

经验模态分解法在大气时间序列 预测中的应用

玄兆燕^{1,2} 杨公训¹

摘要 介绍了一种可以提高非平稳时间序列预测精度的新方法, 该方法应用 Hilbert-Huang 变换的核心内容 — 经验模态分解法 (Empirical mode decomposition, EMD) 对非平稳时间序列进行分解, 以降低被预测信号中的非平稳性, 利用神经网络对分解后的各分量进行预测, 再将预测结果叠加. 利用该方法对石家庄市年逐月降水量进行预测, 预测结果显示, 其预测精度比直接用神经网络预测的预测精度有较明显的提高.

关键词 Hilbert-Huang 变换, 预测, 非平稳性, 非线性, 经验模态分解法 (EMD), 人工神经网络 (ANN), 时间序列

中图分类号 TP13

Application of EMD in the Atmosphere Time Series Prediction

XUAN Zhao-Yan^{1,2} YANG Gong-Xun¹

Abstract In this paper, a new method to improve non-stationary time series prediction accuracy is introduced. The non-stationary time series is decomposed by empirical mode decomposition (EMD) in Hilbert-Huang transform to reduce the non-stationarity in the signals. By using neural network, the component of decomposition is predicted, then the predicted results are added. The author has predicted monthly precipitation data at Shijiazhuang with the method. The study shows that the prediction accuracy of the neural network based on EMD is higher than that of prediction method using the neural network.

Key words Hilbert-Huang transform, prediction, non-stationary, non-linear, empirical mode decomposition (EMD), artificial neural network (ANN), time series

自然界和社会系统的非线性本质决定了非线性时间序列预测的重要性. 到目前为止, 已有一些非线性数学模型被用来进行非平稳时间序列的预测, 如 TAR 模型、基因线性回归、灰色模型、投影寻踪回归、未确知模拟模型、模糊预测、组合预测等预测模型都有各自的特点, 但对于有些复杂非线性系统, 有时很难建立合适、有效的数学模型, 如一些环境系统、气候系统等, 所以使相应这些领域的预测结果总是不能令人满意^[1], 近年来发展起来的模拟智能结构过程的人工神经网络 (Artificial neural network, ANN) 技术, 具有适合于研究复杂非线性系统和不确定过程的优点, 因此在很多领域得到了较为广泛的应用. 但信号的非平稳性给神经网络预测也带来了一些难度, 中国科学院大气物理研究所杨培才最近提出了一种建立在经验模态分解法 (Empirical mode decomposition, EMD) 基础上的隔离预测模式, 该预测模式的基本思想是借助 EMD 方法, 把时间序列分解成有限个具有特征振荡周期的分量, 尽管有些分量仍然保持着不同程度的非平稳性, 但是, 在它们之间的相互影响却被隔离了, 利用这种隔离可以尽可能地减小非平稳行为对预测的影

收稿日期 2006-09-19 收修改稿日期 2007-02-06

Received September 19, 2006; in revised form February 6, 2007

1. 中国矿业大学 北京 100080 2. 河北理工大学机械工程学院 唐山 063009

1. China University of Mining and Technology, Beijing 100080 2. Hebei Polytechnic University, Tangshan 063009

DOI: 10.3724/SP.J.1004.2008.00097

响^[2]. EMD 分解是 Hilbert-Huang 变换中的一种信号分解方法, Hilbert-Huang 变换是美籍华人 Huang 等在 1996 年提出的适用于非平稳时间信号的一种新的分析方法, 这一方法从根本上摆脱了傅里叶变换理论的束缚, 很好地诠释了瞬时频率的概念^[3-4]. EMD 方法将被分析的数据分解成多个固有模态信号 (Intrinsic mode function, IMF) 的组合, 它是 Hilbert-Huang 变换的核心内容之一.

本文将 EMD 方法与神经网络相结合, 用于对石家庄市年逐月降水量进行预测.

1 非平稳信号的经验模态分解 (EMD)

Huang 等人针对非平稳信号提出了一种经验模态分解法 (EMD), 此种方法基于以下的设定:

- 1) 被分析信号至少包含两个极值点: 一个极大值和一个极小值;
- 2) 根据两个相邻极值点的时间距离定义固有振荡模式的特有时间尺度.

