Computers and Mathematics with Applications 61 (2011) 1975–1978

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

Chaotic Bayesian optimal prediction method and its application in hydrological time series

Xiao-Hua Yang [a,*], Ying Mei [a], Dun-Xian She [a], Jian-Qiang Li [b]

[a] The Key Laboratory of Water and Sediment, Ministry of Education, State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, China

[b] Water Resources and Hydropower Planning and Design General Institute, MWR, Beijing 100011, China

ARTICLE INFO

Keywords:
Hydrological time series
Prediction
Embedding dimension
Number of nearest neighbors
Bayesian analysis

ABSTRACT

The embedding dimension and the number of nearest neighbors are very important parameters in the prediction of chaotic time series. To reduce the prediction errors and the uncertainties in the determination of the above parameters, a new chaos Bayesian optimal prediction method (CBOPM) is proposed by choosing optimal parameters in the local linear prediction method (LLPM) and improving the prediction accuracy with Bayesian theory. In the new method, the embedding dimension and the number of nearest neighbors are combined as a parameter set. The optimal parameters are selected by mean relative error (MRE) and correlation coefficient (CC) indices according to optimization criteria. Real hydrological time series are taken to examine the new method. The prediction results indicate that CBOPM can choose the optimal parameters adaptively in the prediction process. Compared with several LLPM models, the CBOPM has higher prediction accuracy in predicting hydrological time series.

1. Introduction

Many hydrological processes, such as runoff, are usually nonlinear, complex, dynamic processes. The simulation of the nonlinear time series turned out to be very difficult with the traditional deterministic mathematic models [1–3]. However, the emergence of chaos theory provides a new way to study highly complex nonlinear systems [4–6], and makes it possible to extract deterministic regulation from the seemingly disordered hydrological phenomenon. With the development of chaos theory and research, many methods have been proposed to predict chaotic time series, which can broadly be divided into two main categories: global method [7] and local method [8]. Farmer and Sidorowich [9] have already proved that the local prediction method is better than the global prediction method. The uncertainties in the determination of the embedding dimension and the number of the nearest neighbors exist in local prediction method. So as to improve the prediction accuracy, the key problem is how to reduce the parameter uncertainties in the prediction processes.

In this study, a parameter set (m, n) is established on the basis of embedding dimension m and the number of nearest neighbors n. A new adaptive local prediction method, chaos Bayesian optimal prediction method (CBOPM), is proposed by choosing optimal structure parameters in the chaos local prediction method and improving the accuracy with Bayesian theory. The monthly discharges time series of Yichang station and Chuntan station at Changjiang river in China are studied to test the prediction efficiency of the CBOPM.

* Corresponding author.
E-mail address: xiaohuayang@bnu.edu.cn (X.-H. Yang).

0898-1221/$ – see front matter Crown Copyright © 2010 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2010.08.041

2. Chaos Bayesian optimal prediction method (CBOPM)

In this paper, we use mean relative error (MRE) and correlation coefficient (CC) indices [1] to select the optimal parameters (m, n). By comparing MRE and CC indices for different (m, n), the optimal parameters can be selected for the local linear method. And then an effective self-learning process with Bayesian processor of forecast (BPF) is used to predict the time series. The new improved local linear method, chaos Bayesian optimal prediction method (CBOPM), is described as follows:

Step 1. Phase-space reconstruction. Determine the embedding dimension m_c [7] and the delay time τ for the original time series. For a scalar time series $x_1, x_2, \ldots, x_N$, the multi-dimensional phase-space Y_i can be reconstructed by Takens' embedding theory, according to

$$Y_i = (x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}),$$

where $i = 1, 2, \ldots, M$, M is the total points number of the phase-space, and $M = N - (m - 1)\tau$, τ is the delay time, m is the dimension of the vector Y_i, called as embedding dimension, and N is the length of the time series.

The delay time is determined by the autocorrelation function method [8] in this study. The autocorrelation function can be described as

$$C(t) = \frac{1}{N-t} \frac{\sum_{i=1}^{N-t} (x_i - \mu)(x_{i+t} - \mu)}{\sigma^2(x)}$$

where $C(t)$ is the autocorrelation coefficient, t is the lag time, μ and σ are the mean and standard variation of the time series, respectively. The delay time τ is selected when the autocorrelation coefficient has dropped to $1 - 1/e$ of its initial value [8].

Step 2. Calculate MRE and CC indices. The definitions of the two indices can be given as follows:

$$\text{MRE} = \frac{1}{p} \sum_{i=1}^{p} |y_i - x_i|/x_i,$$

$$\text{CC} = \frac{\sum_{i=1}^{p} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{p} (x_i - \bar{x})^2 \sum_{i=1}^{p} (y_i - \bar{y})^2}},$$

where p is the number of the time series under investigation. The y_i and x_i are the predicted values and observed values in the time series, respectively, and $\bar{x}$ and $\bar{y}$ are the mean of observed values and predicted values, respectively. Generally, in the above measurement indices, lower values with the MRE indicate better agreement, and higher positive values with the CC indicate better agreement between the observed values and predicted values.

