k_2, \cdots, k_m]$. It is assumed that the parameter vectors B, K are independent of each other, and the prior function can be obtained:

$$f(B, K, \sigma^2) = f(B, K/\sigma^2) \cdot f(\sigma^2) = f(B) \cdot f(K) \cdot f(\sigma^2) \tag{14}$$

If the size of the sample is known and the number of change points is also known, then $f(K)$ is also determined as a constant (Wang W and Wang X, 2009), namely:

$$f(K) \propto C \tag{15}$$

For vector B, assuming that its prior distribution obeys the normal prior whose mean is 0 and the prior covariance matrix is Σ, then:

$$f(B) \propto (\Sigma_B)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(B - B_0)^T {\Sigma_B}^{-1}(B - B_0)\right\} \tag{16}$$

Similar to the previous derivation process, the prior distribution of the t-ARMA model is as follows:

$$f(B, K, \sigma^2) \propto (\Sigma_B)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(B - B_0)^T {\Sigma_B}^{-1}(B - B_0)\right\} \cdot f(\sigma^2) \tag{17}$$

3.3.3 Posterior Distribution Inference

(1) The posteriori function of ARMA model

According to Bayesian theory, $\rho(\beta, \sigma^2 \mid y)$ represents the posterior distribution of parameter β. Combining the information provided by the prior distribution $f(\beta, \sigma^2)$ with the data provided by the likelihood function $L(\beta, \sigma^2 \mid \mathrm{y})$, the posterior distribution $\rho(\beta, \sigma^2 \mid y)$ can be obtained. Therefore, a posteriori distribution can be given by using Bayesian theorem as follows:

$$\rho(\beta, \sigma^2 / y) \propto f(\beta, \sigma^2)\ L(\beta, \sigma^2 / y) \tag{18}$$

Here, the likelihood function of the parameter β, σ^2 can be expressed as:

$$L(\beta, \sigma^2\ / y) \propto \frac{1}{\sigma^n} exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - u_t)^2\right\} \tag{19}$$

In this formula：

$$u_1 = \sum_{i=1}^{p}\phi_i\, y_i + \sum_{i=1}^{q}\theta_i\, \varepsilon_{t-i} \tag{20}$$

$$u_t = \sum_{i=1}^{p}\phi_i\, y_i + \sum_{i=1}^{t-1}\theta_i(y_{t-i} - u_{t-i}) + \sum_{i=1}^{q}\theta_i\, \varepsilon_{t-i}, t = 2, \cdots, n \tag{21}$$

According to the prior distribution and likelihood function, the posterior distribution can be obtained as follows:

$$\rho(\beta, \sigma^2\ / y) \propto \frac{1}{\sigma^n} exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - u_t)^2\right\}\ f(\beta, \sigma^2) \tag{22}$$

Therefore, $\rho(\beta, \sigma^2 \mid y)$ is a posterior distribution density function with multiple parameters (Manoel et al., 2012). For the marginal posterior distribution density of the autoregressive coefficient term β of the model, the posterior distribution density function can be integrated on $(0, \infty)$ as follows:

$$\rho(\beta|y) = \int_0^{\infty}\rho(\beta, \sigma^2|y)\, d\sigma^2 \tag{23}$$

Similarly, for the accuracy parameter σ^2 of the model, the marginal posterior distribution density can be integrated with β on R^K through the posterior distribution density function as follows:

$$\rho(\sigma^2|y) = \int_{\beta\epsilon R^K}\rho(\beta, \sigma^2|y)\, d\,\beta \tag{24}$$

(2) The posteriori function of t-ARMA model

The likelihood function for the t-ARMA model can be expressed as follows:

$$L(B, K, \sigma^2\ / y) \propto |\Sigma_B|^{-1} exp\{-(\frac{(Y-XB)^T\Sigma_B{}^{-2}(Y-XB)}{2})\} \tag{25}$$

where Σ_B is a diagonal matrix with diagonal line $\phi_{it}, \theta_{jt}(i = 1,2, \cdots, p; j = 1,2, \cdots, q)$, and $X_t = \left[y_{t-1}, \cdots, y_{t-p}, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q}\right]$, $B = \left[\phi_{1t}, \cdots, \phi_{pt}, \theta_{1t}, \cdots, \theta_{qt},\right]^T$. Then, the posterior density function of the model can be written in the following form according to the known likelihood function and a priori density function (Zeng, 2012):

$$\rho(B, K, \sigma^2\ / y) \propto f(B, K, \sigma^2)\ \ L(B, K, \sigma^2\ / y) \tag{26}$$

$$\rho(B, K\ / y) \propto |\Sigma_B|^{-1} exp\{-(\frac{(Y-XB)^T\Sigma_B{}^{-2}(Y-XB)}{2})\} \cdot f(B, K, \sigma^2) \cdot \tag{27}$$

$$\rho(B, K|y) = \int_0^{\infty}\rho(B, K, \sigma^2|y)\, d\sigma^2 \tag{28}$$

According to the above Bayesian inference, many parameters are involved, and the calculation of the marginal posterior distribution density is very complex, so it is difficult to solve the high-dimensional probability integral by traditional methods. In this regard, Windows Bayesian Inference Using Gibbs Sampling (WinBUGS) and other tools are used in this paper. With the help of Markov chain Monte Carlo (MCMC) simulation, the marginal posterior distribution density is calculated by the Gibbs sampling algorithm, and its approximate value is obtained by iteration to complete Bayesian inference more easily.

