

基于小波变换和均生函数周期外推组合模式的 非平稳时间序列分析与长期预测

李 晖^{1,2}, 郭 晨¹, 金鸿章²

(1. 大连海事大学 自动化与电气工程学院, 辽宁 大连 116026; 2. 哈尔滨工程大学 自动化学院, 黑龙江 哈尔滨 150001)

摘要: 提出了利用小波变换和均生函数周期外推组合模式进行时间序列长期预测的方法. 基于小波多分辨率分析理论, 非平稳时间序列被分解为多个相对简单的准周期信号, 信号的趋势项、周期项和随机项被分离出来. 然后采用均生函数周期外推预报模式对这些准周期信号进行预报, 此方法能有效的提高预报长度, 并能获得较高的建模及预报精度. 仿真采用两个典型实例进行验证, 结果表明了方法的正确性和有效性.

关键词: 小波变换; 均生函数; 周期外推; 非平稳时间序列; 长期预测

中图分类号: TP273 **文献标识码:** A

Non-stationary time series analysis and long-term forecasting based on the combination model of wavelet transform and mean-generating function period extrapolation

LI Hui^{1,2}, GUO Chen¹, JIN Hong-zhang²

(1. Automation and Electrical Engineering College, Liaoning Dalian Maritime University, Dalian Liaoning 116026, China;

2. Automation College, Harbin Engineering University, Harbin Heilongjiang 150001, China)

Abstract: A combination model forecasting approach combining wavelet transform(WT) and mean-generating function(MGF) period extrapolation is presented in the paper. According to the theory of wavelet multi-resolution analysis(MRA), the non-stationary time series is decomposed into some relative simple and regular periodical signal series. The trend term, periodical term and stochastic term are separated from the original series. Then the mean-generating function period extrapolation forecasting mode is employed to predict these approximate periodical signals. This method can effectively improve the prediction length and has higher modeling and prediction precision. Two representative examples are adopted in the simulation experiments, the simulation results show the correctness and validity of the method.

Key words: wavelet transform; MGF; period extrapolation; non-stationary time series; long-term forecasting

1 引言(Introduction)

非平稳时间序列分析与预测理论研究较为成熟的方法概括起来有两种: 一是Box和Jenkins的ARIMA模型, 它是ARMA模型的推广, ARIMA过程经有限次差分后变为ARMA过程, 以消除其非平稳特性(如趋势性和周期性), 然后进行分析与预测. 这种方法最大的缺点是丢掉了最重要的信息, 即趋势项和周期项. 其二是Kitagawa等人利用状态-空间模型及其相应的Kalman递归关系式对非平稳时间序列进行分析与预测, 其优点是将趋势项和周期项也包括进来考虑, 但是, 其建模过程大都是基于ARIMA模型向状态空间模型的转化^[1].

此外, 由于神经网络具有自学习、自适应的特点和强大的非线性逼近能力, 神经网络预测模型及其与传统预测理论结合的组合预测模型在非平稳时间序列分析与预测中得到了广泛的研究.

魏凤英等人拓展了数理统计中算术平均值的概念, 定义了时间序列的均生函数^[2]. 通过均生函数建立的模型作为一种有效的动态数据建模与预报途径在气象、水文、海洋等领域得到了应用.

目前, 小波分析已成为非线性科学中强有力的工具, 它的多分辨率特性, 且在时频两域都具有表征信号局部特征的能力, 使得小波分析在信号分析与预测领域中得到了越来越多的应用^[3].

本文利用小波分析方法对非平稳时间序列进行分解与重构,将复杂的非平稳时间序列简化为相对简单的准周期信号,即分离出趋势项、周期项和随机项.然后对于这些从低频到高频的准周期信号采用均生函数周期外推模式分别进行建模与预测,最后再集成得到原时间序列的预测值.

相比较而言,基于统计方法或神经网络方法进行预报时,大部分为一次性一步预报,随着一次性预见期的增加,误差会很快增加,影响一次性预报的准确率.另外一种方法则是获得一步或几步预报值之后,利用预报值作为下一次的训练数据再进行预报,这样获得较长时间的预报.但是由于预报值被利用作为训练数据,预报的误差也在训练中被累加,预报准确度也不高.上述方法在预报检验时采用的依然是一步或几步预报模型.本文所采用的方法,则可大幅度增加一次性预见期,实行长期预报.

