

网络出版时间:2018-11-06 16:58

DOI:10.13207/j.cnki.jnwafu.2019.05.019

网络出版地址:http://kns.cnki.net/kcms/detail/61.1390.S.20181106.1657.038.html

# 基于熵谱理论的月径流预报

周正弘, 粟晓玲

(西北农林科技大学 水利与建筑工程学院, 陕西 杨陵 712100)

**【摘要】**【目的】探讨熵谱模型在月径流预报中的应用效果以及训练期长度对模型预报精度的影响,为熵谱模型在径流预报中的应用提供参考。【方法】月径流预报依据黑河莺落峡站月径流资料,采用伯格熵(BESA)和构造熵(CESA)2种熵谱模型进行月径流预报,并用平均相对误差( $RE$ )、均方根误差( $RMSE$ )、相关系数( $R$ )和纳西效率系数( $NSE$ )对模型预报精度进行评价。【结果】训练期长度过短会使模型阶数偏低,模型无法做出准确的预测;训练期长度过长会使模型阶数偏高,此时训练期和验证期精度反而略微下降;适中的训练期长度能够使模型的训练期和验证期精度均相对较高且稳定。对于黑河莺落峡站,BESA模型的最佳训练期长度为13年,CESA模型的最佳训练期长度为19年,CESA模型的训练期拟合精度和验证期预报精度均高于BESA模型,同时CESA模型在汛期预报精度相对较高,而BESA模型在非汛期预报精度相对较高。【结论】BESA和CESA2种模型都可用于月径流预报,但需要合理选择训练期长度,使模型阶数适中且稳定,以提高预报精度和可靠性。

**【关键词】** 月径流预报;熵谱分析;伯格熵;构造熵;时间序列分析

**【中图分类号】** P338

**【文献标志码】** A

**【文章编号】** 1671-9387(2019)05-0146-09

## Monthly streamflow forecasting based on entropy spectral theory

ZHOU Zhenghong, SU Xiaoling

(College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** 【Objective】 This study discussed the application of entropy spectral model in monthly streamflow forecasting and the effect of lengths of calibration period on model performance to provide reference for the application of entropy spectral model in streamflow forecasting. 【Method】 Burg entropy and configurational entropy spectral analysis model were introduced for monthly streamflow forecasting at Yingluoxia station in Heihe river basin. Relative error ( $RE$ ), root mean square error ( $RMSE$ ), correlation coefficient ( $R$ ), and Nash-Sutcliffe efficiency coefficient ( $NSE$ ) were used to evaluate the model performance. 【Result】 Shorter training period resulted in lower model order and the model cannot forecast accurately. Longer training period resulted in higher model order and lower accuracy in the training period and verification period. Moderate length of training period resulted in high and stable accuracy in training period and verification period. At Yingluoxia station in Heihe river basin, the best calibration period length of BESA model and CESA model were 13 years and 19 years, respectively. The accuracy of the CESA model was higher than that of the BESA model. In verification period, the CESA model had higher forecast accuracy in the flood season, while BESA model had higher forecast accuracy in non-flood period. 【Conclusion】 Both BESA and CESA models can be used for monthly streamflow forecasting. However, the length of calibration period needs to be reasonably selected so that the model order can be moderate and stable to improve forecast

〔收稿日期〕 2018-03-28

〔基金项目〕 国家自然科学基金项目(91425302)

〔作者简介〕 周正弘(1993—),男,河南濮阳人,在读硕士,主要从事水文预报研究。E-mail:zzh199302@nwsuaf.edu.cn

〔通信作者〕 粟晓玲(1968—),女,四川开江人,教授,博士,博士生导师,主要从事水文预报与水文模型及水资源规划与管理研究。E-mail:suxiaoling17@126.com

accuracy and reliability.

**Key words:** monthly streamflow forecasting; entropy spectral analysis; Burg entropy; Configurational entropy; time series analysis

径流预报是利用已有信息提前估计未来径流的过程,对河道整治、航运、水库调度、水电站发电计划等有着重要意义。月径流序列具有随机性、季节性与周期性等特点,时间序列分析常用于月径流预报,其中应用最广泛的时间序列模型是自回归模型(AR)或自回归滑动平均模型(ARMA)。然而,这些模型的线性假设限制了其应用<sup>[1]</sup>。熵谱分析能够提取时间序列的周期特征等重要信息<sup>[2-6]</sup>。因此,结合熵谱分析和时间序列分析的熵谱理论,可为水文预报提供新的思路和方法。

Burg<sup>[7]</sup>定义了伯格熵(Burg entropy),并根据最大熵原则(POME)推导了估计时间序列谱密度函数的方法与外推时间序列的预报模型,即伯格熵谱分析(Burg entropy spectral analysis, BESA)。BESA 在国内外各个领域都有广泛应用,并被推荐为水文序列预报的经典模型<sup>[8-11]</sup>。黄忠恕<sup>[12]</sup>以长江宜昌站年最高水位序列为例,对比调和分析、功率谱分析和 BESA 3 种谱分析方法得到的功率谱密度函数,认为当模型定阶合适时, BESA 在分辨率、谱偏移及对短序列的适应性等方面都优于其他 2 种谱分析方法。霍崇仁等<sup>[13]</sup>将 BESA 应用于黄河下游鲁西平原的地下水动态模拟与预报,取得了令人满意的结果。王栋等<sup>[14]</sup>认为,使用 BESA 能较好地识别出水文时间月序列和年序列中隐含的周期成分,并发挥最大熵谱分析假定少、合乎自然、方便快捷的优势。沈慧芳等<sup>[15]</sup>提出了一种更严格的最大熵谱估计的递推算法,该算法除需要满足 BESA 已有的向前和向后误差最小外,还需要满足最优预测误差与信号正交这一条件,并认为该方法得到的谱密度分辨率高于 BESA。然而,由于非平稳条件下 BESA 存在难以确定多峰谱密度的弱点<sup>[16]</sup>,因此有时利用 BESA 并不能很好地模拟具有较强季节性和周期性特点的月径流<sup>[17]</sup>。