3.3.4 Gibbs Sampling

The MCMC method is usually used to solve the above high-dimensional posterior distribution problems, which is a recently developed sampling method for joint distributions of high dimensions (Feng, 2013). This method effectively solves these particularly complex problems in the calculation of high-dimensional probability integrals and has a far-reaching impact on the popularization of Bayesian theory.

The principle is to construct one or more Markov chains with a $\pi(X)$ distribution. When the Markov chain converges, its stationary distribution can be regarded as a posterior distribution, and then the simulation value of the chain can be regarded as an independent sample of the target distribution. Then, the distribution characteristic information of the calculated target can be obtained.

If $\{X_t\}$ satisfies:

$$P\{X_{t+1} = x | X_t, X_{t-1}, \cdots\} = P\{X_{t+1} = x | X_t\} \tag{29}$$

For the MCMC method, the key is the construction of the transfer kernel function, and different transfer kernel functions have different construction methods. Gibbs sampling and the Metropolis–Hastings algorithm are the two main MCMC methods. This paper uses widely used, intuitive and simple Gibbs sampling.

The main ideas of Gibbs sampling are as follows (Geman S and Geman D, 1984):

For the parameter vector $\mathrm{X} = (x_1, x_2, \cdots, x_n)$, X^1 corresponds to the initial value of the first iteration. If the initial value of the i-1 iteration is X^{i-1}, then the steps for the n iterations are as follows: First, x_1^i is extracted from the posterior conditional distribution $f\left(x_1 \middle| x_2^{i-1}, \cdots, x_n^{i-1}\right)$; then, x_2^i is extracted from the posterior conditional distribution $f\left(x_1 \middle| x_1^{i-1}, x_2^{i-1}, \cdots, x_{n-1}^{i-1}\right)$. Then, the above steps are repeated for *i*=(1, 2, ⋯, k), and the samples $X^1, X^2, \cdots, X^k$ are obtained.

4 Results

4.1 Construction of ARMA model

4.1.1 Determination of Model Order

Based on the monthly inflow runoff data of the Panjiakou Reservoir from 1961 to 2010, the hydrological drought index on a 12-month scale can be calculated, and a time series can be established. The sequence is stabilized by first-order differencing and zero-mean processing, and then the stationary sequence is expressed as $md\mathrm{SRI}_t$.

The autocorrelation function (ACF) test, partial autocorrelation function (PACF) test and extended autocorrelation function (EACF) test are used to determine the order of the model. Optimized by the AIC, the order of the model is determined to be ARMA (2, 2). The significance of the model was checked by LB test statistics. The results are shown in Figure 2.

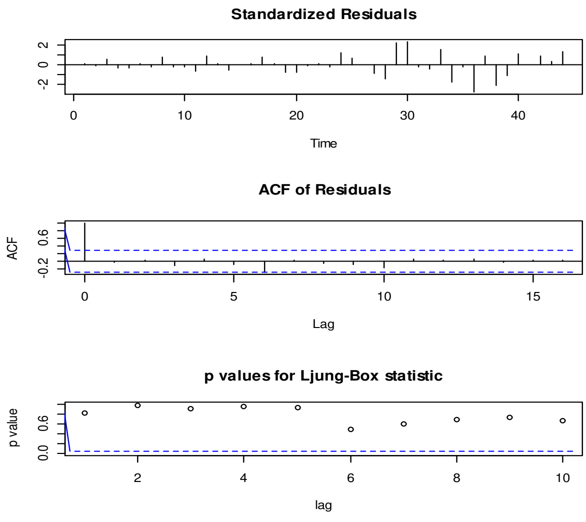

Figure 2 The residual sequence diagram and residual autocorrelation diagram of the ARMA (2, 2)

The autocorrelation map and partial autocorrelation map of the residual of the model ARMA (2, 2) are in the range of two standard deviations, and the p values of the LB statistics under each lag order are all higher than 0.05. The model passed the significance test (Figure 2).

The expression of model ARMA (2, 2) is as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \tag{30}$$

4.1.2 Model Parameter Estimation Based on Bayesian Method

After setting the prior distribution, Gibbs sampling is carried out to obtain the approximate value of the marginal posteriori density of each parameter. The parameter convergence of the model is very important. To ensure the convergence of the Markov chain, WinBUGS is used for multiple iterative analyses. Two different initial values were selected for iteration, with a total of 10000 iterations, of which the first 1000 were used for preheating (Sadia and Boyd, 2018). The iterative trajectory result of the Markov chain is obtained from the same initial values of the two groups, as shown in Figure 3.

