基于 EMD 分解的径流量预测模型

张承全

（河南省水利第一工程局，河南 郑州 450000）

摘 要：径流量是一个非线性、不稳定的时间序列数据。分析了经验模态分解和 Hilbert-Huang 转换的计算方法，构建了多尺度的预测模型，并以黄河高村站 547 个月的径流量数据为资料，探讨了应用该模型进行径流量预测的方法。

关键词：径流量；EMD；Hilbert；预测模型

中图分类号：TV121 **文献标识码**：A **doi**：10.13681/j.cnki.cn41-1282/tv.2017.02.003

0 引言

径流量是一个非线性的、不稳定的时间序列数据。在分析径流量数据时，若采用仅考虑时域的分析方法，往往不能反映瞬时情况。而在一些必要假设的基础上，对观测资料进行处理分析，是体现数据直观价值的重要途径。

数据的分析方法有很多，如小波分析法是一种非线性的处理方法，能够在时域和频域内同时得到较高的分辨率，但会造成许多虚假的谐波，故具有一定的局限性[1]。傅里叶分析法是一种纯时频的数据处理方法，能够在频域内得到非常高的分辨率。但是，它在时域内却失去了分辨能力，对非线性、非稳态数据的处理存在一些不足[2]。1998 年，美国国家宇航局的 Norden E. Huang 首次提出先对一组时间序列数据做经验模态分解（Empirical Mode Decomposition，简称 EMD），再对各个分量进行希尔伯特变换的处理方法，被称为希尔伯特黄变换（Hilbert-Huang-Transformation，HHT）[3]。EMD 能把复杂数据分解成一系列的本征模态函数（Imtrinsic Mode Function，简称 IMF），再对每个 IMF 进行希尔伯特变换，将得到的频率、振幅、相位进行整合[4~10]。HHT 数学模型是利用数据自身的调节机制反映出系统多尺度特性[11~13]。笔者通过对黄河高村站 547 个月径流量进行 EMD 分解，构建了多尺度的预测模型，并用该模型进行径流量预测。

1 基本原理

1.1 经验模态分解(EMD)

经验模态分解过程是一个筛选过程，它对非线性序列进行平稳化处理，提取出不同尺度下的波动和趋势因子，进而筛选出其中的本征模态函数 IMF。本征模态函数必须满足以下 2 个条件：(1)该列数据极值点和过零点数目基本相等。(2)在任意时刻，由局部极大点构成的包络线和由局部极小点构成的包络线的平均值为零。

EMD 的计算步骤为：(1)寻找数据序列 $S(t)$的所有极大值和极小值，并将其用三次样条函数分别拟合成原数据序列的上、下包络线。(2)对上、下包络线求局部的均值。(3)用原数据序列 $S(t)$减去局部均值 $m(t)$，即可得到第一个本征模态函数 $h_1(t)$。(4)判断 $h_1(t)$是否满足 IMF 的 2 个条件，若不满足，就将它看作新的时间序列，重复以上步骤，直到满足 IMF 条件时，即可得到第一个 IMF 函数 $C_1(t)$。(5)将 $r_1(t)=S(t)-h_1(t)$看作新的 $S(t)$，重复以上步骤，即可依次得到 IMF2、IMF3……直到剩余部分 $R(t)$满足给定的终止条件时，筛选结束。此时，原始序列可表示为式(1)。这样，数据序列就由一系列的本征模态函数和一个残值数列组成。

$$S(t)=\sum_{i=1}^{n}C(t)+R(t) \tag{1}$$

标准差 SD 的计算式如式(2)所示。一般情况下，前后两个数据的标准差越小，所得的本征模态函数的稳定性就越好，能够分解出的 IMF 的个数就越多。实践表明，当 $SD\in(0.2,0.3)$时，每个 IMF 分量都反映了信号在不同时间尺度内的模态特征，并具有较好的稳定性。

收稿日期：2016-12-28

作者简介：张承全(1984-)，男，河南新乡人，工程师，主要从事水文及水资源、水利施工方面的工作。

$$SD=\sum_{t=0}^{n}\frac{[h_{k-1}(t)-h_k(t)]^2}{h^2_{k-1}(t)} \quad (2)$$

1.2 Hilbert 变换

原数据序列经过 EMD 分解,再按照式(3)对其每一个 IMF 分量进行 Hilbert 变换。

$$Y(t)=\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{C_n(\tau)}{t-\tau}d\tau \quad (3)$$

