

文章编号 1001-8166(2004)04-0572-05

西太平洋副高形态指数的分解重构与集成预测¹

张 韧^{1,2}, 董兆俊¹, 陈奕德¹, 何金海²

(1. 解放军理工大学气象学院, 江苏 南京 211101; 2. 南京气象学院大气科学系, 江苏 南京 210044)

摘 要:用小波分解和自适应神经模糊推理系统(ANFIS)相结合的方法,建立了西太平洋副热带高压形态指数月、季时间尺度的集成预报模型。由于小波分解可在信号的频域—时域内自由伸缩,准确地分解和重构带通、低通信号,因而能将复杂的副高指数时间序列分解为相对简单的周期分量信号,既简化了系统结构,又突出了信号特征。随后基于ANFIS模糊系统的非线性、容错性、自适应性和联想学习功能,建立各分量信号的独立预报模型,最后对分量预报结果进行集成。试验结果表明,该方法在保留预报对象主要特征的前提下,有效降低了预报难度,预报准确率和预报时效均较传统方法有明显的改进和提高。

关键词:副高指数;小波分解;模糊推理;集成预测

中图分类号 P456

文献标识码 A

西太平洋副高进退活动和形态的异常对夏季东亚地区的天气气候和旱、涝灾害有显著的影响。西太平洋副高强度和形态的预报一直是人们非常关注的重点和难点问题,尽管已揭示出许多副高活动的特征和规律,但对其预报目前仍未取得满意效果。在副高中长期预报中,统计预报方法仍是主要的预报手段,近年来迅速发展的人工神经网络方法在副高预报中也表现出明显的优势和积极的应用前景,但有关副高形态指数的预报研究并不多,效果也不尽如人意^[1-3]。由于常规的统计预报模型大多是建立在线性相关基础之上的,因此在处理复杂的非线性问题时存在较大的局限性;人工神经网络(ANN)是由人工建立的以有向图为拓扑结构的非线性动力系统,它通过对连续或者断续的输入做状态响应而进行信息处理,其特色在于信息的分布式存储和并行协同处理。大气科学中已有不少利用人工神经网络模型进行预报的应用研究^[1-6]。在理论上,基于前传多层网络算法的神经网络模型能够在相当精度上逼近任意非线性映射和复杂动力系统,这一点在

数学上已得到了严格的证明^[7],但鉴于西太平洋副高是一个高度非线性的复杂动力系统,其形态和活动存在较大的随机性和不确定性,因此ANN方法的预报效果也存在较大局限性,其最大的困难是对副高主要趋势和局部细节的描述难以两全其美。如何既能尽量保留预报对象的重要特征,又使预报过程不太过分复杂?本文采用小波分解与自适应神经模糊推理系统相结合的方法对此进行了研究探讨。

1 西太平洋副高指数的小波分解

1.1 小波分解规则

1988年Mallat^[8]在构造正交小波基时提出多分辨率分析(Multi-Resolution Analysis)概念,从空间的概念上形象地说明了小波的多分辨率特性,给出了正交小波的构造方法以及正交小波变换的快速算法,即Mallat算法。对多分辨率的理解,可从一个3层分解结构说明,其小波分解树如图1。小波分解的最终目的旨在构造一个在频率上高度逼近原始信号的正交小波基,这些频率分辨率不同的正交小波基

收稿日期:2003-04-09;修回日期:2003-08-29。

基金项目:国家自然科学基金项目“西太平洋副高中短期数值预报误差修正研究”(编号:40375019)和“夏季副热带高压变化及其影响天气气候异常的机理”(编号:40135020)资助。

