

基于小波分解重构技术的西太平洋热带 气旋生成频数预报模型

王 举^{1,2}, 蒋国荣^{1,2}, 李建东², 张 韧²

(1. 南京气象学院, 江苏 南京 210044, 2. 解放军理工大学气象学院, 江苏 南京 211101)

摘 要: 利用小波的多尺度分辨特性和小波分解技术, 将预测对象按不同时间尺度分解。通过去掉无法预测的高频变化, 并依据时间尺度大小和不同的预报方法分别建立分解预报模型, 再利用小波重构技术进行合成预报, 使原本难以预测的问题变得较易预测。研究结果得出, 利用此方法一方面可以用于预测复杂的气候变化, 另一方面也可以同时提高预测效果。

关键词: 统计预报; 小波分解与重构; 神经网络; 热带气旋

中图分类号: P725 文献标识码: A 文章编号: 1003 - 0239 (2004)01 - 0035 - 08

1 引言

长期天气和短期气候预测理论和方法一直是人们关注的问题。由于影响气候变化的机制和因素十分复杂, 目前长期数值预报仍处于研究发展之中, 业务预报还主要是依靠统计预报方法^[1]。但大部分业务应用的统计预报方法, 本质上是线性的顺序处理技术。具有非线性和信息并行处理特征的人工神经网络是近年来发展较快的应用研究学科之一。特别是 20 世纪 80 年代以来, 许多重要问题的解决和快速有效算法的出现, 使得它在各学科中的广泛应用成为可能。金龙^[2~3]、蒋国荣^[10]、张韧^[4]等运用神经网络进行了一些长期预报研究, 获得了较好的效果。其次, 众所周知, 气候的变化是多尺度的, 不同时间尺度上的气候变化表现形态并不相同, 如何描述这种不同时间尺度的差异呢? 传统的方法是采用滑动平均或高斯滤波等方法, 但这些传统方法往往会造成一些信息的丢失。近几年发展起来的小波分析方法可能是一个更好的方法, 小波分析在时域和频域同时具有良好的局部性质, 它能将信号(气象要素时变序列)分解成交织在一起的多尺度成分, 并对各种不同尺度成分采用相应粗细的时域或空域取同样步长, 从而能够不断地聚集到所研究对象的任意微小细节^[5]。

热带气旋是生成于热带海洋上的剧烈天气系统, 而发生在西北太平洋的热带气旋对我国天气和气候的影响较大, 研究和预测西北太平洋热带气旋生成对于我国有效采取防灾减灾措施具有重要的意义。20 世纪 80 年代开始, 国内外许多学者^[7~9]研究了海洋热

收稿日期: 2003-06-23; 修改日期: 2003-12-10

作者简介: 王 举 (1974-), 男, 南京气象学院博士, 从事海洋气象研究。

基金项目: NSFC 重点基金(40136010)项目赞助。

力状况对西北太平洋热带气旋生成频数的影响。研究发现，赤道东太平洋海温与西太平洋热带气旋生成频数关联最为密切，并主要表现在年际变化尺度上。在年际变化尺度上，相关系数可达 0.6 以上，已可用于建立统计预测模型。

本文拟采用小波分析中的分解重构技术^[6]与神经网络、线性回归方法相结合的方法建立预测西太平洋热带气旋生成频数月变化的模型，为复杂而难以预测的短期气候现象的预测探索新思路和新方法。

2 资料和方法

2.1 资料

本文所使用的西太平洋热带气旋生成频数是根据上海台风研究所整编的《台风年鉴》统计得出的，所用的热带气旋包括了所有的热带低压、热带风暴、台风(用新规定的名称)，资料时间为 1950 年 1 月-1997 年 12 月共 48 年 576 个月。所使用的海温取自南京气象学院大气资料服务中心提供的 1856 年 1 月-1995 年 10 月的全球海温距平场资料。本文主要选择了 Nino3 区(5°N~5°S, 90°~150°W)的平均海温，因为根据以往的研究成果^[9]，Nino3 区的海温与西太平洋热带气旋生成频数关联最好。