Frieden<sup>[18]</sup>定义了构造熵(Configurational entropy),将构造熵谱分析(Configurational entropy spectral analysis, CESA)应用于图像识别领域。Wu<sup>[19]</sup>基于有限长倒谱模型推导出显式谱密度函数估计公式,解决了以往 CESA 计算复杂的问题。Nadeu<sup>[20]</sup>认为,对于自回归滑动平均模型(ARMA)和滑动平均模型(MA),CESA 谱估计精度高于 BE-

SA 谱估计,对于自回归模型(AR)两者精度相当,但 Katsakos-Mavromichalis 等<sup>[21]</sup>认为, BESA 在估计白噪声序列谱密度时精度较高。Cui 等<sup>[17]</sup>利用 Wu<sup>[19]</sup>构造的谱密度估计公式反推出了基于 CESA 的单变量径流预报模型,通过对 19 条不同河流的计算,认为 CESA 预报精度优于 BESA,并在月径流预报方面较 BESA 显示出更好的分辨率和可靠性。

综观已有文献,在使用 BESA 与 CESA 2 种熵谱模型进行月径流预报时选取的训练期总长度较短,一般小于 10 年<sup>[17,22]</sup>,而当模型样本容量过小时,样本可能不能包含径流变化的所有情况,因此有必要探究训练期长度对 BESA 模型和 CESA 模型预报精度的影响。为此,本研究分别应用 BESA 与 CESA 2 种熵谱模型,对黑河莺落峡站月径流序列进行模拟,比较 2 种方法的预报效果,并探讨 3~42 年的不同训练期长度对 2 种方法预报精度的影响,分析 2 种预报模型在黑河莺落峡站的最佳训练期,进而为月径流预报中模型的选择与应用提供参考依据。

## 1 最大熵谱分析预报模型的推导

### 1.1 最大熵模型

将径流序列的频率  $f$  视为一个随机变量,径流序列的谱密度函数  $P(f)$  可视作概率密度函数,因此伯格熵  $H_B(f)$  的定义如下<sup>[7]</sup>:

$$H_B(f) = - \int_{-W}^W \ln[P(f)] df. \quad (1)$$

式中:积分上下限  $W$  为尼奎斯特频率,为样本采样间隔倒数的一半。

而构造熵  $H_C(f)$  的定义如下<sup>[18]</sup>:

$$H_C(f) = - \int_{-W}^W P(f) \ln[P(f)] df. \quad (2)$$

由于自相关函数  $\rho(n)$  是谱密度函数  $P(f)$  的傅里叶逆变换,则模型的约束条件为:

$$\rho(n) = \int_{-W}^W P(f) e^{i2\pi f n \Delta t} df. \quad (3)$$

式中: $i$  为虚数; $n$  为自相关函数的滞时,也表示离散傅里叶变换的第  $n$  个谐波, $n$  为整数,  $-N \leq n \leq N$ ,其中  $N$  通常取样本序列长度的  $1/4 \sim 1/2$ ;  $\Delta t$  为样本采样间隔。

### 1.2 谱密度函数求解

为了得到最大信息熵的谱密度函数  $P(f)$  的无

偏估计,使用拉格朗日乘子法构造拉格朗日函数,基于伯格熵的拉格朗日函数  $L_B(f)$  为:

$$L_B(f) = - \int_{-w}^w \ln[P(f)] df - \sum_{n=-N}^N \lambda_n \left[ \int_{-w}^w P(f) e^{i2\pi f n \Delta t} df - \rho(n) \right]. \quad (4)$$

式中:  $\lambda_n$  为拉格朗日乘子。

基于构造熵的拉格朗日函数  $L_C(f)$  为:

$$L_C(f) = - \int_{-w}^w P(f) \ln[P(f)] df - \sum_{n=-N}^N \lambda_n \left[ \int_{-w}^w P(f) e^{i2\pi f n \Delta t} df - \rho(n) \right]. \quad (5)$$

分别求拉格朗日函数  $L_B(f)$  和  $L_C(f)$  对谱密度函数  $P(f)$  的偏微分并使之等于 0, 得到最大伯格熵谱密度函数  $P_B(f)$  和最大构造熵谱密度函数  $P_C(f)$  为:

$$P_B(f) = - \frac{1}{\sum_{n=-N}^N \lambda_n \exp(-i2\pi f n \Delta t)}, \quad (6)$$

$$P_C(f) = \exp\left(-1 - \sum_{n=-N}^N \lambda_n e^{i2\pi f n \Delta t}\right). \quad (7)$$

### 1.3 BESA 模型求解

$P_B(f)$  的形式与自回归模型的谱密度形式相同, 伯格提出了一种以自回归模型为基础的参数估计方法, 自回归模型参数  $a_k(i)$  的估计公式<sup>[7]</sup> 如下:

$$a_k(i) = \begin{cases} a_{k-1}(i) + k_k a_{k-1}(k-i), & i=1, 2, \dots, k-1; \\ k_k, & i=k. \end{cases} \quad (8)$$

式中:  $a_k(i)$  为  $k$  阶自回归模型的第  $i$  个参数,  $k_k$  为第  $k$  阶反射系数, 若反射系数  $k_1, k_2, \dots, k_k$  已知, 则可以通过式(8)计算出  $1 \sim k$  阶自回归模型参数, 利用向前预测误差和向后预测误差功率最小得到反射系数  $k_k$  的计算公式如下:

$$k_k = \frac{-2 \sum_{i=k}^{N-1} e_{k-1}^f(i) e_{k-1}^b(i-1)}{\sum_{i=k}^{N-1} |e_{k-1}^f(i)|^2 + \sum_{i=k}^{N-1} |e_{k-1}^b(i-1)|^2}. \quad (9)$$