2 小波分析理论(Wavelet analysis theory)

2.1 小波及小波变换(Wavelet and WT)

小波是由一母小波(Mother Wavelet) $\psi(t)$ 经伸缩和平移得到的一小波函数族^[4~7]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}, a \neq 0. \quad (1)$$

在此小波序列中: a 为伸缩因子, b 为平移因子.

在实际运用中,需要对伸缩因子和平移因子进行离散化,同时为使小波变换具有可变化的时间和频率分辨率,适应待分析信号的非平稳性,在实际中采用的是二进制动态采样网格: $a = 2^j, b = 2^j k$,则式(1)变为二进小波(dyadic wavelet):

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), j, k \in \mathbb{Z}. \quad (2)$$

对于任意的函数 $f(t) \in L^2(\mathbb{R})$ ($L^2(\mathbb{R})$ 表示平方可积的实数空间)的离散小波变换为

$$W_f(j, k) = \langle f, \psi_{j,k} \rangle = \frac{1}{\sqrt{C_\psi}} \int f(t) \psi_{j,k}^*(t) dt, \quad (3)$$

其重构公式为

$$f(t) = \frac{1}{\sqrt{C_\psi}} \sum_{j,k} W_f(j, k) \psi_{j,k}(t). \quad (4)$$

式中: $*$ 表示复共轭, C_ψ 表示允许条件.

2.2 多分辨率分析(Multi-resolution analysis)

S.Mallat在构造小波正交基时提出了多分辨率分析(MRA)的概念.多分辨率分析可以一个3层分解结构为例进行说明,其小波分解树如图1所示.对一信号 S ,多分辨率分析只是对低频部分(a_1, a_2, a_3)(不同尺度下的逼近信号)进一步分解,而高频部分(d_1, d_2, d_3)(不同尺度下的细节信号)则不予考虑.分解具有关系式: $S = a_3 + d_3 + d_2 + d_1$ ^[4].小波的多分辨率分析功能就是利用小波的分解和重构,在不同尺度

上提取信号的各种特征,从而可以根据需要以不同的分辨率处理信号.

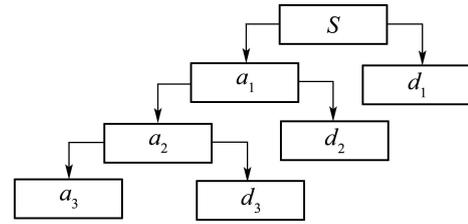


图1 小波分解树

Fig. 1 Wavelet decomposition tree

3 均生函数周期外推预测模式(MGF period extrapolation forecasting model)

3.1 均生函数(Mean-generating function)

在时间序列分析中,常用的是自回归和自回归滑动平均等模型,这些模型都着眼于序列中邻接时刻的联系,即序列的自相关.均生函数模型则从另一角度来建模,即基于序列的周期记忆,构造一组源自序列的周期函数,通过建立时间序列与这组周期函数间的回归,构造预报模型^[8].

设样本量为 N 的时间序列:

$$x(t) = \{x(1), x(2), \dots, x(N)\}, \quad (5)$$

$x(t)$ 的均值为

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x(t). \quad (6)$$

对于式(5)定义均生函数(均值生成函数)

$$\bar{x}_l(i) = \frac{1}{n_l} \sum_{j=1}^{n_l-1} x(i+jt), i = 1, \dots, l, 1 \leq l \leq M. \quad (7)$$

式中: $n_l = \text{int}(N/l)$, $M = \text{int}(N/2)$ 为不超过 $N/2$ 的最大整数.int表示取整.

根据式(7),得到 M 个均生函数,将均生函数定义域延拓到整个数轴上,即作周期性延拓得

$$f_l(t) = \bar{x}_l(t - l * \text{int}(\frac{t-1}{l})). \quad (8)$$

其中: $l = 1, 2, \dots, M, t = 1, 2, \dots, N, N+1, \dots$.笔者称 $f_l(t)$ 为均生函数延拓序列,其含意是这一函数是对观测数据通过用公式(8)计算随后作周期延拓得到.由此可构造出均生函数延拓矩阵

$$F = (f_{ij})_{N \times M}, f_{ij} = f_i(t). \quad (9)$$

3.2 周期外推预测模式(Period extrapolation forecasting model)

将上述时间序列 $x(t)$ 以均生函数为基展开,即

$$x(t) = \sum_{l=1}^k f_l(t) + e(t), k < M. \quad (10)$$

由于 $f_l(t)$ 为周期函数,这也表明 $x(t)$ 是由若干周期函数叠加而成, $e(t)$ 为误差项.