2.2 小波分解与重构技术简介

1988 年，S. Mallat 在构造正交小波基时提出了多分辨分析的概念^[6]，从空间的概念上形象地说明了小波的多分辨率特性，将此之前的所有正交小波基的构造法统一起来，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。对于多分辨分析的理解，可以用一个三层的分解进行说明，设原始序列为 S，其小波分解如图 1 所示：

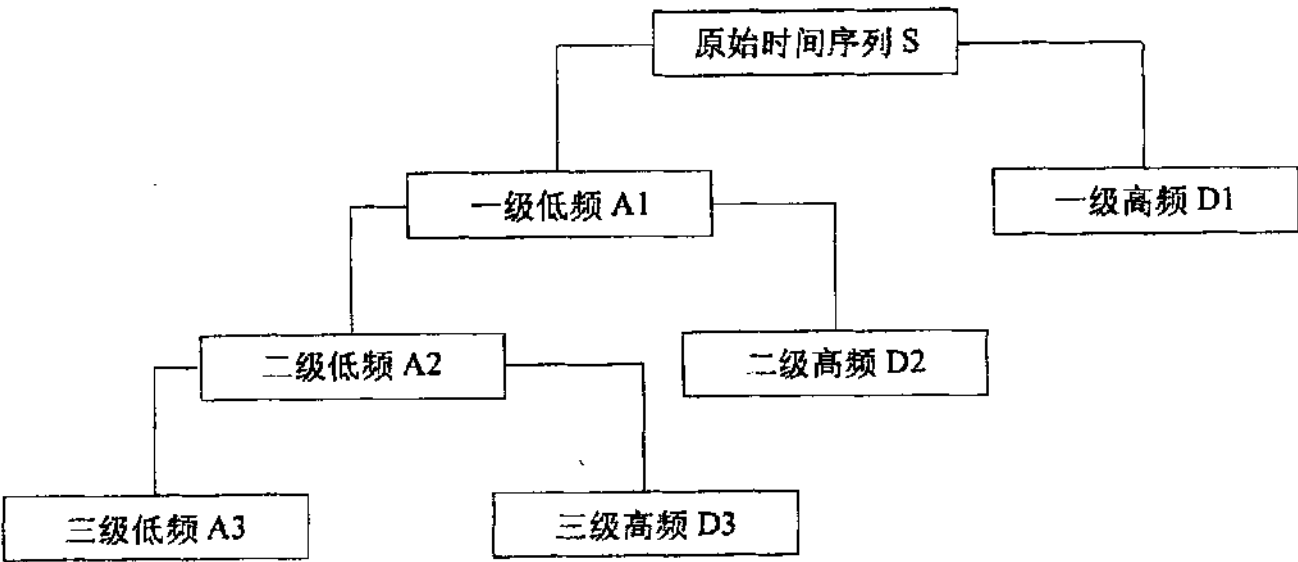


图 1 三层小波分解结构示意图

从图 1 可以明显看出，多分辨分析只是对低频部分进行进一步分解，而高频部份则不予以考虑。分解具有以下关系： $S=A3+D3+D2+D1$ 。若要进一步分解，可将低频部分 A3 分解成低频部分 A4 和高频部分 D4。再分解则依此类推。