该方法的原理是: 根据固有模态信号的定义 (所谓固有模态信号是指满足以下两个条件的信号: 1) 整个数据中, 零点与极点数相等或至多相差 1; 2) 信号上任意一点, 由局部极大值点确定的包络线和由局部极小值点确定的包络线的均值均为 0, 即信号关于时间轴局部对称), 首先找到信号 $X(t)$ 的所有局部极大值和局部极小值, 用三次样条插值分别连接所有极大值点和所有极小值点, 形成信号的上包络线和下包络线, 使信号的所有数据点都在两包络线之间, 然后求出两包络线的平均值 m_1 , 再用 $X(t)$ 减 m_1 得到 h_1 , 即

$$X(t) - m_1 = h_1 \quad (1)$$

这里, h_1 还不是一个 IMF, 筛分过程主要有两个目的: 1) 消除一些附加的波; 2) 使波的轮廓更对称. 为此, 筛分过程可以被重复多次, 第 2 次筛分时 h_1 被作为分解数据, 即

$$h_1 - m_{11} = h_{11} \quad (2)$$

重复筛分过程 k 次, 直到 h_{1k} 满足 IMF 的条件, 此时

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (3)$$

筛分出的第一个 IMF 表示为

$$c_1 = h_{1k} \quad (4)$$

一个信号可以由多个 IMF 组成, 可以从信号中先将 c_1 分离出来, 即

$$X(t) - c_1 = r_1 \quad (5)$$

然后将 r_1 作为分解数据, 再重复以上的筛分过程得到 c_2 , 即

$$r_1 - c_2 = r_2 \quad (6)$$

这个过程可以一直重复, 即

$$r_{n-1} - c_n = r_n \quad (7)$$

直到满足如下条件之一时, 分解过程可以结束: 1) r_n 或 c_n 小于给定值; 2) r_n 成为单调函数, 不可能再从中筛分出 IMF. 最后得到

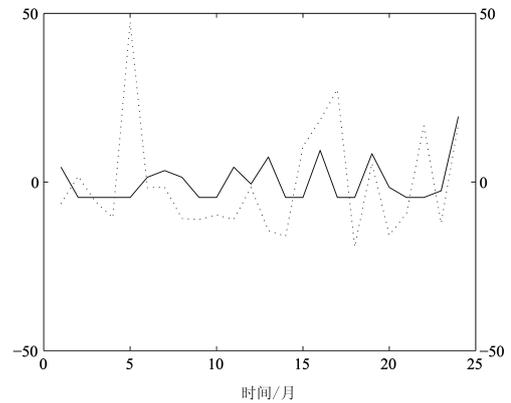
$$X(t) = \sum_{i=1}^n c_i + r_n \quad (8)$$

2 降水量时间序列的神经网络预测

神经网络因为其较强的适应能力和学习能力, 使之在非线性的预测方面得到了广泛的应用. 考虑到大气系统是一类非线性较强的时间序列, 本文选用 RBF 网络, 该网络是一种前向型神经网络, 它具有较高的运算速度和外推能力, 同时具有较强的非线性映射功能, 它通过非线性基函数的线性组合实现输入空间 \mathbf{R}^n 到输出空间 \mathbf{R}^m 的非线性转换.

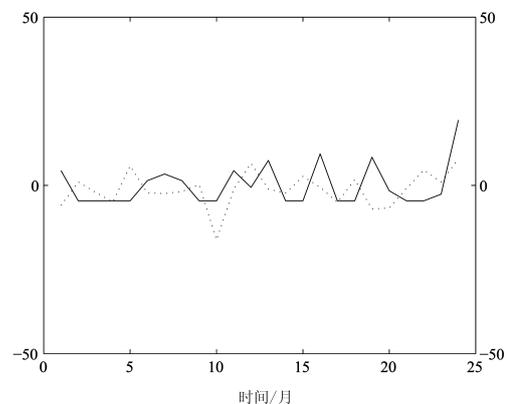
将石家庄市从 1951 年至 2002 年共 52 年的逐月降水距平观测数据进行样本划分. 用前 15 年共 180 个数据来预测后两年 24 个数据. 共抽取 420 个样本进行网络测试.