作者简介:张韧(1963-),男,四川峨边人,教授,主要从事热带海洋气象研究。E-mail: zren63@hotmail.com

相当于带宽各异的带通滤波器。图 1 可以看出,小波变换的多分辨率分析主要对信号的低频空间作细致的分解,使其低频部分的分辨水平越来越高,其分解重构关系为:

$$S = d_1 + d_2 + d_3 + a_3$$

若要作进一步分解可继续将低频部分 a_3 分解为低频 a_4 和 高频 d_4 , 以此类推。

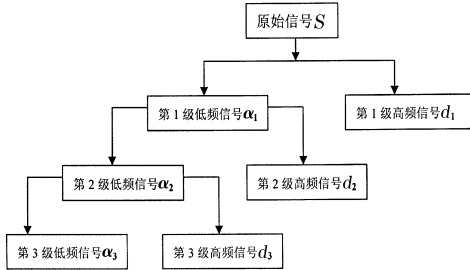


图 1 小波频率分解结构

Fig.1 Structure of decomposed multi-layer wavelet frequency

1.2 副高指数的小波分解

研究表明,副高指数序列具有连续的功率谱和时频变化特性,即具有动力随机性^[9]。同时由于其各频率(周期)分量的相互叠加而表现得混乱无序,尤其是高频扰动既混淆了副高的主要趋势,也使得预报难度加大。因此,准确预报副高的主要困难表现为对副高变化趋势和局部细节的描述难以两全其美,用统计回归和人工神经网络等方法直接建模的预报准确率难以取得实质性的突破^[1-3]。由于小波分解方法能够对复杂信号进行频率(周期)分离,将复杂多周期系统分解为相对简单的准周期信号,从而有效降低预报难度。基于这个思想,我们做以下讨论。

预报对象为中央气象台提供的 1979 年 1 月 ~ 1998 年 12 月共 20 年的 500 hPa 层 110 ~180 E、10 N 以北区域的月平均副高面积指数和副高西伸脊点指数。为表现副高围绕其平均状态的变化和振荡情况,取上述副高指数的距平值进行计算和分析,其中面积指数的负距平表示副高范围和强度弱于其平均水平(最大负值表示副高异常偏弱),正距平表示副高强于其平均水平(最大正值表示副高异常偏强),用副高西脊点值减去 120°,则其正/负距平分别表现副高脊点位于 120° 以东/以西位置,正/负距平的极大值则表现西太副高异常偏东/偏西。

采用具有 5 阶消失矩的 Daubechies 小波基函数

db5 对上述副高指数距平值作 5 层分解和重构。由于是月平均资料,因此数据时间序列的最高采样频率为 1(次/月)。小波分解采用 2 次幂划分,各层小波分解即为带通或低通滤波器,各分解层所占具体的频带如表 1:

表 1 小波频率分解结构

Table 1 Structure of decomposed multi-layer wavelet frequency

| 低频部分 | 频率 | 对应周期 | 高频部分 | 频率 | 对应周期 |
|-------|------------|--------|-------|-----------------|-----------|
| a_1 | 0 ~0.5 | 2 月以上 | d_1 | 0.5 ~1.0 | 1 ~2 个月 |
| a_2 | 0 ~0.25 | 4 月以上 | d_2 | 0.25 ~0.5 | 2 ~4 个月 |
| a_3 | 0 ~0.125 | 8 月以上 | d_3 | 0.125 ~0.25 | 4 ~8 个月 |
| a_4 | 0 ~0.0625 | 16 月以上 | d_4 | 0.0625 ~0.125 | 8 ~16 个月 |
| a_5 | 0 ~0.03125 | 32 月以上 | d_5 | 0.03125 ~0.0625 | 16 ~32 个月 |

原始信号 $S = d_1 + d_2 + d_3 + d_4 + d_5 + a_5$

对副高面积指数距平进行 db5 小波基分解,分别提取其 1 ~5 层的高、低频系数,然后分别对各层高、低频系数进行重构,得到各频带(周期)信号序列(图略)。实际副高指数变化可以通过 1 ~5 层高频和第 5 层低频的重构信号之和精确地获得,但高频段信号(尤其是高频信号 d_1 、 d_2)较为复杂,对其描述和预测有相当难度。滤除了 d_1 、 d_2 高频信号后, d_3 、 d_4 、 d_5 、 a_5 4 种周期信号之和(图 2a)与实际副高面积指数(图 2b)的对比结果(图 2c)表明,分解重构信号 $d = d_3 + d_4 + d_5 + a_5$ 与实际副高强度变化信号 S 之间拟合的很好,不仅趋势变化与实际信号相符,而且相当部分细节也能较好地刻画和描述。这表明副高强度变化可通过 4 ~8 个月(d_3)、8 ~16 个月(d_4)、16 ~32 个月(d_5)及 32 个月以上(a_5) 4 个周期分解信号的叠加来加以描述,滤除 d_1 、 d_2 等高频扰动对副高变化的主要趋势不会产生显著影响。这样,对副高的预测就转化为对其 4 个相对简单的带通、低通分量的预测。由于 d_3 、 d_4 、 d_5 、 a_5 分解信号较实际信号要规则和简单的多,因此可有效地降低预报难度。