Let the embedding dimension m change from $m_{\min}$ to $m_{\max}$, the number of the nearest neighbor change from $n_{\min}$ to $n_{\max}$ (In this study, $m_{\min}$ and $m_{\max}$ are selected as 2 and m_c, respectively). $n_{\min}$ and $n_{\max}$ are taken from $m + 1$ to $m + 10$, the detail operation technology is the same as Jayawardena's method [10]. $(m_{\max} - m_{\min} + 1) \times (n_{\max} - n_{\min} + 1)$ local linear models can be obtained.

Step 3. Choose the optimal parameters (m_{opt}, n_{opt}). In the selection of the optimal parameters for each prediction step, MRE and CC indices are calculated under different combination of (m, n). The optimal parameter set is chosen according to the optimization criteria when MRE obtains the minimum value and CC gets maximum value.

Step 4. Use parameters to reconstruct the original time series and construct an optimal local linear prediction method (OLLPM) to predict the next value. And the traditional local linear prediction method (LLPM) [9] with the optimal parameters (m_{opt}, n_{opt}) is regarded as the OLLPM method.

Step 5. Combine Bayesian processor of forecast (BPF) technology into the above OLLPM method. Thus, the Bayesian processor can automatically learn by the posterior mean substituting the predicted value from the above OLLPM. This detail operation of the BPF technology is the same as the relational reference [1]. Finally the BPF is used to improve the prediction accuracy by treating the prediction results in OLLPM. And then the new prediction results are regarded as the results of the CBOPM.

3. Application in hydrological time series

The real hydrological time series in this study are chosen as the monthly discharges of Yichang station and Chuntan station at Changjiang river in China for the period January 1890–December 2000. For the monthly discharge, the first training data set for Yichang and Chuntan is the data during the period January 1890–October 1996. The prediction time for both of the time series is 50 months.

In this study, the delay time and the embedding dimension of the above two monthly discharge time series are determined by the autocorrelation function method. The changes of the autocorrelation functions for the time series are

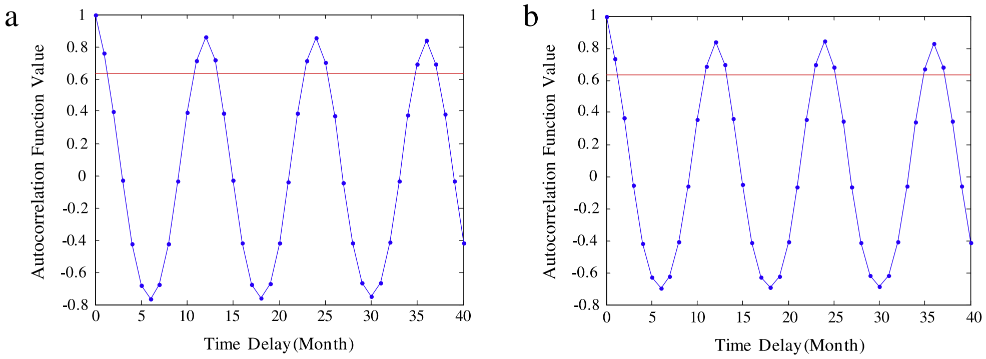

Fig. 1. Autocorrelation functions for monthly discharge time series observed: (a) at Yichang; (b) at Chuntan.

Table 1
The comparison of the prediction results with LLPM1, LLPM2, LLPM3, OLLPM and CBOPM at Yichang and Chuntan.

Methods	Yichang		Chuntan	
	MRE (%)	CC	MRE (%)	CC
LLPM1	20.19	0.9229	22.68	0.9162
LLPM2	18.78	0.9046	19.75	0.9067
LLPM3	18.60	0.9319	24.65	0.8658
OLLPM	17.76	0.9356	17.99	0.9162
CBOPM	17.32	0.9515	17.43	0.9379

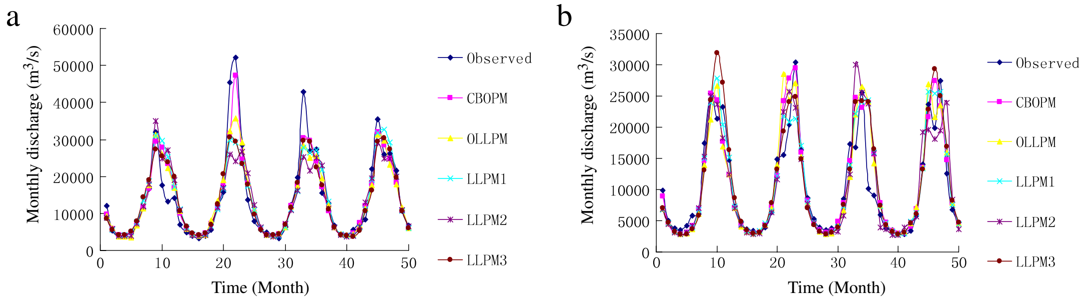

Fig. 2. Comparison of prediction results with the methods of CBOPM, OLLPM, LLPM1, LLPM2 and LLPM3: (a) at Yichang, (b) at Chuntan.

presented in Fig. 1. The delay time is selected when the autocorrelation coefficient has dropped to $1 - 1/e$ of its initial value. From Fig. 1, the delay time for Yichang and Chuntan can be determined as $\tau = 2$, respectively.