若将各周期所对应的均生函数作 q 步外延, 则可得到均生函数延拓矩阵

$$F = (f_{ij})_{(N+q) \times M} \quad (11)$$

和 q 步预报值

$$x(N+q) = \sum_{l=1}^k f_l(N+q). \quad (12)$$

3.3 组合模式预报方法(Combination model forecasting approach)

基于小波变换和均生函数周期外推组合模式预报结构原理如图2所示.

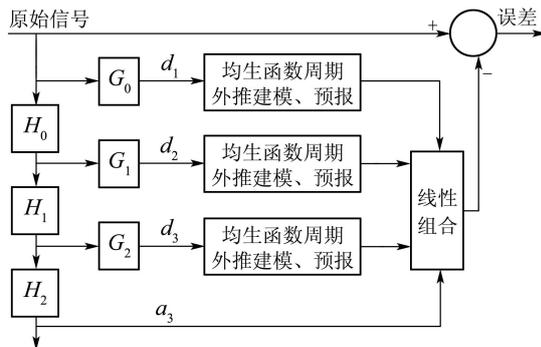


图 2 组合模式预报结构图

Fig. 2 Scheme of combination model forecasting

组合模式预报的步骤为

1) 对给定时间序列进行多尺度小波分解, 将复杂多周期信号简化为若干相对简单的准周期信号. 假设分解尺度为 J 则输出信号为 $d_1, d_2, \dots, d_J, a_J$. 低频信号 a_J 反映了信号的发展变化趋势, 相对高频信号 d_1, d_2, \dots, d_J 为周期性和随机性成分.

2) 对于小波分解序列分别采用基于均生函数模型周期外推预报方法进行建模预报, 若原函数具有明显线性趋势, 则可采用线性趋势进行预报, 可参考有关文献. 基于均生函数模型周期外推预报方法如下:

① 对数据分别进行规格化处理, 可采用标准差规格化处理方法进行规格化.

② 计算各规格化序列的均生函数, 并利用公式(8)和(9)作周期性延拓, 得到均生函数延拓序列 $f_i(t)$ 和延拓矩阵 F , 然后利用公式(12)对均生函数进行外推, 获得各变量的建模与预报值. 为提高建模及预报精度, 本方法结合最小二乘算法(LMS)进行建模拟合并修正预报值, 即在获得均生函数延拓矩阵 F 后, 利用最小二乘算法对建模数据进行拟合辨识出模型系数:

$$W = (F^T F)^{-1} F^T X(t). \quad (13)$$

然后利用辨识出的模型系数修正预报值.

3) 对所获得的各变量建模及预报值进行线性集成, 即可得到原序列的拟合和预报值.

4 应用举例(Application examples)

本文选取两组时间序列进行预报验证. 第1组为多个不同频率的标准正弦函数的叠加信号; 第2组为长江三角洲1951 ~ 1998年的降水量的实际值.

实例 1 不同频率的标准正弦函数叠加信号:

$$x = \sin t + \sin(10 * t) + \sin(100 * t), \quad (14)$$

从上式可知, 信号的最大频率为16 Hz, 则它的奈奎斯特频率为32 Hz, 采样频率应选择32 Hz以上, 本仿真采样时间间隔选为0.03, 以保证用于小波分析的采样数据的真实完整性.

对原信号进行小波分解, 本文选择db3小波, 该小波具有正交性、紧支撑性和近似对称性, 可进行离散小波变换. 分解结果如图3所示.