式中:  $e_{k-1}^f$  和  $e_{k-1}^b(i-1)$  分别为向前和向后预测误差, 上标  $f, b$  分别表示向前或向后预测, 下标  $k-1$  表示预测误差为  $k-1$  阶, 自变量  $i$  和  $i-1$  为计算误差的输入长度。第  $k-1$  阶的向前和向后预测误差可以由第  $k-2$  阶向前和向后预测误差递推得到, 其公式如下:

$$e_{k-1}^f(i) = \begin{cases} y(i), & k=1; \\ e_{k-2}^f(i) + k_{k-1} e_{k-2}^b(i-1), & k>1. \end{cases} \quad (10)$$

$$e_{k-1}^b(i) = \begin{cases} y(i), & k=1; \\ e_{k-2}^b(i-1) + k_{k-1} e_{k-2}^f(i), & k>1. \end{cases} \quad (11)$$

式中:  $y(i)$  为径流时间序列。

### 1.4 CESA 模型求解

对于构造熵谱密度函数, Wu<sup>[19]</sup> 将倒谱分析应用于求解拉格朗日乘子与外推自相关函数, 提出了最大构造熵的显式解。对式(7)取自然对数, 然后进行傅里叶逆变换可得:

$$\int_{-w}^w \{1 + \ln[P(f)]\} \exp(i2\pi f j \Delta t) df = \int_{-w}^w \left[ - \sum_{n=-N}^N \lambda_n \exp(i2\pi f n \Delta t) \right] \exp(i2\pi f j \Delta t) df. \quad (12)$$

式中:  $j$  表示离散序列傅里叶逆变换的第  $j$  个谐波,  $j$  的取值范围与  $n$  相同。

功率谱密度函数  $P(f)$  取自然对数后进行傅里叶逆变换得到倒谱  $e(j)$ , 计算公式如下:

$$e(j) = \int_{-w}^w \ln[P(f)] \exp(i2\pi f j \Delta t) df. \quad (13)$$

将式(13)代入式(12), 则式(12)可简化为:

$$\delta(j) + e(j) = - \sum_{n=-N}^N \lambda_n \delta(j-n). \quad (14)$$

式中:  $\delta(j)$  和  $\delta(j-n)$  为单位脉冲函数, 只有当自变量为 0 时, 函数值为 1, 当自变量不为 0 时, 函数值为 0。

对于  $j$  取  $0 \sim N$ , 可以得到一个求解拉格朗日乘子的方程组:

$$\begin{cases} \lambda_0 = -e(0) - 1; \\ \lambda_1 = -e(1), \\ \vdots \\ \lambda_N = -e(N). \end{cases} \quad (15)$$

为求解拉格朗日乘子  $\lambda_n$ , Nadeu<sup>[20]</sup> 提出了一种简单的估计有限长序列的倒谱  $e(n)$  方法, 其原理是使所需被估计的谱密度与先验信息的谱密度差距最小, 当  $-N \leq n \leq N$  时, 倒谱估计公式如下:

$$e(n) = \begin{cases} 2[\rho(n) - \sum_{j=1}^{n-1} \frac{j}{n} e(j) \rho(n-j)], & n>0; \\ 0, & n \leq 0. \end{cases} \quad (16)$$

反之, 由倒谱反推自相关函数, 有:

$$\rho(n) = \frac{e(n)}{2} + \sum_{j=1}^{n-1} \frac{j}{n} e(j) \rho(n-j). \quad (17)$$

对于一个自回归模型, 若其阶数确定为  $m$ , 则自相关系数  $\rho(m)$  为:

$$\rho(m) = \sum_{j=1}^m a_m(j) \rho(m-j), \quad m \leq N. \quad (18)$$

式中:  $a_m(j)$  为  $m$  阶自相关模型的第  $j$  个参数,  $a_m(j)$  的计算公式为:

$$a_m(j) = \frac{j}{m} e(j), \quad j < m. \quad (19)$$

在进行径流预报时,外推自相关函数:

$$\rho(t) = \sum_{j=1}^m a_m(j) \rho(t-j), t > T. \quad (20)$$

式中:  $T$  为径流时间序列样本容量,  $t$  为所需预测径流的相对时刻。

时间径流序列  $y(t)$  的预报模型可以表示为:

$$y(t) = \sum_{j=1}^m a_m(j) y(t-j), t > T. \quad (21)$$

### 1.5 自回归模型阶数

自回归模型阶数  $m$  通过 BIC 准则<sup>[23]</sup> 确定,有:

$$BIC(m) = N \ln \sigma_e^2 + m \ln N. \quad (22)$$

式中:  $\sigma_e^2$  为残差的方差,  $N$  为时间序列长度。使 BIC 最小的阶数  $m$  为模型的最佳阶数。

### 1.6 模型评价指标

选择平均相对误差 (RE)、均方根误差 (RMSE)、相关系数 (R) 和纳西效率系数 (NSE) 作为模型预测精度的评价指标,其计算公式分别为:

$$RE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x(t) - f(t)}{x(t)} \right|. \quad (23)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n [x(t) - f(t)]^2}{n-1}}. \quad (24)$$

$$R = \frac{\sum_{t=1}^n (x(t) - \bar{x})(f(t) - \bar{f})}{\left[ \sum_{t=1}^n (x(t) - \bar{x})^2 \sum_{t=1}^n (f(t) - \bar{f})^2 \right]^{0.5}}. \quad (25)$$

$$NSE = 1 - \frac{\sum_{t=1}^n |x(t) - f(t)|^2}{\sum_{t=1}^n |x(t) - \bar{x}|^2}. \quad (26)$$

式中:  $x(t)$  为径流序列的观测值,  $f(t)$  为径流序列预测值,  $n$  为总预测月数,  $\bar{x}$  为  $x(t)$  的均值,  $\bar{f}$  为  $f(t)$  的均值。在评价模型精度时, RE 和 RMSE 指标越小或 R 和 NSE 指标越大,则精度越高。