用测试好的网络预测 2001 年、2002 年两年共 24 个数据, 预测结果如图 1(a) 所示. 图中虚线为预测值, 实线为实际观测值. 预测误差曲线如图 2(a) 所示. 由图中可以看出, 因大气系统的非平稳性使得该预测产生了较大的误差.



(a) 直接神经网络预测

(a) Neural network prediction



(b) EMD 分解后的预测

(b) Neural network prediction based on EMD

图 1 石家庄市逐月降水序列的预测结果

Fig. 1 Prediction of monthly precipitation data at Shijiazhuang

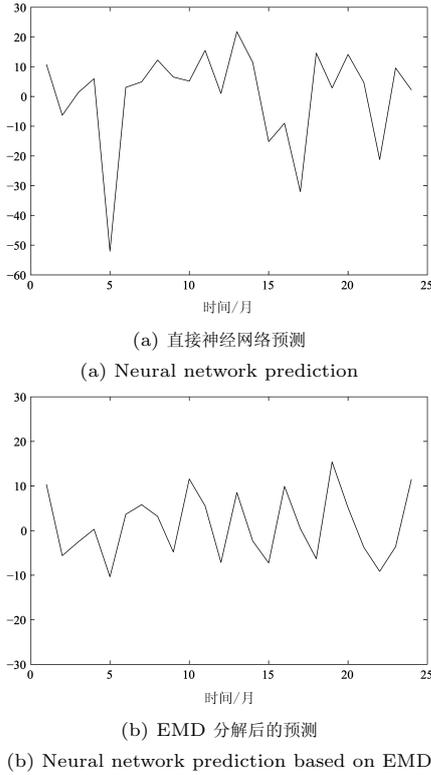


图 2 石家庄市逐月降水序列的预测误差

3 经验模态分解在石家庄市年逐月降水量预测中的应用

据一些大气科学家分析, 控制天气和气候系统的外部条件不是一成不变的, 随着时间的变化, 大气的基本状态或平稳性质会随之改变, 所以大气系统的动力学本质上是非平稳的, 因此利用神经网络对该数据进行预测, 数据的非平稳性使得其预测结果很不理想, 如图 1(a) 所示. 为了提高预测精度, 用 EMD 方法对信号进行分解, 对分解后的各分量分别进行预测, 以降低其非平稳性对预测精度的影响.

3.1 石家庄市年逐月降水量时间序列的经验模态分解

图 3 为石家庄市从 1951 年至 2002 年共 52 年的逐月降水距平观测数据, 该数据长度为 624 (52 年共 624 个月).

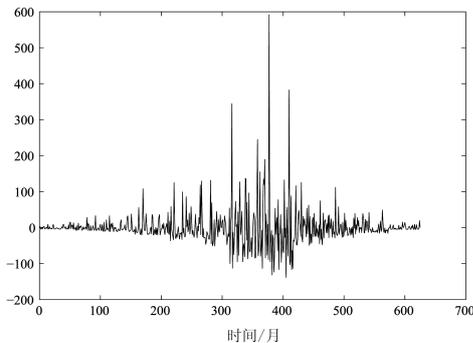


图 3 石家庄市 52 年逐月降水距平数据

根据式 (1)~(8) 的分解算法, 利用 HHT-DSP 软件将其进行 EMD 分解, 分解结果如图 4 所示. 共分解出 8 个 IMF, 图 4 中从上至下分别为 IMF 1 到 IMF 8, 最后一个为剩余量. 若将原始数据用 $X(t)$ 表示, 则图 4 的分解结果可以表示为

$$X(t) = \sum_{i=1}^n c_i + r_n \quad (9)$$

其中 c_i 代表 IMF i , r_n 代表剩余分量.