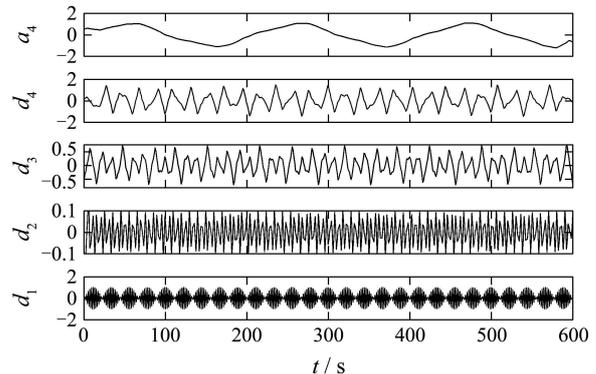


图 3 正弦叠加信号的小波分解

Fig. 3 Wavelet decomposition of sum of sinusoidal signals

若将信号中的最高频率看作是1, 则各层小波分解便是带通或低通滤波器, 且各层所占的频率为

$$\begin{aligned} a_1 &: 0 \sim 0.5, & d_1 &: 0.5 \sim 1, \\ a_2 &: 0 \sim 0.25, & d_2 &: 0.25 \sim 0.5, \\ a_3 &: 0 \sim 0.125, & d_3 &: 0.125 \sim 0.25, \\ a_4 &: 0 \sim 0.0625, & d_4 &: 0.0625 \sim 0.125. \end{aligned}$$

该信号在小波分解下的相对频率分别为: 高频正弦(1), 定位于 d_1 层; 中频正弦(0.1), 定位于 d_4 层; 低频正弦(0.01), 定位于 a_4 层.

本仿真实验选取600个数据点, 前500个点为建模序列, 后100个点为预报检验序列. 各序列预报结果如图4所示, 图中实线为实际信号, 虚线为预报输出.

以 a_4 预报为例, 它反映 $\sin t$ 的曲线的变化, 信号的频率为 $\omega_{a_4} = 1/2\pi$, 则周期为 $T_{a_4} = 2\pi$, 故信号周期占 $2\pi/0.03 = 209$ 个点, 在均生函数延拓建模中, 初始选择 $M = 500/2 = 250$, 选取不同个数均生函数建模均方根误差(MSE)如图4左半部分所示. 实际上在 $M = 190$ 点时, $MSE_{a_4} = 0.0041$ 已经稳定, 基本不会随 M 的增加而减小多少. 190点和209的差异为小波分解后的 a_4 与原信号中低频成分 $\sin(t)$ 的差异, 所以用于均生函数的建模点与周期基本吻合.

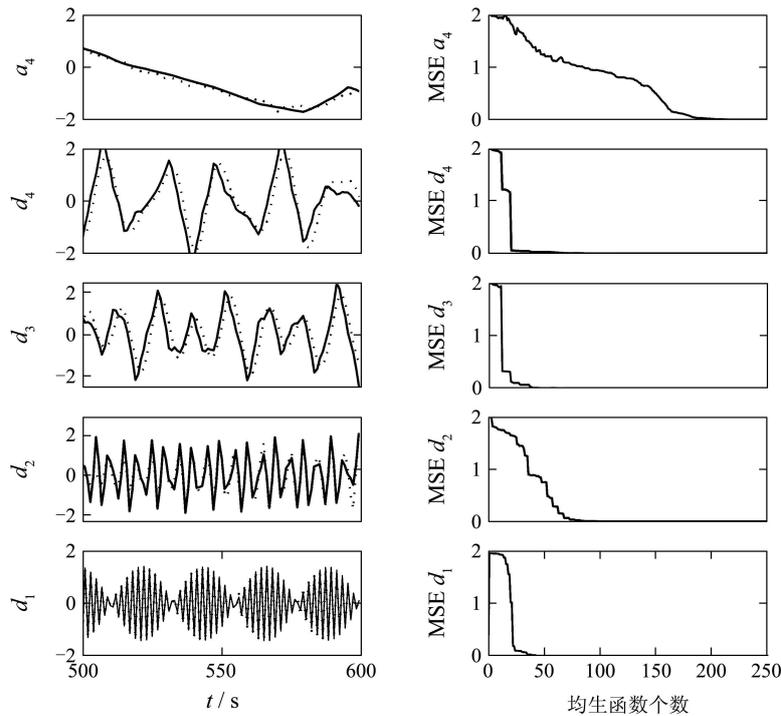


图4 预报及建模均方根误差曲线

Fig. 4 Prediction and modeling MSE curve

各信号分析情况如表1(表中建模和预报均方根误差数据为仿真序列经过标准化处理后所得结果).