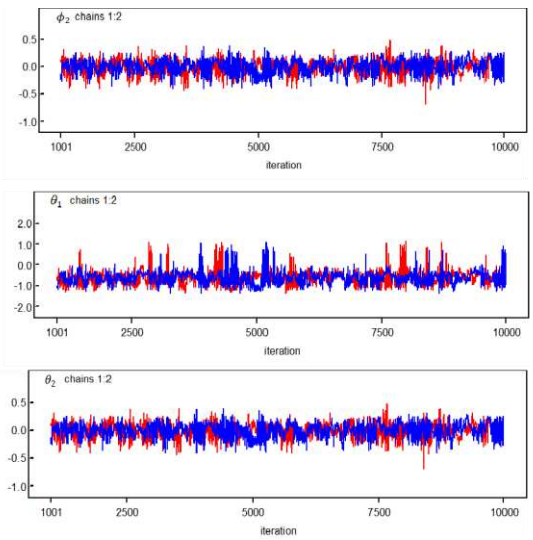

Figure 3 ϕ_2, θ_1, θ_2 Markov chain iteration trajectory

The convergence diagnosis of each parameter is shown in Figure 4.

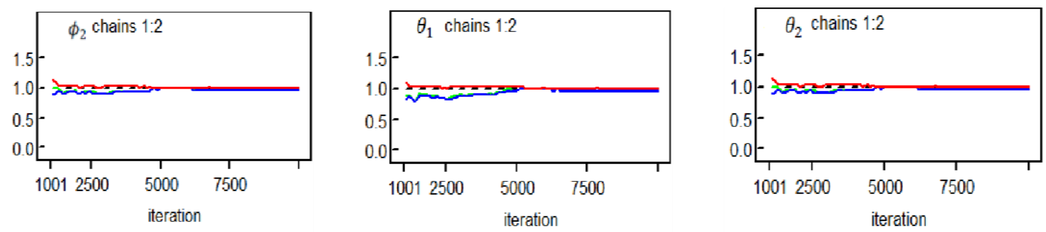

Figure 4 Convergence diagnosis ofϕ_2, θ_1, θ_2

The trajectories of the two chains obtained by iterating with two groups of different initial values and the convergence diagnostic graph of the parameters also tend to coincide.

After judging the convergence of the model, the marginal posterior probability distribution of each parameter can be obtained, as shown in Figure 5.

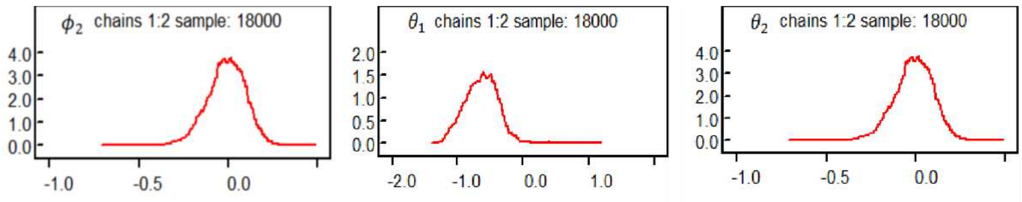

Figure 5 Simulation diagram of the marginal posterior probability distribution ofϕ_2, θ_1, θ_2

Thus, the ARMA (2, 2) model can be constructed as follows:

$$y_t = 0.019 - 0.009y_{t-2} + \varepsilon_t - 0.621\varepsilon_{t-1} - 0.009\varepsilon_{t-2} \tag{31}$$

Based on the ARMA (2, 2) model, the first 44 numbers of zero-mean stationary random sequences $md\mathrm{SRI}_t$ can be used to fit the model to forecast the sequence values in the next 5 years. The simulation results are shown in Figure 6. In the figure, the black dot represents the observed value, the red curve represents the simulated value of the model, the blue curve is the forecast value of drought index, the blue dotted line represents 95% confidence interval of forecasts, and the red dotted line represents 95% confidence interval of simulation.