### 1.7 径流预测步骤

第 1 步:为使径流序列满足自回归模型要求,分别利用 BOX-COX 变换<sup>[24]</sup> 和标准化变换,使实测径流序列  $x(t)$  变换为正态分布的标准化序列  $y(t)$ :

$$y(t) = \text{zscore}\{\text{boxcox}[x(t)]\}. \quad (27)$$

式中: boxcox 函数的输出为输入序列的 BOX-COX 变换, zscore 函数的输出为输入序列的标准化序列, boxcox 函数和 zscore 函数分别来自 MATLAB 2017b 金融工具箱和统计与机器学习工具箱。

第 2 步:根据式(8)~(11)计算 1 到  $T/2$  阶 BESA 模型参数;根据式(16)与式(19)计算倒谱和 1 到  $T/2$  阶 CESA 模型参数;

第 3 步:将训练期拟合结果带入式(22)计算 BIC 值, BIC 最小值对应的模型阶数  $m$  为最优模型阶数,将  $m$  代入模型,对月径流进行预报;

第 4 步:用式(27)对预报结果进行逆变换,得到径流序列的预测值  $f(t)$ 。

## 2 黑河流域莺落峡站月径流预报

### 2.1 黑河流域概况

黑河是我国西北地区第二大内陆河,发源于祁连山北坡,由莺落峡和正义峡两个水文断面将流域分为上中下游。随着中游地区经济的发展,水资源供需矛盾和用水结构不合理、下游生态急剧恶化等问题日益突出,为改善下游生态状况,促进中游用水结构优化,1997 年国务院审批通过了《黑河干流水量分配方案》(后称“九七”分水方案),但自分水方案实施以来,正义峡实际下泄水量在大部分年份难以完成分水方案下泄指标的要求<sup>[25]</sup>。莺落峡来水是黑河中游地区最主要的供给水源,因此,准确预报莺落峡来水可以为流域管理部门制定合理水量分配方案提供理论依据。

依据黑河莺落峡站 1966—2012 年共 47 年实测月径流序列分别建立 BESA 和 CESA 预报模型,选取 3~42 年的月径流序列作为训练期分别对模型进行拟合,训练期长度设置见表 1。用 2008—2012 年的月径流序列进行检验,采用 RE、RMSE、R 和 NSE 作为评价指标。

表 1 训练期长度与相应模型阶数

Table 1 Selection of different calibration periods and corresponding model orders

序号 No.	训练期 Calibration period	训练期长度/年 Length of calibration period	模型阶数 Model order		序号 No.	训练期 Calibration period	训练期长度/年 Length of calibration period	模型阶数 Model order	
			BESA	CESA				BESA	CESA
1	2005—2007	3	6	9	16	1990—2007	18	12	14
2	2004—2007	4	6	9	17	1989—2007	19	12	14
3	2003—2007	5	8	10	18	1988—2007	20	12	20
4	2002—2007	6	6	13	19	1987—2007	21	12	20
5	2001—2007	7	11	14	20	1986—2007	22	12	20
6	2000—2007	8	17	14	21	1984—2007	24	12	20
7	1999—2007	9	7	14	22	1982—2007	26	12	14

表 1(续) Continued table 1

序号 No.	训练期 Calibration period	训练期长度/年 Length of calibration period	模型阶数 Model order		序号 No.	训练期 Calibration period	训练期长度/年 Length of calibration period	模型阶数 Model order	
			BESA	CESA				BESA	CESA
8	1998—2007	10	11	10	23	1980—2007	28	24	14
9	1997—2007	11	15	20	24	1978—2007	30	16	20
10	1996—2007	12	12	20	25	1976—2007	32	12	20
11	1995—2007	13	12	20	26	1974—2007	34	18	27
12	1994—2007	14	12	14	27	1972—2007	36	12	19
13	1993—2007	15	12	20	28	1970—2007	38	24	39
14	1992—2007	16	12	14	29	1968—2007	40	18	40
15	1991—2007	17	12	14	30	1966—2007	42	21	19

## 2.2 数据预处理

用式(27)对实测月径流序列进行变换,在 MATLAB 中使用 `adftest` 函数对变换后序列进行置信度为 95% 的单位根 (ADF) 检验,结果拒绝原假设,即不存在单位根,认为处理后的序列为平稳随机序列,可以使用自回归模型。

## 2.3 训练期精度比较

分别以不同长度的月径流序列作为训练期,对模型参数进行拟合,依据 BIC 准则确定的自回归模型阶数见表 1。从表 1 可以看出,对于 BESA 和 CESA 2 种模型,当训练期长度小于 11 年时,求出

的自回归模型阶数较低,这可能会导致模型无法模拟径流序列的内在变化规律;当训练期长度大于 32 年时,求出的自回归模型阶数偏大,这可能会导致模型出现过拟合现象;当训练期长度为 12~26 年时, BESA 模型阶数适中,且稳定为 12 阶;当训练期长度为 11~32 年时, CESA 模型阶数适中,且稳定为 14 或 20 阶。由于受模型阶数的限制, BESA 和 CESA 的最优训练期长度应选择模型阶数适中且稳定时的训练期。不同训练期长度对 BESA 和 CESA 模型拟合精度的影响见图 1。