式中:P 表示柯西主值。

由于 $C(t)$和 $Y(t)$为共轭复数对,于是可以得到一个解析信号 $Z(t)$,如式(6)所示。

$$Z(t)=C(t)+iY(t)=a(t)e^{i\theta(t)} \quad (4)$$

式中:$a(t)$是瞬时振幅函数;$\theta(t)$是相位函数,$\theta(t)=\arctan\frac{Y(t)}{C(t)}$。

相位函数对时间求导,可得到瞬时频率 ω,即 $\omega=\frac{d\theta}{dt}$。周期 $T=\frac{2\pi}{\omega}$。

由此可知,经 Hilbert 变换得到的瞬时振幅和瞬时频率都是随时间变化的函数,所得的瞬时频率能够揭示数据的真实物理意义。

2 模型分析

2.1 技术路线

模型操作的具体流程如图 1 所示。

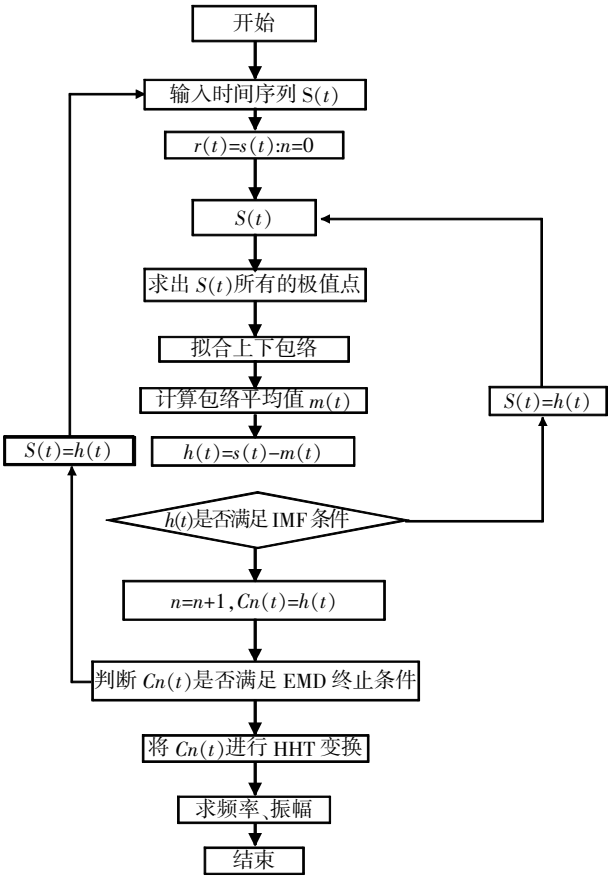

图 1 技术路线图

Fig.1 Technical circuit

2.2 模型建立

根据 EMD 和 HHT 变换理论,结合数据特征,构建预测模型,如式(5)所示。

$$\frac{dS(t)}{d(t)}=r(t)+\sum_{i=1}^{n}a\sin[\frac{2\pi}{T}(b-t)] \quad (5)$$

式中:$S(t)$为月均径流量,亿/m³;t 为时间,年;a 是各 IMF 分量的平均振幅;T 是各 IMF 的平均周期;R(t)是残值项;n 是变换后分量的个数;b 是各函数的初相位。

对式(5)求解可得式(6)。

$$S(t)=R(t)+\sum_{i=1}^{n}C(t)=R(t)+\sum_{i=1}^{n}A\cos[\frac{2\pi}{T}(b-t)] \quad (6)$$

3 山东高村水文站月径流量多尺度分析及预测模型

高村水文站是黄河流入山东的重要控制站,测验河段位于黄河下游游荡河段的末端,断面距河口 579.1 km,集水面积为 734 146 km²。本文选取黄河高村水文站 547 个月的径流量数据进行分析,数据来源于《中国河流泥沙公报》和《黄河水资源公报》。

3.1 多尺度分析

将山东高村水文站 547 个月的月均径流量进行 EMD 分解,结果如图 2 所示。由图 2 可以看出,山东高村水文站 547 个月的月均径流量经 EMD 分解,成为 5 个本征函数 IMF 分量和 1 个残余函数。IMF1 振荡频率最高,从 IMF1 到 IMF5 的频率逐渐变低,振幅逐渐减小。但是,在 IMF4 和 IMF5 中,振幅出现了异常。这种情况时常会出现在低频部分,但不会影响整体的变化趋势。最后一项是残差相,它单调地体现序列的趋势,即径流量有减弱的趋势。