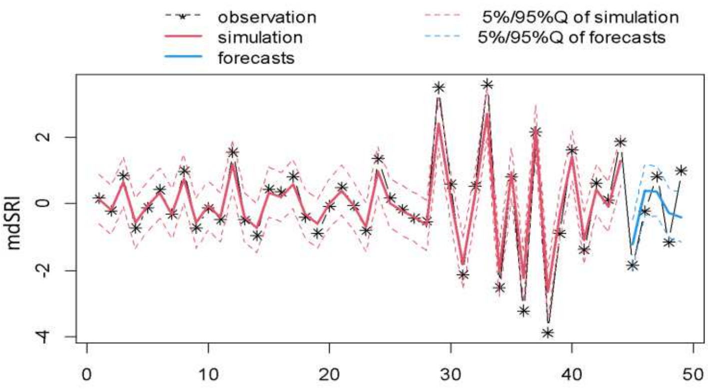

Figure 6 Simulation results of ARMA model

4.2. Construction of t-ARMA model

Since the sequence has 44 samples and the sequence is short, it is reasonable to assume that there is one change point. The WinBUGS Software package was used to estimate the Bayesian parameters of the t-ARMA model. The model is executed for 10000 iterations, and the first 5000 simulated values are discarded to ensure the convergence of the model. It can be seen from the convergence statistical diagnosis graph (Figure 7) that these lines converge to 1. Therefore, it is considered that the parameters of the model are convergent.

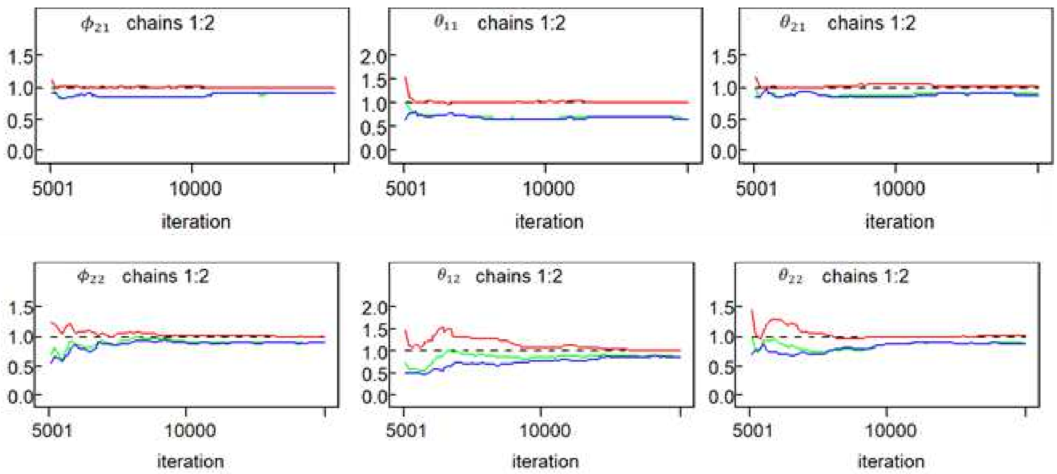

Figure 7 Convergence diagnosis of ϕ_{21}, θ_{11}, θ_{21}, ϕ_{22}, θ_{12}, θ_{22}

The iterative trajectory results of the Markov chain are obtained, as shown in Figures 8 and 9. Figures 8 and 9 show that the two chains tend to coincide, and it is preliminarily judged that the model parameters rs after sampling are convergent.

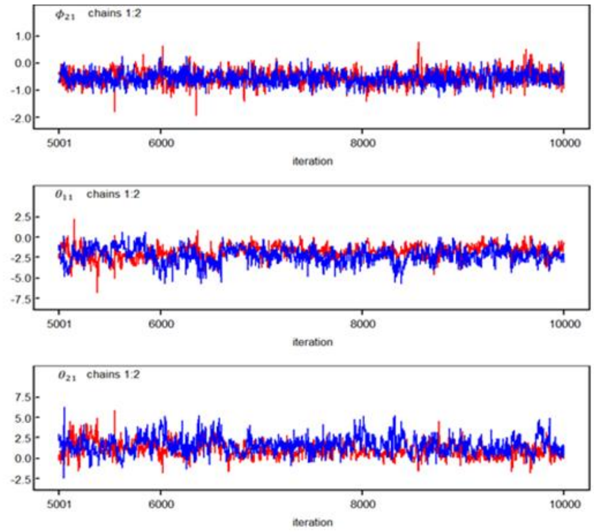

Figure 8 ϕ_{21}, θ_{11}, θ_{21} Markov chain iteration trajectory

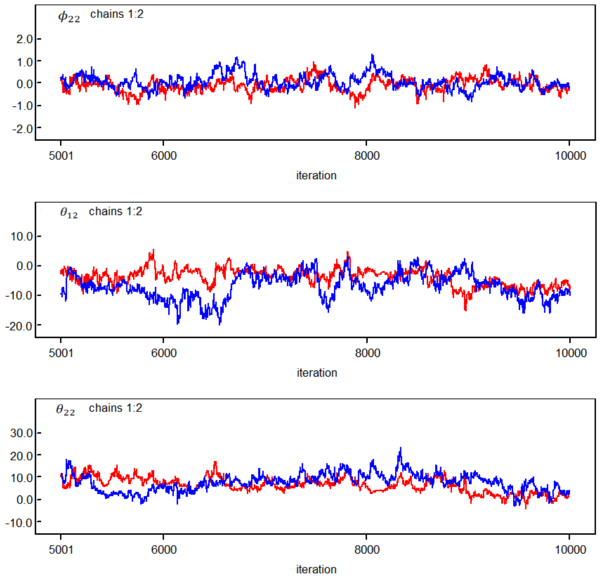

Figure 9 ϕ_{22}, θ_{12}, θ_{22} Markov chain iteration trajectory

The running results of WinBUGS are shown in Table 1. It can be seen from the table that the posterior mean values of ϕ_{21}, ϕ_{22}, $\theta_{11}, \theta_{12}, \theta_{21}$ and θ_{22} are 0.53, -0.03, -2.05, -4.98, 1.27 and 7.86, respectively. The kernel density of the posterior distribution is estimated as shown in Figures 10 and 11.