表1 建模及预报分析表

Table 1 Modeling and predication analysis

信号	图示对应周期	采用均生函数个数	MSE (建模)	MSE (预报)
a_4	209点	190	0.0041	0.0107
d_4	21点	22	0.0670	0.1984
d_3	约12点	13	0.3300	0.6056
d_2	约5点	85	0.0172	0.1969
d_1	<2.5点	44	0.0002	0.0148

从图4中对各变量的均方根误差分布曲线可以观察到, a_4 和 d_4 的均方根误差呈均匀下降趋势, 采用均生函数建模的个数与数据周期所占点数吻合, 这是因为 a_4 和 d_4 均反映的是单一频率信号 $\sin t$ 和 $\sin(10t)$; d_3 的均方根误差有两次跳跃式下降, 即采用均生函数建模点数为13点和22点, 前者与数据周期所占点数吻合, 后者为两倍, 这表明数据 d_3 由两个周期的数据叠加而成. d_2 的均方根误差约有两次跳跃式下降, 即采用建模点数为36点、85点, d_1 的均方根误差在44点时呈直线下降, 即采用建模点数为44点时建模与预报误差均很小, 为什么会产生这样的结果呢? 从小波分解的情况看, d_1 信号存在着多个信号包络, 每个信号包络占有约22个点, 这也是为什么建模的信

号点为44(两个包络)的缘故, 同时可以看到, 每个信号包络有约十个正弦振荡, 每个正弦振荡占有约2个点, 它与 d_1 信号的周期相同, 故 d_1 信号反映的是 $\sin(100 * t)$ 的高频信号.

图5为小波分解序列预报值经过线性集成得到的结果, 图中实线为实际信号, 虚线为预报输出. 仿真得出的均方根预报误差为0.0204.

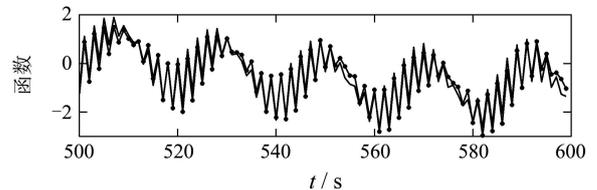


图5 集成预报曲线

Fig. 5 Integration prediction curve

表2 不同预报方法的比较

Table 2 Comparison of different forecasting approaches

预报方法	采用均生函数个数	MSE (建模)	MSE (预报)
MGFPE	209(最佳数, 实际为信号低频成分周期)	0.0067	0.0916
WT and MGFPE	a_4, d_4, d_3, d_2, d_1 分别为: 190, 22, 13, 85, 44	0.0063	0.0204

为了验证方法的有效性, 表2将预报结果与仅用均生函数算法预报进行了比较, 结果表明: 小波与均生函数组合预报方法在建模拟合和预报上优

于传统均生函数方法.

实例 2 长江三角洲降水数据序列从1951 ~ 1998年6 ~ 8月的降水总量为^[8]

[513, 447, 461, 764, 485, 601, 631, 356, 307, 508, 385, 519, 419, 320, 487, 323, 221, 290, 497, 474, 365, 423, 372, 556, 516, 408, 460, 169, 391, 717, 407, 517, 456, 488, 424, 511, 669, 403, 546, 413, 847, 382, 660, 361, 534, 622, 516, 469].

利用前42个数据建模, 后6个进行预报检验. 序列经小波变换后如图6所示, a_4 反映的是序列的趋势变化.

各信号的预报及建模误差随均生函数建模点数不同的误差变化情况如图7所示, 图中实线为实际信号, 虚线为预报输出.