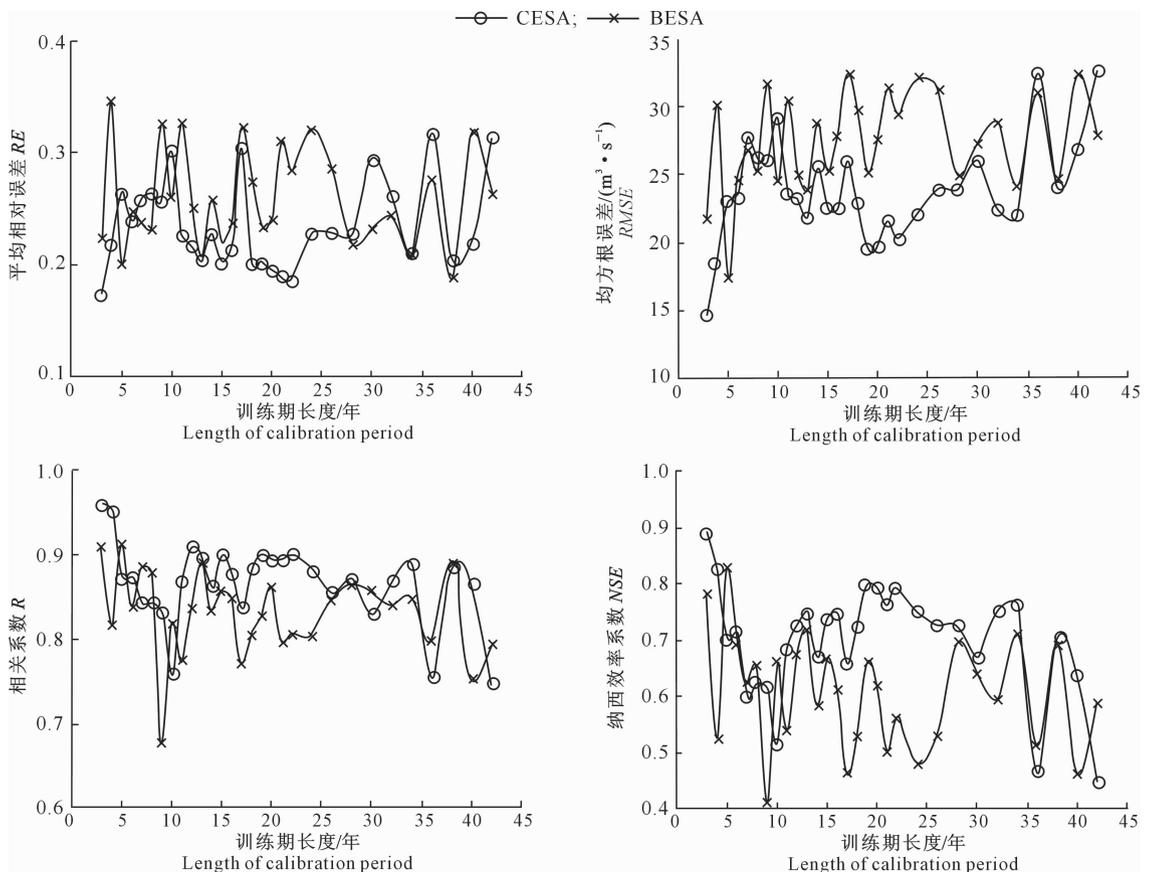


图 1 不同训练期长度下 BESA 和 CESA 2 种熵谱模型训练期精度比较

Fig. 1 Comparison of simulation accuracy of BESA and CESA models in calibration during different lengths of calibration period

由图 1 可知,当训练期长度为 5 年时,BESA 模型拟合精度达到最优,其  $RE$  为 20.2%, $RMSE$  为  $17.59 \text{ m}^3/\text{s}$ , $R$  为 0.913, $NSE$  为 0.829;但在模型阶数稳定后,训练期长度为 13 年时 BESA 模型拟合精度也接近最优,其  $RE$  为 20.8%, $RMSE$  为  $23.85 \text{ m}^3/\text{s}$ , $R$  为 0.890, $NSE$  为 0.719。当训练期长度为 3 年时,CESA 模型拟合精度达到最优,其  $RE$  为 17.4%, $RMSE$  为  $14.59 \text{ m}^3/\text{s}$ , $R$  为 0.961, $NSE$  为 0.893;但在模型阶数稳定后,训练期长度为 19 年时 CESA 模型的拟合精度也接近最优,其  $RE$  为 20.2%, $RMSE$  为  $19.49 \text{ m}^3/\text{s}$ , $R$  为 0.901, $NSE$  为

0.800。对于 BESA 和 CESA 模型,训练期长度分别为 5 年和 3 年时训练期拟合精度达到最优,但此时模型阶数较低,分别只有 8 阶和 9 阶,可能无法模拟径流序列的变化规律,在模型阶数稳定的前提下,BESA 和 CESA 模型的训练期长度分别为 13 年和 19 年时训练期拟合精度达到最优,并且 CESA 模型的训练期拟合精度优于 BESA 模型。