Table 1 Bayesian parameter estimation results of the t-ARMA (2, 2) model

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
ϕ_{21}	-0.53	0.23	0.01	-0.96	-0.54	-0.07	5001	10000
ϕ_{22}	-0.03	0.31	0.02	-0.64	-0.04	0.62	5001	10000
θ_{11}	-2.05	0.99	0.05	-3.97	-2.05	-0.30	5001	10000
θ_{12}	-4.98	3.53	0.23	-12.92	-4.57	0.83	5001	10000
θ_{21}	1.27	1.06	0.05	-0.56	1.20	3.51	5001	10000
θ_{22}	7.86	3.84	0.25	0.18	8.04	15.43	5001	10000

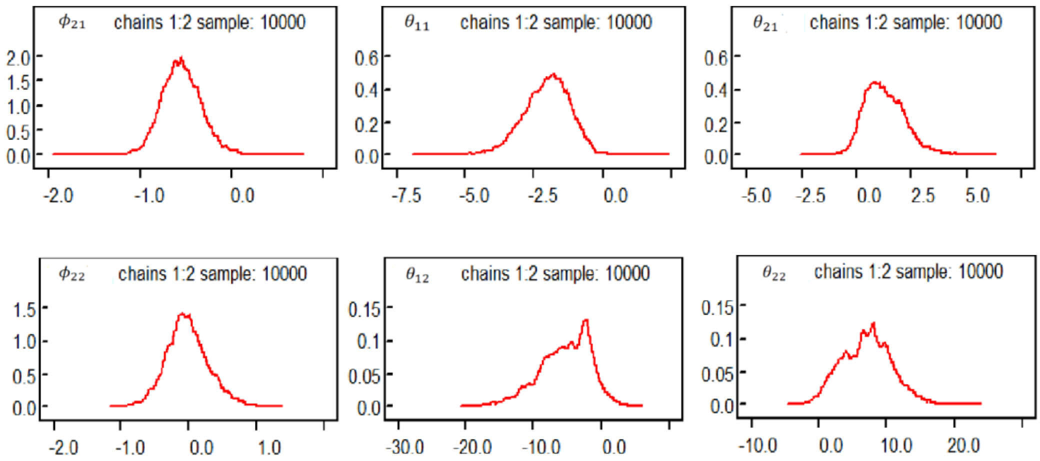

Figure 10 Kernel density of the posterior distribution of ϕ_{21}, θ_{11}, θ_{21}, ϕ_{22}, θ_{12} θ_{22}

According to the marginal posteriori probability distribution graph of change-point k (Figure 11), the maximum posteriori estimate of change-point k is 29.

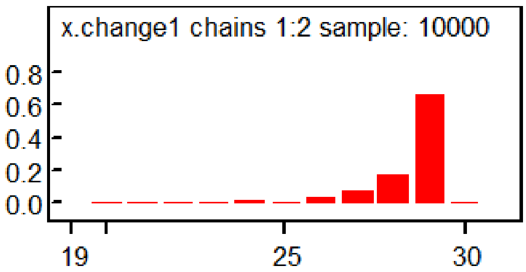

Figure 11 Posteriori probability distribution of change-point k

Based on the established t-ARMA model, the first 44 values of the zero-mean stationary random series $md\mathrm{SRI}_t$ can be used to forecast the sequence values of the next 5 years. The simulation results are shown in Figure 12.

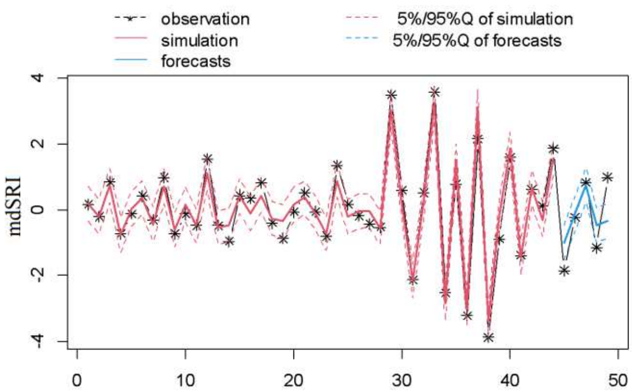

Figure 12 Simulation results of the t-ARMA model

4.3 Evaluation of the Two Models

In this paper, the simulation effects of the two models are compared and evaluated by using the Nash coefficient NSE. The Nash coefficient NSE is a parameter used to evaluate the quality of the model, and its expression is as follows (Hao, 2021):

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_0^t - Q_m^t)^2}{\sum_{t=1}^{T}(Q_0^t - \overline{Q_0})^2} \tag{32}$$

In the equation, Q_0 is the observed value, Q_m is the simulated value, Q^t refers to a value at time t, and $\overline{Q_0}$ refers to the mean of the observed value.