另外, 本文还使用了人工神经网络中 BP 网络的算法, 有关 BP 网络的算法和规则, 许多文献都做了详细介绍, 可参考相应的文献。

3 西太平洋热带气旋生成频数月变化的预测模型

3.1 Nino3 区海温与西太平洋热带气旋生成频数变化的多尺度分析

由于海温资料和热带气旋生成频数资料的时间序列长度并不一致, 为方便比较以及后面预测模型的建立, 本文仅选择了二个资料序列共有的时间段: 1950 年 1 月-1995 年 10 月共 550 个月的资料。在进行小波分解和重构之前, 本文采用了对称延伸法^[5]消除边界效应。由于热带气旋生成的季节变化较为明显, 为消除季节变化的影响, 我们采用了正规化方法对热带气旋的原始资料序列进行了预处理: 即, 用原始资料与季节平均值的差再除以各月平均方差。所使用的海温值本身就是去除了季节变化的距平值。利用具有 5 阶消失矩的 Daubechies 小波函数 db5^[6]对上述进行了预处理的海温与热带气旋生成频数时间序列做 6 层分解与重构。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 幂次方式划分, 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(周期段)为: d1: 0.5~0.1(1~2 个月); d2: 0.25~0.5(2~4 个月); d3: 0.125~0.25(4~8 个月); d4: 0.0625~0.125(8~16 个月); d5: 0.03125~0.0625(16~32 个月); d6: 0.015625~0.03125(32~64 个月); a6: 0.0~0.015625(64 个月以上); 原始序列 $S=a6+d6+d5+d4+d3+d2+d1$ 。图 2(abcdefg)为分解后各个频带(周期段)的信号序列。