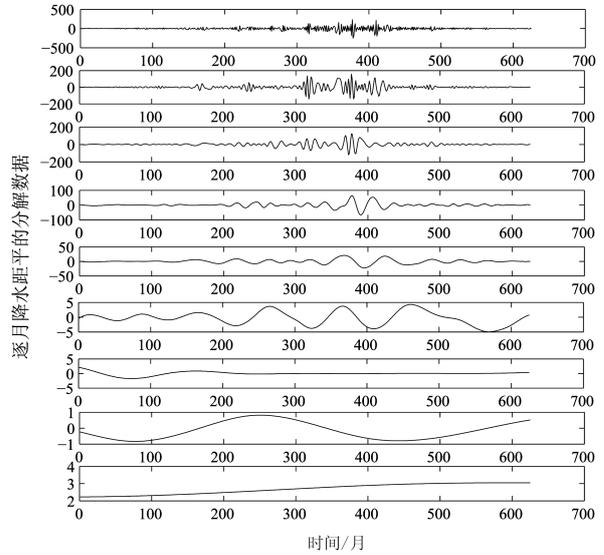


图 4 图 3 数据的 EMD 分解结果

根据杨培才提出的 EMD 基础上的隔离预测模式, 以上的分解隔离了 IMF i 之间的相互影响, 随着分解次数的增加, 分量的非平稳行为被减小, 将各个分量分别预测, 再将其预测结果叠加, 会增强信号的可预报性.

3.2 基于 EMD 的神经网络预测

EMD 将石家庄市逐月降水序列分解成 8 个 IMF 和一个剩余量, 如图 4 所示. 利用上面创建的神经网络对 9 个分量分别进行预测, 结果如图 5 所示, 图 5(a)~5(i) 依次为剩余量、IMF 8~IMF 1 的预测结果, 图中实线为实际观测值, 虚线为预测值.

从图中可以看出, 低频分量的预测结果很理想, 最小预测误差达到 10^{-5} 数量级, 预测误差最大的为 IMF 1, 这也说明了高频分量中信号的非平稳性要高一些, 将各分量合成后的预测结果及预测误差曲线见图 1(b) 和图 2(b) 所示.

直接神经网络预测和基于 EMD 的神经网络预测的 24 点数据的预测误差如表 1 所示, 从表 1 中可知: 1) 直接神经网络预测的最大预测误差绝对值出现在第 5 个预测数据点上, 误差值为 -52.025; 2) 基于 EMD 的神经网络预测的最大预测误差绝对值出现在第 19 个预测数据点上, 误差值为 15.406; 3) 直接神经网络预测的误差值, 分别第 5、11、13、15、17、18、20、22 共 8 个预测数据点上的预测误差绝对值大于或接近基于 EMD 的神经网络预测的最大误差值; 4) 直接神经网络预测的最小预测误差绝对值出现在第 12 个预测数据点上, 其值为 1.0088; 5) 基于 EMD 的神经网络预测的最小预测误差绝对值出现在第 4 个预测数据点上, 其值为 0.30043; 6) 直接神经网络预测的预测误差绝对值的

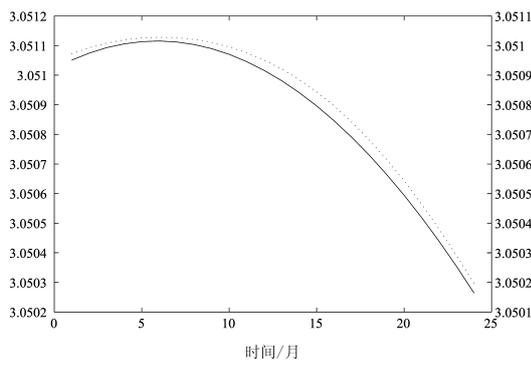
平均值为 11.6; 7) 基于 EMD 神经网络预测的预测误差绝对值的平均值为 6.43.

从以上比较中可知, 对于具有非平稳性的降水量时间序列的预测, 基于 EMD 的神经网络预测要优于直接神经网络预测.

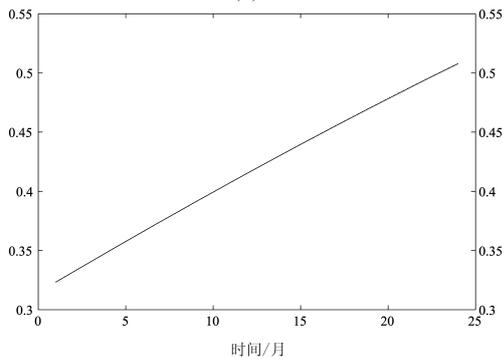
表 1 直接神经网络预测误差值与基于 EMD 神经网络预测误差值比较

Table 1 Neural network prediction error and neural network prediction error based on EMD