For the Nash coefficient NSE, its value ranges from negative infinity to 1. If the value of the coefficient is closer to 1, the simulation results are better, and the model is more reliable. The closer the NSE is to 0, the closer the simulation results are to the average level of the observations. Namely, the overall results are close to the average level of the observations, but the process simulation error is large and the credibility of the model is not high. The smaller the value, the less reliable the model.

The NSE values of the simulation period and forecast lead time n = 1, 2, ⋯, 5 years of the two models are calculated. The model evaluation results are shown in Table 2. According to Table 2, in the simulation period, the NSE values of the two models are 0.93 and 0.94, which are close to 1 and prove that the simulation results of the two models are good. However, the NSE value of the t-ARMA model is greater than that of the traditional ARMA model, which proves that the simulation effect of t-ARMA in the simulation period is better than that of the traditional ARMA model. In the forecast period, the NSE values of the traditional ARMA model are 0.39, 0.72, 0.56 and 0.39 in the forecast lead time n = 2, 3, 4, and 5 years, and the NSE values of the t-ARMA model are 0.45, 0.80, 0.70 and 0.51, respectively. The NSE values of the two models show a decreasing trend with the extension of the forecast lead time, and the NSE value of the t-ARMA model is larger than that of the traditional ARMA as a whole, indicating that the forecast accuracy of the t-ARMA model is higher. In conclusion, the t-ARMA model is more reliable in both the simulation period and the forecast lead time.

Table 2 NSE values in the simulation period and forecast lead times between the two models

Model	Simulation period	Forecast lead time (year)				
		1	2	3	4	5
ARMA	0.93	/	0.39	0.72	0.56	0.39
t-ARMA	0.94	/	0.45	0.80	0.70	0.51

5 Conclusions and discussion

In recent years, drought has become an increasingly prominent challenge due to global climate change and socioeconomic development and therefore has attracted much attention from the international community as a global environmental problem. How to reasonably forecast the development of drought has important practical significance and application value to ensure the rational utilization of water resources and the sustainable development of the social economy. In this paper, Bayesian theory and the ARMA model are combined to simulate the hydrological drought time series of 44 years in the Panjiakou Reservoir Basin, and the drought development trend in the next 5 years is forecasted. The influence of climate change and human activities may cause change points in hydrological series. ARMA models both with and without considering structural change are established. The simulation and forecast results of the two models are compared and evaluated, and the main conclusions are as follows:

(1) During the simulation period from 1961 to 2004, the NSE values of the ARMA model without considering the change point and the dynamic t-ARMA model considering the change point are 0.93 and 0.94, respectively, and both are close to 1. This shows that the ARMA model based on the Bayesian method can better simulate the hydrological drought series of the Panjiakou reservoir basin; however, the NSE value of the dynamic t-ARMA model is slightly higher than that of the ARMA model without considering change points, which shows that the simulation effect of the t-ARMA model considering change points is better.

(2) For the forecast lead time, the NSE values simulated by the two models tend

to decrease as the forecast time increases. Furthermore, the NSE mean value of the t-ARMA model considering the change point is larger than that of the ARMA model without considering the change point. Therefore, the forecast effect of the t-ARMA model is more robust than that of the ARMA model without considering the change point, and the forecast accuracy of the two models decreases with the extension of the forecast time as a whole.

In summary, the t-ARMA model based on the Bayesian method has high credibility and a better simulation effect. Compared with the ARMA model without considering time variation, it is more suitable for drought simulation in the Panjiakou reservoir basin, and the model is more suitable for short-term drought forecasting.

Declarations:

Ethical Approval: This work meets the ethical and moral requirements.

Consent to Participate: M L. MF Z. LM Z. ZY P and RX C all agreed to participate in the research for the article.

Consent to Publish: M L. MF Z. LM Z. ZY P and RX C all agreed to publish this article.

Authors Contributions:

M L(First Author and Corresponding Author)： Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing-Original Draft;

MF Z：Data Curation, Writing-Original Draft;

LM Z: Visualization, Investigation;

ZY P: Resources, Supervision;

RX C: Visualization, Writing-Review & Editing.

Funding: This work was supported by the Jiangsu Provincial Double innovation Program (No. JSSCBS20211029) and National Natural Science Foundations and Postdoctoral Research Program of Jiangsu Province (No. BK20200942; 2020Z223).