同理,我们对副高西脊点指数序列也做类似的周期分解。对副高西脊点距平值进行 db5 小波基 4 层分解后所得 d_2 、 d_3 、 d_4 、 a_4 周期信号之和(滤除了高频 d_1)与实际副高西脊点信号序列两者趋势变化和细节结构上都拟合得很好(图略),表明用 2 ~4 个月(d_2)、4 ~8 个月(d_3)、8 ~16 个月(d_4)及 16 个月以上(a_4) 4 个周期分解信号可以表征副高西脊点变化的主要特征和基本规则,对副高西脊点活动的

预测可通过对这 4 个带通、低通分量的预测集成来加以实现。

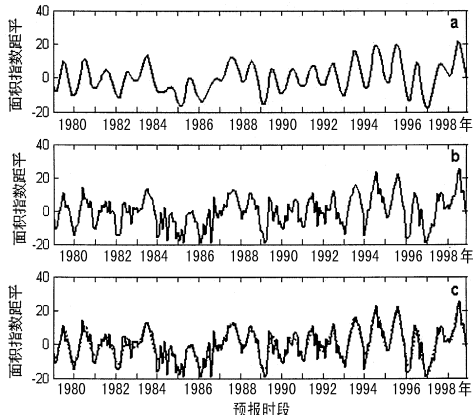


图 2 分解重构信号 d(a)与实际信号 S(b)及拟合情况(c)(采样时间 1979.01~1998.12)

Fig.2 Reconstructed signal d(a) actual signals S(b) and their fitting(c) (Sampling range from January 1979 to December 1998)

2 ANFIS 模型与分解信号的集成预测

2.1 模糊推理与 ANFIS

模糊推理的核心是对复杂系统或过程建立一种语言分析数学模式,将专家经验或实践知识从自然描述语言转化为用模糊规则和模糊集合表示的定量计算机算法和控制系统^[10]。自适应神经模糊推理系统(Adaptive Neuro-Fuzzy Inference System, ANFIS)能够通过训练和自适应学习,完成传统模糊系统中靠经验调整隶属函数以减小误差、提高逼近效率的工作。它以混合式学习过程为基础,分别运用最小平方方法和后传梯度下降法来完成推理结果与训练目标间的误差计算,以确定出模型最小指定误差所对应的线性和非线性参数,并逐渐调配出相应的隶属函数来满足所需模糊推理系统输入输出关系,从而建立起一套“if...then...”规则的模糊推理系统^[10,11]。该模糊推理系统对应规则为:

规则 1 :if x is A_1 and y is B_1 ,
then $h_1 = p_1x + q_1y + r_1$;

规则 2 :if x is A_2 and y is B_2 ,
then $h_2 = p_2x + q_2y + r_2$;

A_1, B_1 为隶属函数映射值, x, y 为模糊推理的假设和训练输入, p_1, q_1, r_1 为模糊推理结论参数, $i = 1, 2$ 其推理过程如图 3 所示。

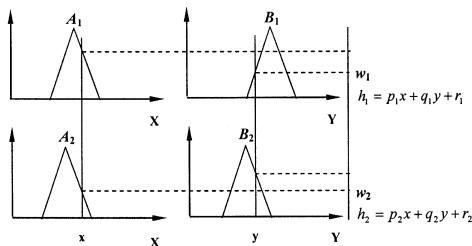


图 3 ANFIS 模糊推理过程

Fig.3 Fuzzy inference process of ANFIS model

采用加权平均法非模糊化,则模糊推理输出为

$$h = \frac{w_1}{w_1 + w_2} h_1 + \frac{w_2}{w_1 + w_2} h_2 = \text{模糊}h_1 + \text{模糊}h_2, w_i \text{ 为第 } i \text{ 个}$$