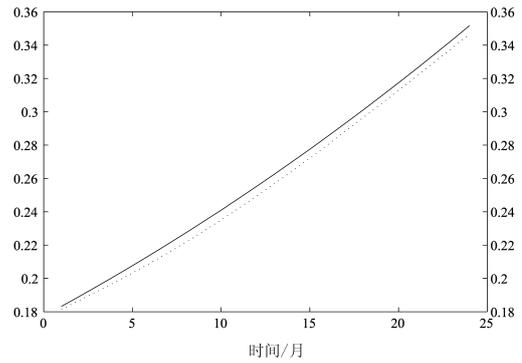
预测点	直接神经网络误差	EMD 神经网络误差
1	10.835	10.344
2	-6.3117	-5.637
3	1.2885	-2.5368
4	5.9994	0.30043
5	-52.025	-10.369
6	3.0768	3.6436
7	4.9185	5.804
8	12.229	3.1893
9	6.521	-4.8219
10	5.1852	11.581
11	15.52	5.5306
12	1.0088	-7.1427
13	21.777	8.5342
14	11.486	-2.3439
15	-15.179	-7.2665
16	-8.9956	9.9331
17	-32.009	0.44684
18	14.599	-6.3175
19	2.8751	15.406
20	14.15	5.1468
21	4.7314	-3.7659
22	-21.241	-9.1476
23	9.5693	-3.6321
24	2.2286	11.502



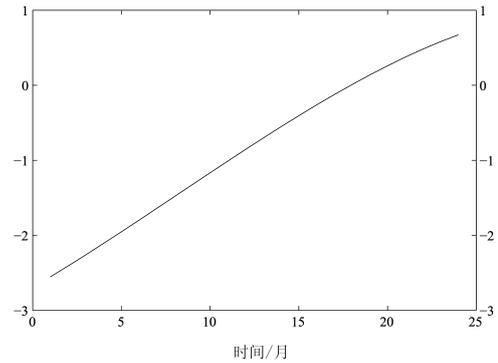
(a)



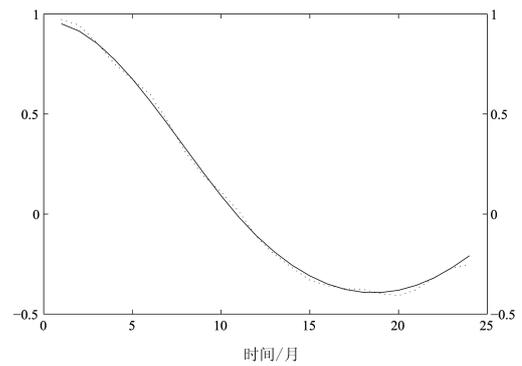
(b)



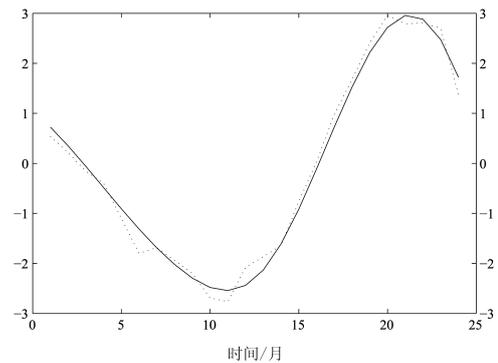
(c)



(d)



(e)



(f)