### 2.4 验证期精度比较

为验证最优训练期长度的预测效果,利用 2 种熵谱模型对验证期月径流序列进行预测,将预测值与观测值进行对比,其预报精度见图 2。

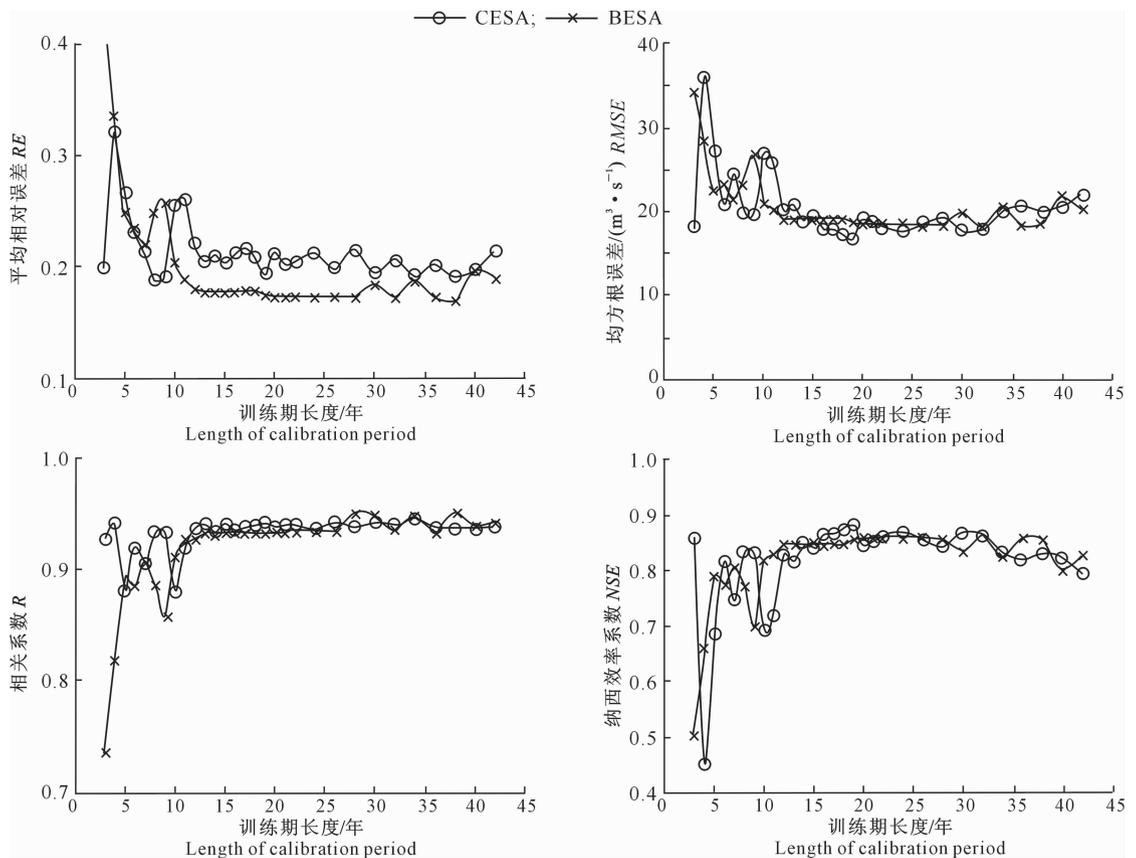


图 2 不同训练期长度下 BESA 和 CESA 2 种熵谱模型验证期精度比较

Fig. 2 Comparison of forecasting accuracy of BESA and CESA models in validation during different lengths of calibration period

由图 2 可知,对于 BESA 和 CESA 2 种模型,当训练期长度为 3~11 年时,BESA 和 CESA 模型阶数较低,此时模型验证期预报精度相对较低且不稳定;当训练期长度为 12~26 年时,模型 BESA 和 CESA 模型阶数适中且稳定,此时模型验证期预报精度相对较高且稳定;当训练期长度大于 26 年时,BESA 模型和 CESA 模型阶数相对偏高,此时模型验证期精度略低于上一阶段且不稳定。结果表明,

适当的模型阶数对 BESA 和 CESA 模型非常重要。

对于 BESA 模型,当训练期长度为 26 年时,模型验证期预报精度最高,其  $RE$  为 17.3%, $RMSE$  为  $18.35 \text{ m}^3/\text{s}$ , $R$  为 0.934, $NSE$  为 0.859;当训练期长度为 13 年时,模型验证期预报精度略低于训练期长度为 26 年时,其  $RE$  为 17.7%, $RMSE$  为  $19.19 \text{ m}^3/\text{s}$ , $R$  为 0.933, $NSE$  为 0.846。对于 CESA 模型,当训练期长度为 19 年时,模型验证期预报

精度最高,其  $RE$  为 19.4%, $RMSE$  为  $16.81 \text{ m}^3/\text{s}$ , $R$  为 0.942, $NSE$  为 0.882。通过比较验证期预报精度可以发现,在 BESA 模型训练期拟合精度最高时(训练期长度为 13 年),验证期预报精度接近但未达到最高;在 CESA 模型训练期拟合精度最高时(训练期长度为 19 年),验证期预报精度达到最高;CESA 模型验证期预报精度优于 BESA 模型,但 BESA 模型的验证期平均相对误差  $RE$  小于 CESA 模型。由于  $RE$  所反映的是观测值与预测值之间的

线性误差,而  $RMSE$ 、 $R$ 、 $NSE$  所反映的是观测值与预测值之间的二次幂误差,若汛期预测误差相对较小,就会使  $RMSE$ 、 $R$ 、 $NSE$  等评价指标更优;若非汛期预测误差相对较小,就会使  $RE$  更优。

为进一步分析最优训练期长度下 BESA 和 CESA 模型的预报效果,分别将最优训练期带入 2 种模型,得验证期预测值,将其与观测值进行对比,结果见图 3。

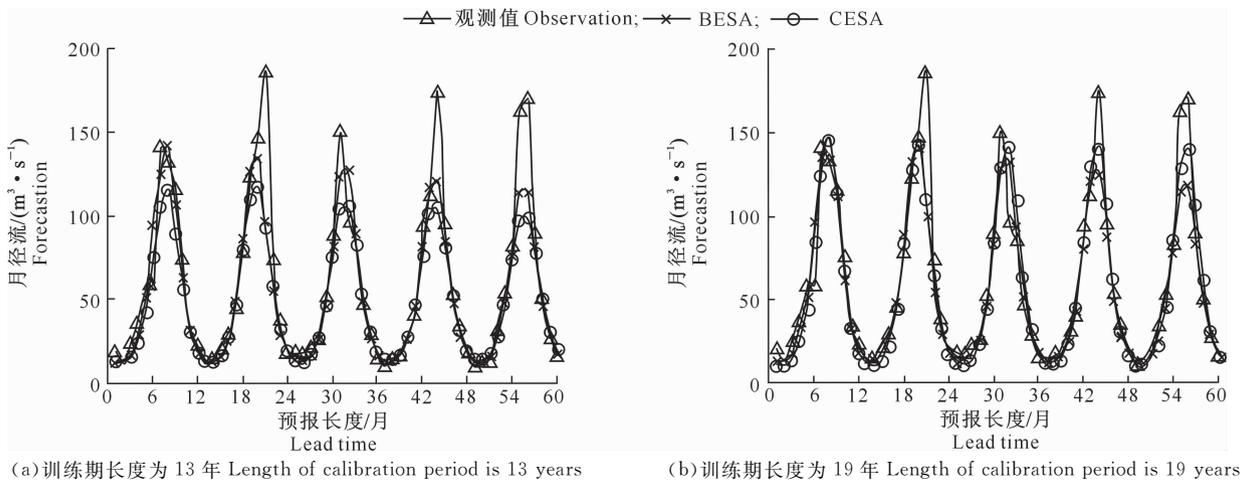


图 3 最佳训练期下验证期莺落峡站月径流观测值与预测值的比较

Fig. 3 Comparison of observed and forecasted monthly streamflow at Yingluoxia station with the best calibration period