Competing interest: M L. MF Z. LM Z. ZY P and RX C all declare that there is no conflict of interest.

Availability of data and materials: We are grateful to the Hydrology and Water Resource Survey Bureau of Hebei Province for providing runoff data. The data and materials of the research are available.

Reference

[1] Bezerra M , Moala F A , Iano Y . A Bayesian Analysis of Spectral ARMA Model[J]. Mathematical Problems in Engineering,2012,(2012-7-5), 2012, 2012(pt.7):857-868.

[2] Burlando P, Rosso R, Cadavid LG, Salas JD (1993) Forecasting of short-term rainfall using ARMA models. J Hydrol 144(1–4):193–211

[3] David J. Warne,Ruth E. Baker,Matthew J. Simpson. Multilevel rejection sampling for approximate Bayesian computation[J]. Computational Statistics and Data Analysis,2018,124:

[4] FANG Lichong. Research on change-points in Shanghai composite index based on Bayesian inference[D]. Huaqiao University,2015.(in Chinese)

[5] FENG Na. Research on change-points in Shanghai composite index: based on structure mutation AR(p) model with lag coefficient mutation[D].Huaqiao University, 2013. (in Chinese)

[6] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1984 (6): 721-741.

[7] HAO Gairui. Research on non-point source pollution characteristics and model simulation in Shaanxi section of the Hanjiang river basin[D]. Xi'an University of Technology, 2021.DOI:10.27398/d.cnki.gxalu.2021.000048. (in Chinese)

[8] HU Kang, GE Ruiyang. Price forecast of CSI 300 index based on ARMA model[J]. Economy and Management Digest, 2021(18):169-171. (in Chinese)

[9] Hu T, Wu F, Zhang X (2007) Rainfall-runoff modeling using principal component

analysis and neural network. Nord Hydrol 38(3):235–248

[10] José-Luis Molina and Santiago Zazo. Causal Reasoning for the Analysis of Rivers Runoff Temporal Behavior[J]. Water Resources Management, 2017, 31(14) : 4669-4681.

[11] Kacem Gairaa, Abdallah Khellaf,Youcef Messlem,Farouk Chellali. Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach[J]. Renewable and Sustainable Energy Reviews,2016,57: 238–249.

[12] Kadri Yurekli, Ahmet Kurunc,Fazli Ozturk. Application of linear stochastic models to monthly flow data of Kelkit Stream[J]. Ecological Modelling,2004,183(1), 67–75.

[13] Koutroumanidis T, Sylaios G, Zafeiriou E, Tsihrintzis VA (2009) Genetic modeling for the optimal forecasting of hydrologic time-series: application in Nestos River. J Hydrol 368(1–4):156–164

[14] Leng G, Tang Q, Huang S, et al. Assessments of joint hydrological extreme risks in a warming climate in China[J]. International Journal of Climatology, 2016, 36(4):1632-1642.

[15] LI Mei, LI Jing, WEI Zijian, et al. Short-time passenger flow forecasting at subway station based on deep learning LSTM structure[J]. Urban Mass Transit, 2018, 21(11) : 42 － 46, 77．(in Chinese)

[16] Manoel I. Silvestre Bezerra and Fernando Antonio Moala and Yuzo Iano. A Bayesian Analysis of Spectral ARMA Model[J].Mathematical Problems in Engineering, 2012, 2012(Pt.7) : 565894.1-565894.15.

[17] Maryam Shafaei and Ozgur Kisi. Lake Level Forecasting Using Wavelet-SVR, Wavelet-ANFIS and Wavelet-ARMA Conjunction Models[J]. Water Resources Management, 2016, 30(1) : 79-97

[18] McLeod AI, Hipel KW, Lennox WC (1977) Advances in Box-Jenkins modeling: 2. Applications. Water Resour Res 13(3):577–586

[19] Milenkovic M, vadlenka L, Melichar V, et al． SAＲIMAmodelling approach for railway passenger flow forecasting［J］． Transport, 2015, 33(5) : 1113 － 1120．

[20] Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., Dettinger, M.D., Krysanova, V., 2015. On critiques of “Stationarity is Dead: Whither Water Management?”. Water Resour.

Res. 51 (9), 7785–7789.

[21] Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: whither water management? Science 319 (5863), 573–574.

[22] Mohammad Ebrahim Banihabib and Arezoo Ahmadian and Mohammad Valipour. Hybrid MARMA-NARX model for flow forecasting based on the large-scale climate signals, sea-surface temperatures, and rainfall[J]. Hydrology Research, 2018, 49(6) : 1788-1803.

[23] Mohammad Valipour and Mohammad Ebrahim Banihabib and Seyyed Mahmood Reza Behbahani. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir[J].

[24] Mohammad Valipour. Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms[J]. Meteorological Applications,2016,23(1): , 91–100.