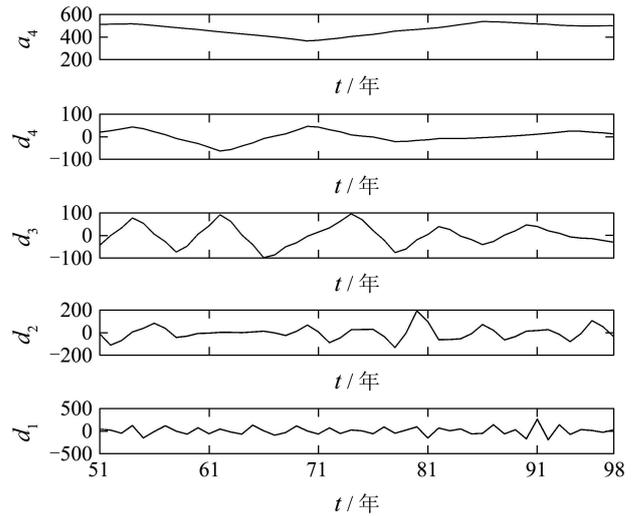


图 6 信号小波分解

Fig. 6 Wavelet decomposition of signal

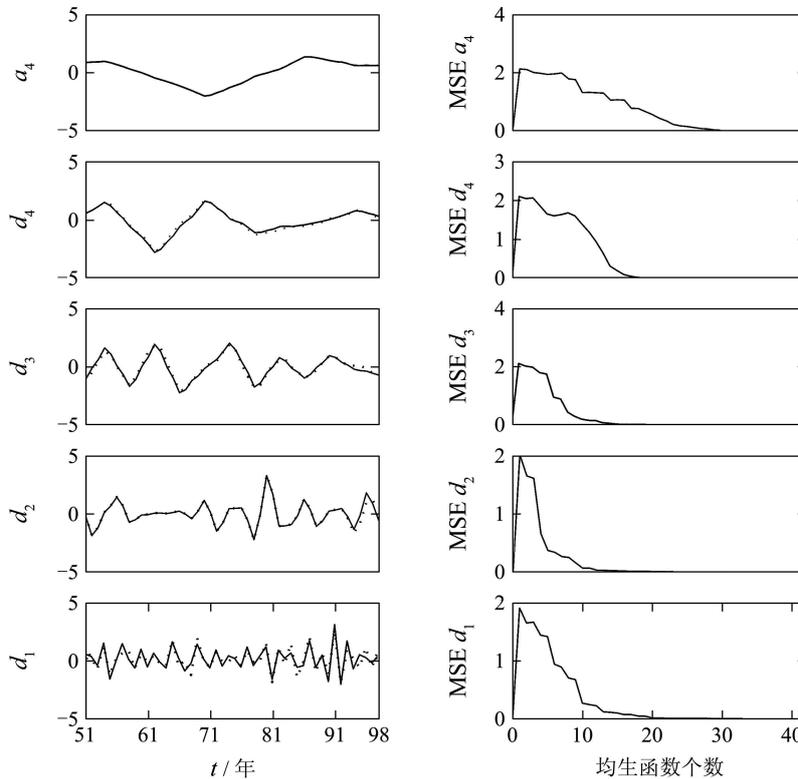


图 7 预报及建模均方根误差曲线

Fig. 7 Prediction and modeling MSE curve

在图7中 a_4 的周期大于建模点的一半, 如在均生函数延拓建模中, 取初始选择 $M = 42/2 = 21$, 则不能很好反映周期外推的思想, 使得建模与预报误差较大. 本实验初始选取 $M = 42$, 最后, 建模点到30时, 建模及预报误差基本不再下降, 也与 a_4 的周期相吻合. d_4, d_3, d_2, d_1 在均生函数延拓建模中, 均取初始选择 $M = 42/2 = 21$. d_4 和 d_3 的建模点与序列周期点基本吻合; d_2 的均方误差有

两次剧烈变化, 分别约为5点和13点, 说明 d_2 由这两个周期的信号叠加, 周期外推是对这两周期信号外推的, 其中5点周期在原信号上基本可以观测到. d_1 的均方误差有多次剧烈变化, 分别约为3点、5点、11点, 其中3点为显示周期, 可从分解图中观测得到. 信号分析情况如表3(表中建模和预报均方根误差数据为仿真序列经过正则化处理后所得结果).

表3 建模及预报分析表

Table 3 Modeling and predication analysis

信号	图示对应周期	采用均生函数个数	MSE (建模)	MSE (预报)
a_4	约30点	30	0.0002	0.0116
d_4	约20点	19	0.0222	0.0202
d_3	约10点	14	0.0303	0.0646
d_2	约5点	13	0.0002	0.4078
d_1	约3点	11	0.2264	0.6354

图8为各序列预报经过线性集成得到的结果,图中实线为实际信号,虚线为预报输出.从图中可以观测得到:建模及预报效果都相当好,尤其是极值年份都能准确拟合和预报.