结点的输出权重(取隶属函数的极大映射值)模糊系统可表现为 Takagi-Sugeno 形式^[11]。 $h = \text{模糊}h_1 + \text{模糊}h_2 = (\text{模糊}x)p_1 + (\text{模糊}y)q_1 + (\text{模糊})r_1 + (\text{模糊}x)p_2 + (\text{模糊}y)q_2 + (\text{模糊})r_2$ 在混合式学习过程中,结论参数 p_i, q_i, r_i 由最小二乘估计法求得,通过反向学习最大梯度法,返回误差变换并更新前提输入量 x, y 。在改变这些参数的过程中,各种对应 A_i, B_i 的适当的隶属函数也相应出现。由于前提与结论参数已经解耦(decoupled),且 ANFIS 又是放射性网络,所以它的学习效率比神经网络高^[11,12]。

2.2 副高面积指数的分量集成预测

基于副高面积指数的周期分解结果,采用自适应模糊推理系统 ANFIS 分别建立 4~8 个月(d_3)、8~16 个月(d_4)、16~32 个月(d_5)及 32 个月以上(a_5) 4 个周期分解信号的外推预报模型,即用前 3、4、5 个月的副高面积指数距平值预测其后的副高指数形态,实际预报时效为 3 个月。设 P, T 分别为 ANFIS 模型的预报因子输入和预报结果输出序列,即:

$$P = [x(t-5) \ x(t-4) \ x(t-3)] \quad T = [x(t)]$$

则每一时次的训练数据对为:

$$[x(t-5) \ x(t-4) \ x(t-3) \ x(t)]$$

为便于检验所建 ANFIS 预报模型的独立预报效果,将 240 个月的副高面积指数分为两部分:第一部分取 1979 年 1 月~1995 年 8 月时段共 200 个月,用于预报模型建立和模型拟合试验;第二部分取 1995 年 9 月~1998 年 12 月共 40 个月,用于对所建模型做独立的预报效果检测。

模型计算结果表明,副高面积指数的小波分解信号 d_3, d_4, d_5, a_5 与所建 ANFIS 预报模型的相应分量计算值之间达到相当精确的拟合效果,各分量信

号几乎重合。ANFIS 模型的分量集成计算结果与实际小波分解重构信号 $d = d_3 + d_4 + d_5 + a_5$ 和实际副高面积指数 s 之间均达到较好的拟合效果 相关系数分别达 0.9795 和 0.8716 (置信度 $=0.05$)。

ANFIS 模型独立的集成预报效果如图 4, 其中小波分解重构信号 d (图 4a 实线) 和实际信号 s (图 4b 实线) 与 ANFIS 模型集成预报结果 (图 4 点线) 相比较 两者的目标趋势特征基本吻合, 其中模型预报值与 d 的相关系数达 0.9395 (置信度 $=0.05$); 由于模型中滤除了 d_1 (1 ~ 2 个月周期) 和 d_2 (2 ~ 4 个月周期) 等高频信号, 因此预报结果对实际信号细节的描述略有欠缺, 但其主要变化特征和升降趋势把握基本正确, 模型预报值与实际值 s 的相关系数为 0.8761 (置信度 $=0.05$)。

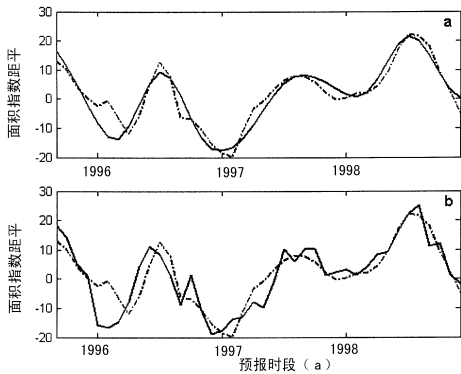


图 4 分解重构信号(a 实线)/实际信号(b 实线)与模型独立预报值(点线)(预报时段 1995.09 ~1998.12)

Fig.4 Reconstructed signal (solid line of a) / observed signal (solid line of b) and independent prediction by the ANFIS model (dotted line) (The prediction ranges from September 1995 to December 1998)