[25] Mohammadi K, Eslami HR, Kahawita R (2006) Parameter estimation of an ARMA model for river flow forecasting using goal programming. J Hydrol 331(1–2):293–299

[26] Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., 2013. "Panta Rhei Everything Flows": change in hydrology and society-The IAHS Scientific Decade 2013–2022. Hydrol. Sci. J. 58 (6), 1256–1275.

[27] NAN Jiefeng. Bayesian methods and applications on structural damage and restoring force model parameter identification[D]. Bei Jiaotong Unibersity, 2018. (in Chinese)

[28] Noori R et al (2011) Assessment of input variables determinationon the SVM model performance using PCA,Gamma test, and forward selection techniques for monthly stream flow prediction. J Hydrol 401(3–4):177–189

[29] Pektaş AO, Kerem Cigizoglu H (2013) ANN hybrid model versus ARIMA and ARIMAX models of runoff coefficient. J Hydrol 500:21–36

[30] Roberto Rocchetta,Matteo Broggi,Quentin Huchet,Edoardo Patelli. On-line Bayesian model updating for structural health monitoring[J]. Mechanical Systems and Signal Processing,2018,103:

[31] Rubing Liang et al. Testing a linear ARMA model against threshold-ARMA models: A Bayesian approach[J]. Communications in Statistics - Simulation and Computation, 2016, 46(2) : 1302-1317.
[32] Sadia Farhana and Boyd Sarah and Keith Jonathan M. Bayesian change-point modeling with segmented ARMA model.[J]. PloS one, 2018, 13(12) : e0208927.
[33] Saeid Mehdizadeh and Javad Behmanesh and Keivan Khalili. A comparison of monthly precipitation point estimates at 6 locations in Iran using integration of soft computing methods and GARCH time series model[J]. Journal of Hydrology, 2017, 554 : 721-742.
[34] Said S E , Dickey D A . Testing for unit roots in autoregressive-moving average models of unknown order[J]. Biometrika, 1984, 71(3):599-607.
[35] Sang Y-F, Wang Z, Liu C (2012) Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis. J Hydrol 424–425:154–164
[36] Shiri J, Kisi O (2010) Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. J Hydrol 394(3–4):486–493
[37] Stedinger JR, Lettenmaier DP, Vogel RM (1985) Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. Water Resour Res 21(4):497–509
[38] Sun Yue, Song Xiaoyu, Jin Li-ting, LIU Tong. Prediction of Railway Passenger Flow Based on ARMA-LSTM Combined Model [J]. Computer Applications and Software, 201,38 (12):262-267+273.(in Chinese)
[39] Suparman. Hierarchical Bayesian Choice of Laplacian ARMA Models Based on Reversible Jump MCMC Computation[J]. International Journal of Computational Intelligence Systems, 2020, 13(1) : 310-317.
[40] Turan M, Yurdusev M (2014) Predicting monthly river flows by genetic fuzzy systems. Water Resour Manag 28(13):4685–4697
[41] Valipour M (2012) Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting. J Math Stat 8(3):330–338
[42] Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J Hydrol 476:433–441
[43] WANG Weiguo,WANG Xia. Research on Change-Points in Shanghai Composite

Index Based on Bayesian Inference[J]. Chinese Journal of Management Science,2009,17(03):8-17.DOI:10.16381/j.cnki.issn1003-207x.2009.03.022. (in Chinese)

[44] Wang Wen-chuan et al. Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition[J]. Water Resources Management, 2015, 29(8) : 2655-2675.

[45] Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol 399(3–4):394–409

[46] Wu Z, Huang NE (2009) EMAEmble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 1(1):1–41

[47] Yarar A (2014) A hybrid wavelet and neuro-fuzzy model for forecasting the monthly streamflow data. Water Resour Manag 28(2):553–565

[48] Yoshihiro Ohtsuka and Takashi Oga and Kazuhiko Kakamu. Forecasting electricity demand in Japan: A Bayesian spatial autoregressive ARMA approach[J]. Computational Statistics and Data Analysis, 2009, 54(11) : 2721-2735.

[49] Youcun Liu et al. Analyzing effects of climate change on streamflow in a glacier mountain catchment using an ARMA model[J]. Quaternary International, 2015, 358 : 137-145.

[50] ZENG Guoxiang. Theoretical method and application of catastrophe point detection in time series [D].Yangzhou university,2012. (in Chinese)

[51] ZENG Zhaofa. YIN Siyu. Research on structural mutation of stock market based on Bayesian statistics[J]. Statistics & Decision, 2016(13):160-162.DOI:10.13546/j.cnki.tjyjc.2016.13.042. (in Chinese)

[52] ZHEN Jincheng, ZHU Huiming. Bayesian AR(P) model analysis based on MCMC method[J]. Statistics & Decision, 2005(10X):3. (in Chinese)

[53] Zhou YUAN. Prediction of net worth by ARMA Model based on Bayes Analysis [D]. Beijing Jiaotong University, 2011.