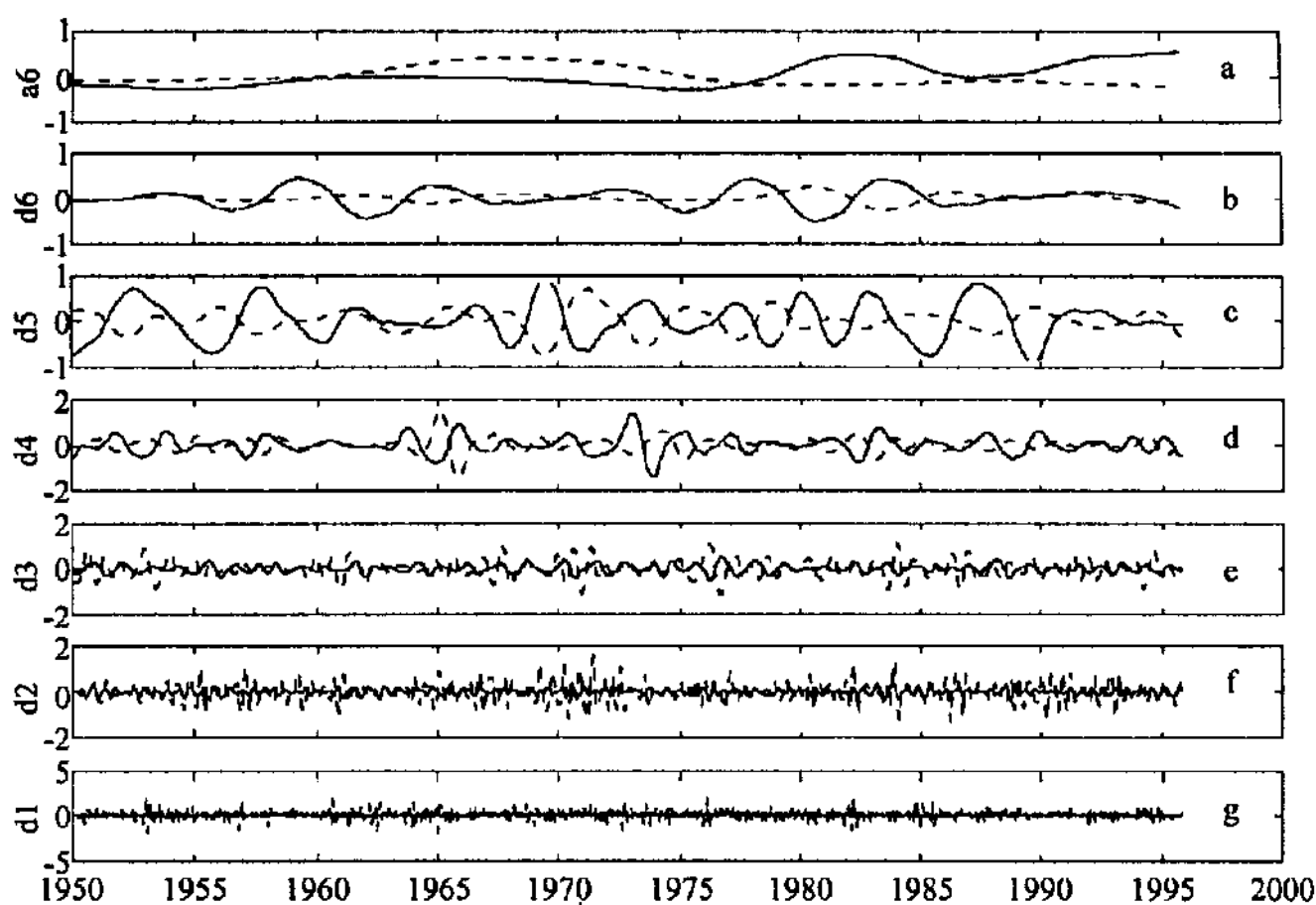


图 2 西太平洋热带气旋生成频数与 Nino3 区海温小波分解图

(实线代表热带气旋, 虚线代表海温)

由图 2a 可以发现,从长期趋势来看(5 年以上尺度),Nino3 区海温的变化趋势与西太平洋热带气旋生成频数的变化趋势相关并不太好(相关系数为-0.456),表明西太平洋热带气旋生成频数的长期变化不能单独由 Nino3 区海温的长期变化确定。由图 2b、2c、2d 则可以看出,在 8 个月至 64 个月(1 年~5 年尺度上)周期段上,Nino3 区海温的变化与西太平洋热带气旋生成频数的变化有相反变化趋势(相关系数分别为-0.490,-0.715,-0.622),其中,在 16~32 个月尺度上(准二年周期)相关最好;由图 2e、2f、2g,在较小尺度上,Nino3 区海温的变化趋势与西太平洋热带气旋生成频数的变化趋势关联最差(相关系数分别是-0.123,0.042,-0.029),反映出西太平洋热带气旋生成频数的短时间尺度变化与 Nino3 区海温相同尺度变化无关;这与以往的研究结果一致,只是这里的分析结果在时间尺度上的划分更为精细,这也充分反映了小波分析良好的多尺度分辨特性。

3.2 热带气旋生成频数月变化预测模型的建立

影响热带气旋生成的因素很多,造成了热带气旋生成频数月变化的复杂性和不规则性,我们曾试图利用人工神经网络良好的函数逼近能力和非线性系统特性建立热带气旋生成频数原始序列月变化预测模型,预测因子选择热带气旋生成频数的前期量。结果发现,人工神经网络虽然保证了预测模型非常好的拟合效果(可达 0.9 以上),但预报效果极差,几乎没有预报能力。究其原因,我们认为,就在于过份追究了热带气旋生成频数月变化的具体细节,这些具体细节可能含有太多的难以预测的短周期变化,这从上面小波分解中图 2g 中 d1 分量的变化就可看出。因此,为了确保预测的可行,我们必须忽略一些具体的细节而主要保障对大趋势的拟合和预测。要做到这一点,只需去掉小尺度变化部分,利用上面的小波分解重构技术,重构序列中去掉了 1~2 个月尺度变化的 d1 部