为进一步提高预报稳定性和准确率 根据实际可用资料, 我们增加了训练数据的样本序列, 取 1954 年 1 月 ~1987 年 4 月共 400 个月的副高面积指数距平值作为建模数据; 取 1987 年 5 月 ~1999 年 10 月共 150 个月的副高面积指数距平用作独立预报效果检验。按上述相同方法和过程所得副高面积指数的独立预报结果见图 5。模型独立预报值与分解重构信号 d 的相关系数达 0.9767, 与实际信号 s 的相关系数达 0.9029 (置信度均为 $=0.05$)。上述结果表明 随着建模样本的增加 预报结果对实际信号细节的描述和总体趋势的把握可得到较明显的改进和提高。

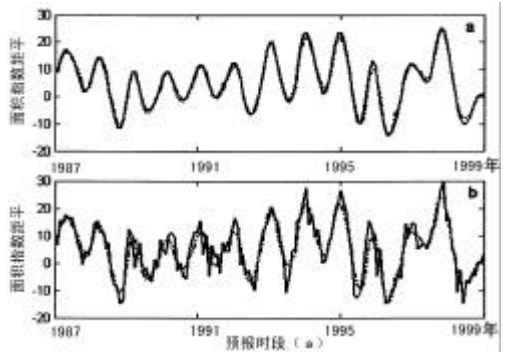


图 5 分解重构信号(a 实线)/实际信号(b 实线)与独立预报值(点线)(预报时段 1987.05 ~1999.10)

Fig.5 Reconstructed signal (solid line of a) / observed signal (solid line of b) and independent prediction by the ANFIS model (dotted line) (The prediction ranges from May 1987 to October 1999)

2.3 副高西脊点的分量集成预测

采用与副高面积指数相同的建模方法和过程以及资料范围(亦取 1979 年 1 月 ~1995 年 8 月共 200 个月用于预报建模和模型拟合 1995 年 9 月 ~1998 年 12 月共 40 个月用于独立预报检测), 用 ANFIS 模型分别建立 2 ~ 4 个月(d_2), 4 ~ 8 个月(d_3), 8 ~ 16 个月(d_4) 及 16 个月以上(a_4) 周期 4 个分解信号的外推预报模型 用前 3、4、5 个月的副高西脊点指数预测其后的副高西伸状况 实际预报时效为 3 个月。

试验结果表明, ANFIS 模型的分量集成结果与小波分解重构信号 $d = d_2 + d_3 + d_4 + a_4$ 和实际副高西脊点信号 s 均达到较理想的拟合精度 拟合相关系数分别为 0.9642 和 0.8522 (置信度 $=0.05$, 图略)。图 6 是该模型的独立集成预报结果(虚线)与小波分解重构信号(a 实线)和实际副高西脊点(b 实线)的对比, 预报结果除部分波谷区表现欠佳外, 其余基本能够正确描述和反映实际副高脊点的活动趋势和变化特征 相关系数分别达 0.9266 和 0.9149 (置信度 $=0.05$)。

3 小结

本文引入小波分解和 ANFIS 模型相结合的集成预报方法用于西太平洋副高面积指数和西伸脊点指数月、季时间尺度的预测研究。该方法在保留预报对象主要特征的同时, 有效降低了问题的复杂性, 因而显著提高了预报准确率。由于小波方法能够准

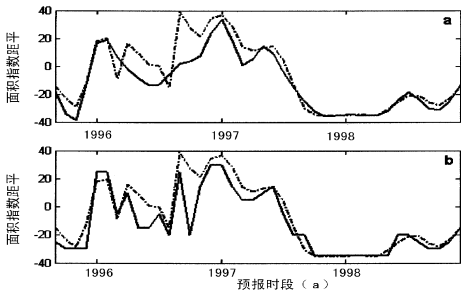


图 6 分解重构信号(a 实线)/实际信号(b 实线)与独立预报值(虚线)(预报时段 1995.09~1998.12)

Fig.6 Reconstructed signal (solid line of a) /observed signal (solid line of b) and independent prediction by the ANFIS model (dotted line) (The prediction ranges from September 1995 to December 1998)

确分解和重构带通、低通信号,其频域—时域分辨率能够自由伸缩,因此利于简化系统和提取特征,而模糊系统及 ANFIS 模型具有高度的非线性、容错性、自适应性和联想学习功能,因此小波分解重构和 ANFIS 模型优势互补的集成预测方法在预测西太平洋副高等复杂现象时往往表现出较明显的优越性。

参考文献 (References):

[1] Feng Minxue (冯民学), Chen Biyun (陈必云), Yuan Ruiqin (袁瑞琴), et al. West Pacific subtropical high forecast operation system based on artificial neural network [J]. Scientia Meteorologica

Sinica (气象科学), 1998, 18(4): 396-401 (in Chinese).