对各个 IMF 分量进行 HHT 变换,可得到 IMF 分量平均幅度和平均频率,如表 1 所示。

由表 1 可知,IMF1 是从时间序列中分解出的振幅最大、频率最高、波长最短的波动,其平均周期在 4.78 个月左右波动,平均振幅为 511.82 亿 m³/月;IMF2 平均周期在 13.84 个月左右波动,平均振幅为 381.67 亿 m³/月;IMF3 平均周期在 29.15 个月左右波动,平均振幅为 338.99 亿 m³/月;IMF4 平均周期在 86.8 个月左右波动,平均振幅为 408.32 亿 m³/月;IMF5 平均周期在 313.86 个月左右波动,平均振幅为 385.41 亿 m³/月。IMF1~IMF5 各本征模态函数的振幅逐渐变小、频率逐渐变低,波长越来越长。这种变化趋势一直持续到频率已经很低的 IMF4。IMF5 的波动振

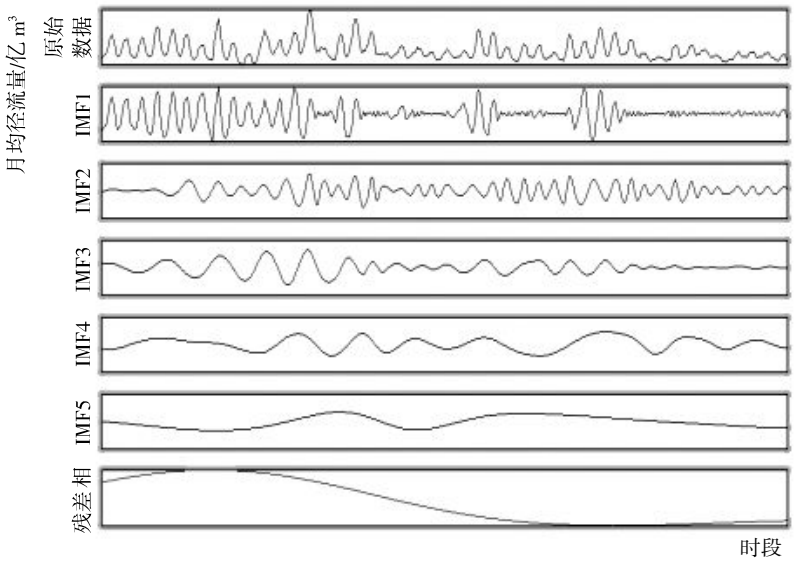

图 2 高村站 547 个月月均径流量分解的 IMF 分量及趋势项

Fig.2 IMF component and trend of Gao village station average runoff decomposition for 547 months

表 1 IMF 分量平均幅度和平均频率

Tab.1 Average amplitude and average frequency of IMF component

分量	平均周期 T_i/月	平均振幅 a_i/(亿 m³/月)	初相位 bi
IMF1	4.78	511.82	−0.54
IMF2	13.84	381.67	−0.99
IMF3	29.15	338.99	−0.95
IMF4	86.8	408.32	−0.51
IMF5	313.86	385.41	1.55

幅比 IMF4 的波动振幅大。这是 EMD 中经常会出现的异常情况,但它只可能出现在频率较低的部分,不会影响整体的变化趋势。可见,高村站 547 个月月均径流量主要以 4.78 和 13.84 为主进行波动。

3.2 预报模型

根据 EMD 分解和 HHT 变换后得到黄河高村站径流量的周期、振幅和初相位,建立 2 种不同的非线性水文模型。

模型一为未考虑初相位的模型,表达式如式(7)所示。模型二为考虑初相位的模型,表达式如式(8)所示。

$$Y=\beta_0+\beta_1 t+511.82\cos\frac{2\pi}{4.78}t+381.67\cos\frac{2\pi}{13.84}t+338.99\cos\frac{2\pi}{29.15}t+408.32\cos\frac{2\pi}{86.8}t+385.41\cos\frac{2\pi}{313.86}t \quad (7)$$

$$Y=\beta_0+\beta_1 t+511.82\cos(\frac{2\pi}{4.78}t-0.54)+381.67\cos(\frac{2\pi}{13.84}t-0.99)+338.99\cos(\frac{2\pi}{29.15}t-0.95)+408.32\cos(\frac{2\pi}{86.8}t-0.51)+385.41\cos(\frac{2\pi}{313.86}t+1.55) \quad (8)$$