The embedding dimension can be determined as $m_c = 20$ for Yichang and Chuntan in the CBOPM. In this study, only the traditional prediction method (LLPM) needs to give this parameter as a prior, the rest methods, optimal LLPM (OLLPM) based on optimal structure parameter (m_{opt}, n_{opt}) and CBOPM, can obtain the number of nearest neighbors adaptively in the prediction procedure.

All of the time series were predicted by OLLPM (at Yichang, $(m_{opt}, n_{opt}) = (6, 10)$; at Chuntan, $(m_{opt}, n_{opt}) = (6, 15)$), LLPM1 ($n = m_{opt} + 6$), LLPM2 ($n = m_{opt} + 7$), LLPM3 ($n = m_{opt} + 8$) and CBOPM, respectively. In order to compare the forecast results under the same condition, the number of nearest neighbors n for LLPM is changed from $m + 1$ to $m + 10$. And the comparison of the forecast results between the LLPM1, LLPM2, LLPM3, OLLPM and CBOPM, is shown in Table 1.

From Table 1, it can be seen that, the OLLPM gets better prediction results (with MRE $=$ 17.76 and CC $=$ 0.9356) at Yichang among the LLPM methods. The CBOPM gets the best prediction results (with MRE $=$ 17.32 and CC $=$ 0.9515) at Yichang.

For Chuntan time series, the OLLPM also gets better prediction results (with MRE $=$ 17.99 and CC $=$ 0.9162) among the LLPM methods. The CBOPM also gets the best prediction results (with MRE $=$ 17.43 and CC $=$ 0.9379).

Comparison of predicted and observed monthly runoffs at Yichang and Chuntan is given in Fig. 2. From the Fig. 2, we can see that prediction results of CBOPM are better than that of the LLPM1, LLPM2, LLPM3 and OLLPM in the extreme values (maximum values) and other values.

4. Conclusions

This paper presents a new chaos Bayesian optimal prediction method (CBOPM) and applies it to predict the monthly discharges time series.

The optimal parameters can be obtained by using MRE, CC indices and chaos theory. CBOPM has the ability to learn in its forecast process by using BPF technology. It provides a new way to improve the prediction accuracy in predicting monthly hydrological time series.

The application results indicate that the new CBOPM can choose the embedding dimension and the number of the nearest neighbors adaptively in the prediction process. Compared with the LLPM1, LLPM2, LLPM3 and OLLPM, CBOPM is the best method for predicting monthly hydrological time series.

Acknowledgements

The present work is supported by the Program for Changjiang Scholars and Innovative Research Team in University under Grant No. IRT0809, the Program of the National Science Foundation of PR China under Grant Nos. 50939001 and 51079004, the National Basic Research Program of PR China under Grant Nos. 2010CB429003 and 2010CB951104.

References

[1] X.H. Yang, D.X. She, Z.F. Yang, Q.H. Tang, J.Q. Li, Chaotic Bayesian method based on multiple criteria decision making (MCDM) for forecasting nonlinear hydrological time series, International Journal of Nonlinear Sciences and Numerical Simulation 10 (11–12) (2009) 1595–1610.
[2] K.W. Chau, Calibration of flow and water quality modeling using genetic algorithm, Journal Lecture Notes in Artificial Intelligence 2557 (2002) 720–724.
[3] M. Li, Modeling autocorrelation functions of long-range dependent teletraffic series based on optimal approximation in Hilbert space—a further study, Applied Mathematical Modelling 31 (3) (2007) 625–631.
[4] E.N. Lorenz, Deterministic nonperiodic flow, Journal of the Atmospheric Sciences 20 (1) (1963) 130–141.
[5] M. Mackey, L. Glass, Oscillations and chaos in physiological control systems, Science 197 (1977) 287–289.
[6] R.M. May, Simple mathematical models with very complicated dynamics, Nature 261 (1976) 459–467.
[7] L.Y. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Physica D 110 (1) (1997) 43–50.
[8] B. Sivakumar, Chaos theory in hydrology: important issues and interpretations, Journal of Hydrology 227 (1) (2000) 1–20.
[9] D.J. Farmer, J.J. Sidorowich, Predicting chaotic time series, Physical Review Letters 59 (1) (1987) 845–848.
[10] A.W. Jayawardena, W.K. Li, P. Xu, Neighbourhood selection for local modelling and prediction of hydrological time series, Journal of Hydrology 258 (3) (2002) 40–57.