- [2] Zhang Ren (张韧). Diagnosis and prediction research on the Pacific subtropical high based on the forward-propagation network approximation [J]. Acta Scientia Atmospherica Sinica (大气科学), 2001, 25(5): 650-660 (in Chinese).
- [3] Zhang Ren (张韧), Jiang Guorong (蒋国荣), Yu Zhihao (余志豪), et al. Establish prediction model of the Pacific subtropical high by ANN method [J]. Quarterly Journal of Applied Meteorology (应用气象学报), 2000, 11(4): 474-482 (in Chinese).
- [4] Zhang Ren, Yu Zhihao. Neural network BP model approximation and prediction of complicated weather system [J]. Acta Meteorologica Sinica, 2001, 15(1): 105-115.
- [5] Jin Long (金龙), Chen Ning (陈宁), Lin Zhenshan (林振山), et al. Research and comparison of ANN-based consensus forecasting [J]. Acta Meteorologica Sinica (气象学报), 1999, 57(2): 198-207 (in Chinese).
- [6] Tang Xunchang (唐洵昌), Ge Wenzhong (葛文忠), Wang Xiangdong (王向东), et al. ANN techniques for weather radar identified severe convective weather [J]. Scientia Meteorologica Sinica (气象科学), 1997, 17(4): 393-399 (in Chinese).
- [7] Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks [J]. Neural Networks, 1990, 3(4): 551-560.
- [8] Mallat S. A theory for multiresolution signal decomposition: The wavelet representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674-693.
- [9] Li Chongyin (李崇银). Introduction of Climate Dynamics [M]. Beijing: China Meteorology Press, 1995 (in Chinese).
- [10] Li Shiyong (李士勇). Fuzzy Control, Neurocontrol and Intelligent Cybernetics [M]. Harbin: Harbin Industry University Press, 1998 (in Chinese).
- [11] Roger Jang J S, Sun C T, Mizutani E. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence [M]. Englewood Cliffs: Prentice-Hall, Inc, 1997.
- [12] Takagi T, Sugeno M. Fuzzy identification and its application to modeling and control [J]. IEEE SMC, 1985, 15(1): 116-132.

WAVELET DECOMPOSITION AND COMPOSITIVE PREDICTION ON THE MODALITY INDEX OF THE WEST-PACIFIC SUBTROPICAL HIGH

ZHANG Ren^{1,2}, DONG Zhao-jun¹, CHEN Yi-de¹, HE Jin-hai²

(1. Institute of Meteorology PLA University of Sciences and Technology Nanjing 211101, China; 2. Nanjing Institute of Meteorology, Nanjing 210044, China)

Abstract: Based on the method of associating wavelet decomposition with adaptive neuro-fuzzy inference system (ANFIS), a composite prediction model on the modality index of the west-Pacific subtropical high (WPSH) on the monthly-seasonal scale was established. Signals can be freely extended/shrunked in frequency time domain and any pass-band and pass-low frequency branch can be accurately produced and reconstructed by means of wavelet decomposition. Therefore, the complex WPSH modality index time series signals can be separated into several relative simple band-pass signals, which both simplify the system structure and stand out the chief characters of signals. Subsequently, the independent prediction model of the decomposed signals were established based on the advantages of ANFIS model, such as non-linear, bearing-error, self-adapting and association-learning and the independent predicted results were integrated finally. The test results showed that under the premise of keeping the main characters of forecast body, the prediction difficulty on WPSH system had effectively been decreased, the precision and durative of the composite prediction model were evidently improved and promoted compared with that of traditional prediction technique.

Key words: WPSH index; Wavelet decomposition; Fuzzy inference; Composite prediction.