从图 3 可以看出,在非汛期,2 种模型都能较好地进行预测,在每年的汛期 7—9 月,2 种模型预报结果均偏小,但相对而言 CESA 模型的预报结果更接近观测值,表明 CESA 模型在汛期的预报效果相对更好。造成汛期预报效果不好的原因可能是由于 7—9 月的径流量主要由暴雨组成,在无降水资料的条件很难对汛期月径流做出准确预测。提取最优

训练期长度下 BESA 模型和 CESA 模型非汛期的预测值,结合相应的观测值绘制散点图,结果见图 4。从图 4 可以看出,2 种模型的趋势线斜率都接近 1,其中 BESA 模型斜率更接近 1,说明 BESA 模型在非汛期的预报效果相对更好,这也验证了 BESA 模型的  $RE$  较 CESA 模型更小的合理性。

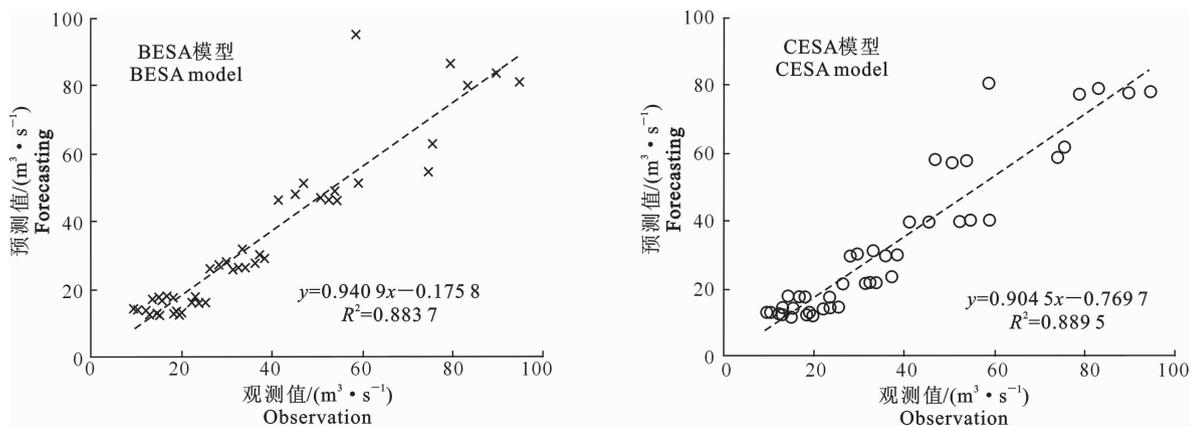


图 4 最佳训练期长度时 BESA 模型和 CESA 模型非汛期月径流观测值与预测值的散点图

Fig. 4 Comparison of observed non-flood period monthly streamflow and predictions by CESA and BESA with the best calibration period

### 3 讨论与结论

本研究选择 BESA 和 CESA 2 种熵谱模型,依据黑河莺落峡站实测月径流资料进行月径流预报,并对比分析了训练期长度与预报精度的关系,结果表明:

1)对于 2 种熵谱模型,当训练期长度较短时,依据 BIC 准则求出的最佳模型阶数较小,并且不稳定,此时模型的训练期评价指标虽然能够达到最优,但不稳定,且模型的验证期评价指标未达到最优;当训练期长度适中时模型阶数趋于稳定,此时模型的训练期和验证期精度均能达到较高水平;当训练期长度较长时,模型的阶数偏大并且不稳定,训练期拟合精度和验证期预报精度低于上一阶段。

2)在模型阶数适中且稳定的条件下,当训练期长度为 13 年时,BESA 模型训练期评价指标达到最优,验证期评价指标略低于最优结果;当训练期长度为 19 年时,CESA 模型训练期和验证期评价指标均达到最优。表明 BESA 和 CESA 模型的最佳训练期长度分别为 13 和 19 年。在训练期长度最佳时,BESA 和 CESA 模型预报精度均较高,但 CESA 模型总体预报精度高于 BESA 模型,其中汛期时 CESA 模型的报精度更高,而非汛期时 BESA 模型的预报精度相对更高。本研究得到的莺落峡站最佳训练期长度较 Cui 等<sup>[17]</sup>和 Zhou 等<sup>[22]</sup>使用的训练期长度更长,但验证期预报精度也更高。这表明利用熵谱模型进行预报时,恰当的训练期长度能够提高预报精度和稳定性。

#### [参考文献]

[1] Elshorbagy A, Simonovic S P, Panu U S. Noise reduction in chaotic hydrologic time series: facts and doubts [J]. *Journal of Hydrology*, 2002, 256(3): 147-165.

[2] Singh V P, Cui H. Entropy theory for streamflow forecasting [J]. *Environmental Processes*, 2015, 2(3): 449-460.

[3] Fleming S W, Marsh Lavenue A, Aly A H, et al. Practical applications of spectral analysis to hydrologic time series [J]. *Hydrological Processes*, 2002, 16(2): 565-574.

[4] Labat D. Recent advances in wavelet analyses: I. A review of concepts [J]. *Journal of Hydrology*, 2005, 314(1): 275-288.

[5] Marques C A F, Ferreira J A, Rocha A, et al. Singular spectrum analysis and forecasting of hydrological time series [J]. *Physics and Chemistry of the Earth*, 2006, 31(18): 1172-1179.

[6] Sang Y F, Wang Z G, Liu C M. Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis [J]. *Journal of Hydrology*,

2012, 424(10): 154-164.

[7] Burg J P. Maximum entropy spectral analysis [D]. Palo Alto, USA: Stanford University, 1975.

[8] Krstanovic P F, Singh V P. A univariate model for long-term streamflow forecasting [J]. *Stochastic Hydrology and Hydraulics*, 1991, 5(3): 173-188.

[9] Krstanovic P F, Singh V P. A real-time flood forecasting model based on maximum-entropy spectral analysis: I. Development [J]. *Water Resources Management*, 1993, 7(2): 109-129.