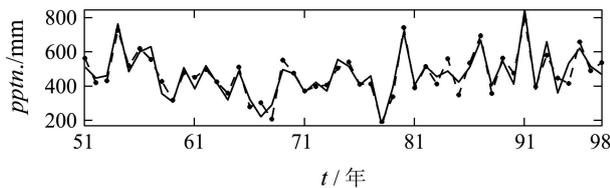


图8 集成预报曲线

Fig. 8 Integration prediction curve

表4为利用小波与均生函数结合进行预报同仅用均生函数的建模与预报结果比较,可以看出,本文提出的方法具有一定的优势,同时对于这些多周期混叠的时间序列,小波分析能够避免均生函数寻找优势周期的困难.

表4 不同预报方法的比较

Table 4 Comparison of different forecasting approaches

预报方法	采用均生函数个数	MSE (建模)	MSE (预报)
MGFPE	30	15.1781	33.7527
WT 和 MGFPE	a_4, d_4, d_3, d_2, d_1 分别为: 30, 19, 14, 13, 11	8.4565	16.5821

5 结论(Conclusion)

本文提出了利用小波变换和均生函数周期外推组合模式进行时间序列长期预测的方法,通过小波变换可把不确定及多周期混叠的信号转化为多个相对简单的准周期信号,获得原信号的趋势、周期和随机序列,然后采用均生函数周期外推建模预报模式对这些准周期信号进行预报.由于

序列的周期性和采用周期外推模式,故预报长度可大幅度增加.本文采用两个典型实例进行验证,结果都非常理想.但在仿真时发现:采用不同类型的小波或不同级别的小波分解都会对预报结果有一定的影响,本人在仿真中多是结合经验选择小波,如何从理论上解决这一问题也是下一步要研究的问题.

参考文献(References):

- [1] 马社祥,刘贵忠,曾召华.基于小波分析的非平稳时间序列分析预测[J].系统工程学报,2000,15(4):305-311.
(MA Shexiang, LIU Guizhong, ZENG Zhaohua. Wavelet-based analysis and forecasting for non-stationary time series[J]. *Journal of Systems Engineering*, 2000, 15(4): 305-311.)
- [2] 魏凤英,曹鸿兴.长期预测的数学模型及应用[M].北京:气象出版社,1990.
(WEI Fengying, CAO Hongxin. *Mathematical Models of Long-term Prediction and Its Application*[M]. Beijing: Meteorology Press, 1990.)
- [3] 胡昌华,张军波,夏军等.基于MATLAB的系统分析与设计—小波分析[M].西安:西安电子科技大学出版社,1999,12.
(HU Changhua, ZHANG Junbo, XIA Jun. *System Analysis and Design Based on MATLAB—Wavelet Analysis*[M]. Xi'an: Xidian University Press, 1999, 12.)
- [4] MALLAT S G. A theory of multi-resolution signal decomposition: The wavelets representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674-693.
- [5] DAUBECHIES I. The wavelet transform, time-frequency localization and signal analysis[J]. *IEEE Transactions on Information Theory*, 1990, 36(5): 961-1005.
- [6] 程正兴.小波分析算法与应用[M].西安:西安交通大学出版社,1998.
(CHENG Zhengxing. *Wavelet Analysis Algorithm and Application*[M]. Xian: Xi'an Jiaotong University Press, 1998.)
- [7] 彭玉华.小波变换与工程应用[M].北京:科学出版社,1999.
(PENG Yuhua. *Wavelet Transform and Engineering Application*[M]. Beijing: Science Press, 1999.)
- [8] 曹鸿兴.动力系统自忆原理—预报和计算应用[M].北京:地质出版社,2002.
(CAO Hongxing. *Self-memorization Principle of Dynamic Systems—Prediction and Computation Application*[M]. Beijing: Geology Press, 2002.)

作者简介:

李 晖 (1968—),男,博士,大连海事大学自动化与电气工程学院副教授,目前研究方向为系统建模与仿真、船舶运动控制与虚拟现实仿真, E-mail: delxf@263.net;

郭 晨 (1956—),男,博士,大连海事大学自动化与电气工程学院教授,博士生导师,目前研究方向为船舶自动控制系统、船舶轮机系统仿真、智能控制理论与应用、虚拟现实技术及应用;

金鸿章 (1946—),男,哈尔滨工程大学自动化学院教授,博士生导师,目前研究方向为船舶控制与船舶装置、控制理论、复杂系统.