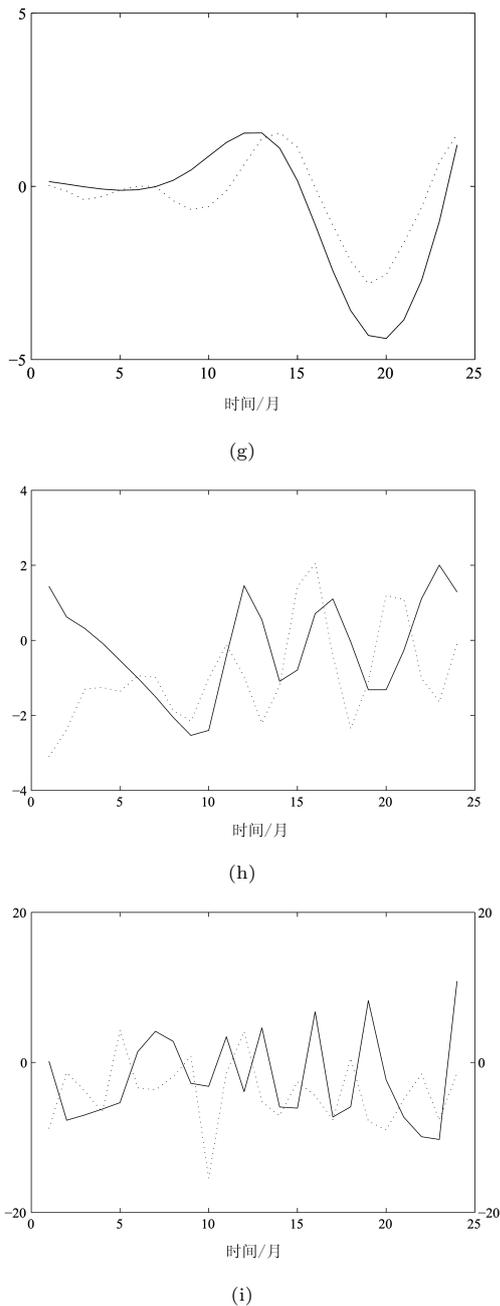


图5 石家庄市逐月降水序列分解的8个IMF和一个剩余量的预测结果(实线:观测值,虚线:预测值)

Fig. 5 Prediction of monthly precipitation data at Shijiazhuang based on EMD (Solid line is representation of observation, while the dotted line is the prediction.)

4 结论

1) 对于非线性系统,神经网络预测有着明显的优势,但复杂非线性系统中,信号的非平稳行为确实给预测带来了一定的困难。

2) EMD分解确实降低了各分量的非平稳性,特别是较低频的IMF分量预测结果非常理想,从剩余分量到IMF4的

最小预测误差为 10^{-5} 数量级,最大误差也在 10^{-1} 数量级以下。

3) 从图5中可以看到,最终的预测误差,很大程度上取决于最高频率分量IMF1,如果对IMF1分量的预测方法再作进一步改进,整个系统的预测精度会得到更大的提高。

实际上在诸如大气等复杂系统的预测中很少取得令人满意的效果,从对石家庄市年逐月降水量进行预测的过程中可以看到,利用EMD分解的神经网络预测与直接神经网络预测相比其预测精度确实有了明显的提高,通过对该方法的进一步研究,可以使其成为非线性复杂系统预测的发展方向。

References

- 1 Yang Xiao-Hua, Yang Zhi-Feng, Shen Zhen-Yao, Lu Gui-Hua, Li Jian-Qiang. High precision radial basis function network model for predicting water environment nonlinear time series. *Advances in Water Science*, 2005, **16**(6): 788–791 (杨晓华, 杨志峰, 沈珍瑶, 陆桂华, 郦建强. 水环境非线性时序预测的高精度RBF网络模型. *水科学进展*, 2005, **16**(6): 788–791)
- 2 Yang Pei-Cai, Zhou Xiu-Ji. On non-stationary behaviors and prediction theory of climate systems. *Acta Meteorologica Sinica*, 2005, (5): 556–570 (杨培才, 周秀骥. 气候系统的非平稳行为和预测理论. *气象学报*. 2005, (5): 556–570)
- 3 Huang N E, Wu M L, Qu W L, Shen S R. Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 2003, **19**(3): 246–268
- 4 Huang N E, Zheng S, Long S R, Wu M L, Shih H H, Zheng Q A. *The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis*. UK: Royal Society, 1998. 904–962

玄兆燕 中国矿业大学博士研究生. 主要研究方向为非线性复杂系统的预测. 本文通信作者. E-mail: xzyanan@126.com

(XUAN Zhao-Yan Ph.D. candidate at China University of Mining and Technology. His research interest covers prediction of nonlinearity complex system. Corresponding author of this paper.)

杨公训 中国矿业大学教授. 主要研究方向为控制理论与控制工程, 信号采集与处理. E-mail: ygx@cumtb.edu.cn

(YANG Gong-Xun Professor at China University of Mining and Technology. His research interest covers control theory and control engineering, signal acquisition and processing.)