[10] Krstanovic P F, Singh V P. A real-time flood forecasting model based on maximum-entropy spectral analysis: II. Application [J]. *Water Resources Management*, 1993, 7(2): 131-151.

[11] Singh V P. Entropy theory and its application in environmental and water engineering [M]. Hoboken, USA: John Wiley & Sons, 2013.

[12] 黄忠恕. 波谱分析方法在水文上的应用 [J]. *水文*, 1983(3): 13-20.

Huang Z S. The application of spectrum analysis method in hydrology [J]. *Hydrology*, 1983(3): 13-20.

[13] 霍崇仁, 陈南祥, 刘恒岐. 时间序列的自由回归模型在灌区地下水动态模拟与预报中的应用 [J]. *工程勘察*, 1990(1): 36-38.

Huo C R, Chen N X, Liu H Q. Autogression model of time series is used to simulation of groundwater dynamic in irrigation areas [J]. *Geotechnical Investigation and Surveying*, 1990(1): 36-38.

[14] 王 栋, 朱元甍. 基于 MEM1 谱分析的水文时间序列隐含周期特性研究 [J]. *水文*, 2002, 22(2): 19-23.

Wang D, Zhu Y S. Research on cryptic period of hydrologic time series based on MEM1 spectral analysis [J]. *Hydrology*, 2002, 22(2): 19-23.

[15] 沈慧芳, 李民生, 罗 丰. 基于递推算法的严格最大熵谱估计 [J]. *雷达科学与技术*, 2008, 6(4): 288-291.

Shen H F, Li M S, Luo F. Strict maximum entropy spectral estimation based on recursive algorithm [J]. *Radar Science and Technology*, 2008, 6(4): 288-291.

[16] Boshnakov G N, Lambert-Lacroix S. A periodic Levinson-Durbin algorithm for entropy maximization [J]. *Computational Statistics & Data Analysis*, 2012, 56(1): 15-24.

[17] Cui H, Singh V P. Configurational entropy theory for streamflow forecasting [J]. *Journal of Hydrology*, 2015, 2(3): 449-460.

[18] Frieden B R. Restoring with maximum likelihood and maximum entropy [J]. *JOSA*, 1972, 62(4): 511-518.

[19] Wu N L. An explicit solution and data extension in the maximum entropy method [J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1983, 31(2): 486-491.

[20] Nadeu C. Finite length cepstrum modelling: a simple spectrum estimation technique [J]. *Signal Processing*, 1992, 26(1): 49-59.

[21] Katsakos-Mavromichalis N A, Tzannes M A, Tzannes N S. Frequency resolution: a comparative study of four entropy

- methods [J]. *Kybernetes*, 1986, 15(1): 25-32.
- [22] Zhou Z H, Ju J L, Su X L, et al. Comparison of two entropy spectral analysis methods for streamflow forecasting in Northwest China [J]. *Entropy*, 2017, 19(11): 597-611.
- [23] Schwarz G. Estimating the dimension of a model [J]. *Annals of Statistics*, 1978, 6(2): 15-18.
- [24] Box G E, Cox D R. An analysis of transformations [J]. *Journal of the Royal Statistical Society*, 1964, 26(2): 211-252.
- [25] 赵梦杰, 蒋晓辉, 姚文艺, 等. 影响黑河正义峡断面下泄水量的主导因子分析 [J]. *人民黄河*, 2016, 38(1): 56-59.
- Zhao M J, Jiang X H, Yao W Y, et al. Analysis of dominant factors of affecting the released water amount of Zhengyixia section of Heihe River [J]. *Yellow River*, 2016, 38(1): 56-59

(上接第 145 页)

- [20] 杨利国, 骆亚生, 李 炎, 等. 初始应力条件对压实黄土动强度影响的研究 [J]. *岩土力学*, 2010, 31(1): 87-91.
- Yang L G, Luo Y S, Li Y, et al. Research on effect of initial stress conditions on dynamic strength of compacted loess [J]. *Rock and Soil Mechanics*, 2010, 31(1): 87-91.
- [21] 高玉峰, 黎 冰. 黏土与 EPS 颗粒混合轻质土的动强度特性试验研究 [J]. *岩石力学与工程学报*, 2007, 26(2): 4276-4283.
- Gao Y F, Li B. Experimental study on dynamic strength properties of lightweight clay mixed with EPS beads soil [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2007, 26(2): 4276-4283.
- [22] 辛 凌, 刘汉龙, 沈 杨, 等. 废弃轮胎橡胶颗粒轻质混合土强度特性试验研究 [J]. *岩石力学与工程学报*, 2010, 32(3): 428-433.
- Xin L, Liu H L, Shen Y, et al. Consolidated undrained triaxial compression tests on lightweight soil mixed with rubber chips of scrap tires [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2010, 32(3): 428-433.
- [23] 姬凤玲, 李 强. 疏浚淤泥 EPS 颗粒轻质土等向压缩变形特性研究 [J]. *环境科学与技术*, 2016, 39(S1): 16-19.
- Ji F L, Li Q. Study on the isotropic compression deformation properties of lightweight EPS-bead-treated soil made from dredging silt [J]. *Environmental Science & Technology*, 2016, 39(S1): 16-19.
- [24] 董金梅, 刘汉龙, 洪振舜, 等. 聚苯乙烯轻质混合土三轴压缩试验研究 [J]. *河海大学学报(自然科学版)*, 2005, 33(1): 99-103.
- Dong J M, Liu H L, Hong Z S, et al. Study on properties of lightweight polystyrene heterogeneous soil with triaxial compression tests [J]. *Journal of Hohai University(Natural Sciences Edition)*, 2005, 33(1): 99-103.
- [25] 黎 冰. 动荷载下粘土与 EPS 颗粒混合轻质土的变形和强度特性试验研究 [D]. 南京: 河海大学, 2007.
- Li B. Experimental study on the deformation and strength properties of lightweight clay-EPS beads soil under cyclic loading [D]. Nanjing: Hohai University, 2007.