3.3 模型的预测

采用模型对黄河高村站第 548~552 月径流量进行预测检验,结果如表 2 所示。由表 2 可知,预测结果与实测值接近,误差在−4%~10%之间。

4 结语

(1)EMD 分解法是对非线性序列进行平稳化的过程,提取数据的客观规律。HHT 变换是一种在时域和频域中应用较好的处理方法。将他们结合起来,用于径流量的预测,可为防洪工作提供依据。

表 2 第 548~552 月预测结果和相对误差

Tab.2 Prediction results and relative error from 548−552 month

时间/月	原始数据	模拟值 1	模拟值 1 的相对误差/%	模拟值 2	模拟值 2 的相对误差/%
548	2 600	2 543.0	−2.19	2 569.3	−1.18
549	1 450	1 442.3	−0.53	1 462.0	0.83
550	683	659.2	−3.48	692.21	1.35
551	821	899	9.50	800.2	−2.53
552	376	371.1	−1.30	378.2	0.60

(2) 用该模型预测黄河高村水文站径流量, 结果为: 高村水文站月平均径流量的波动准周期为 4.78 月、13.84 月、29.15 月、86.8 月、313.86 月。其中, 以 4.78 月和 13.84 月为主要震荡周期。预测误差在-4%~10%之间,预测精度较高,满足相关要求。

(3)模型在进行趋势项数据处理时,做了简化处理,HHT 缺乏完善的数学理论支持。对于这些不足之处,还需要进一步研究。

参考文献:

[1] Long S R,Lai R J, Huang N E, et al. Blocking and trapping of waves in an inhornogeneous flow [J]. Dyu Atmos Oceans,1993,20:79-106.

[2] Farge M. Wavelet transforms and their applications to turbulence[J]. Ann Rev Fluid Mech, 1992, 24:395-457.

[3] Huang N E,Zheng S, Steven R,et al.The empirical mode decomposition and the Hilbert spectrum for nonlinear non-stationary time Series Analysis [A]. Proc R Soc Lond A,1998:454:903- 995.

[4] 李新杰,胡铁松,郭旭宁,等.不同时间尺度的径流时间序列混沌特性分析[J]. 水利学报,2013,44(5):516-520.

[5] 李艳玲,张云鹏,张国玮,等.非线性多尺度模型及其应用[J].数学的实践与认识,2015,45(1):203-208.

[6] Falconer I, Gottwald G A, Melbourne I, et al . Application of the 0-1 Test for chaos to experimental data [J]. SI?AM Journal on Applied Dynamical Systems, 2007, 6(2):395-402 .

[7] 邓拥军,王伟,钱成春,等. EMD 方法及 Hilbert 变换中边界问题的处理[J].科学通报,2001,46(3):257-263.

[8] 周念成,王予疆,陈刚,等. 低频振荡模式辨识中信号非线性去除的平滑先验方法[J].电力系统保护与控制,2012,40(11):1-5.

[9] 戴吾蛟,丁晓利,朱建军,等.基于经验模式分解的滤波去噪法及其在 GPS 多路径效应中的应用[J].测绘学报,2006,35(4):321-327.

[10] 孙娴,林振山.HHT 新技术及其对 ENSO 的诊断[J]. 气象,2006,32(9):17-22.

[11] Norden E Huang, Zheng Shen,Steven R Long,et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis[J].The royal Society,1998,454:903-995.

[12] Bogdan M,Maarten D V,Ivan G,et al. Source separation from single-channel recordings by combining empirical -mode decomposition and independent component analysis [J]. IEEE Trans on Bio-Medical Engineering,2010,57(9):2188-96.

[13] P Flandrin, G. Rilling,P Goncalves.Empirical modedecomposition as a filter bank[J].IEEE Signal Processing Letters,2004,2:112-114.

[责任编辑 杨明庆]

Runoff Prediction Model Based on EMD

Zhang Chengquan

(First Water Conservancy Engineering Bureau of Henan Province, Zhengzhou 45000, Henan, China)

Abstract: Runoff is the nonlinear and unstable time series data. It analyzes the Empirical Mode Decomposition (EMD) and Hilbert-Huang transformation calculation method, establishes multiscale prediction model. Based on runoff data of the Yellow River Gao village hydrologic station for 547 months, it also discusses the method of using this model to predict the runoff.

Key Words: Runoff; EMD; Hilbert; prediction model