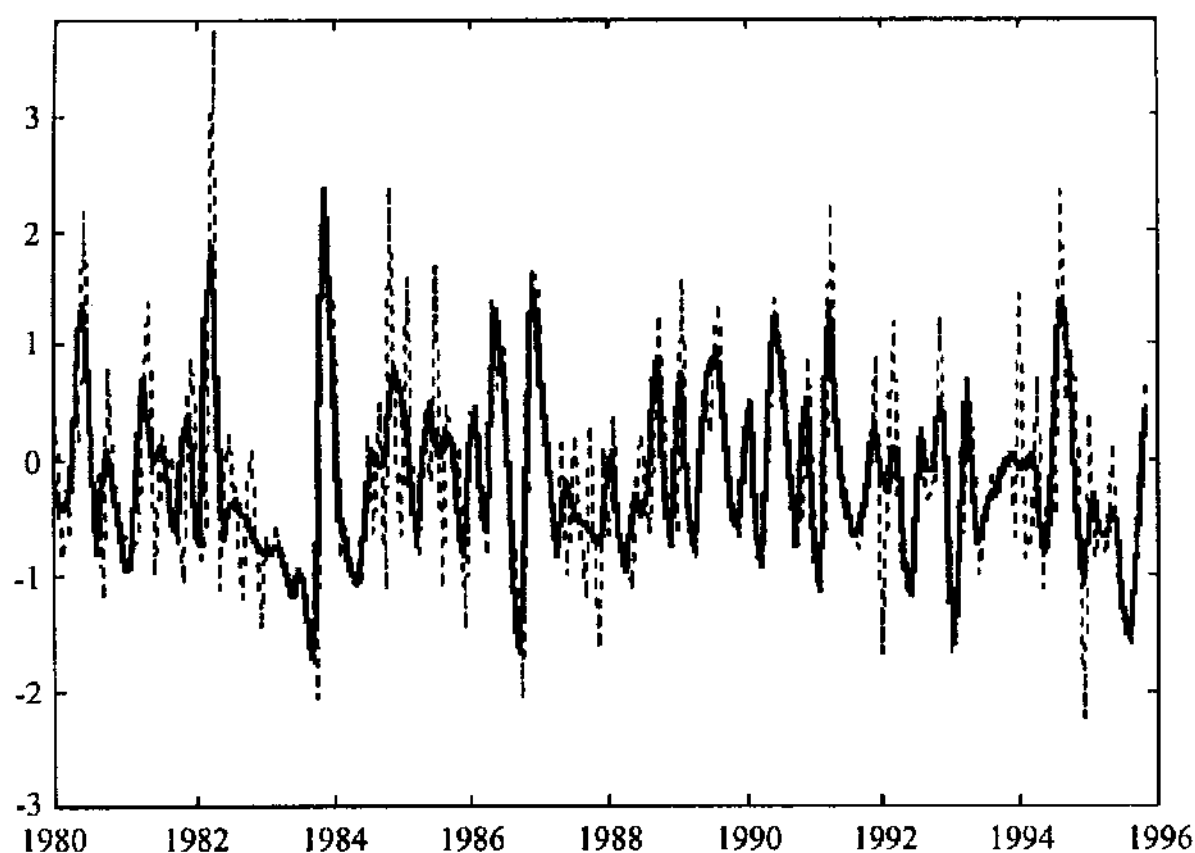


图 3 部分时段原始序列 S 与预测序列 S1 的对比曲线

(粗实线为 S1 序列,虚线为 S 序列)

分, 也就是我们的预测对象是 $S1 = a6 + d6 + d5 + d4 + d3 + d2$, 而不是原始序列 S 。图 3 给出了原始序列 S 和预测序列 $S1$ 的对比曲线 (为使二者的对比更为清晰, 图中仅选择了部分时间段的情况)。由图可以看出, 序列 $S1$ 基本上再现了序列 S 的主要变化趋势, 但在一些细小环节存在差异, 二者的相关系数值为 0.749。

根据上面海温与热带气旋生成频数在不同尺度上的对比分析结果, 如果我们把 $S1$ 作为一个整体进行预测, 预报因子中选择海温反而可能会降低预报效果, 这是因为海温的变化在长期趋势和短尺度变化与热带气旋频数的相同尺度的变化相关不是太好。由于神经网络的拟合会再现预报因子的行为, 反而可能降低预报效果。因此, 为进一步提高预报效果, 我们把预测对象 $S1$ 依据不同尺度再分成三个不同的预报序列: $Y1 = a6$; $Y2 = d6 + d5 + d4$; $Y3 = d3 + d2$ 。这里, $S1 = Y1 + Y2 + Y3$ 。对不同尺度的预报序列采用不同的预报因子和预报方法, 最后将三个预报序列进行合成。

为便于预报效果比较, 将数据资料分为独立的两部分, 第一部分用于模型的建立和拟合检验, 资料时段为 1950.1~1983.7 共 403 个月; 第二部分用于预报模型预报效果的评估, 资料时段为 1983.8~1995.10 共 147 个月;

4 预报效果和拟合效果

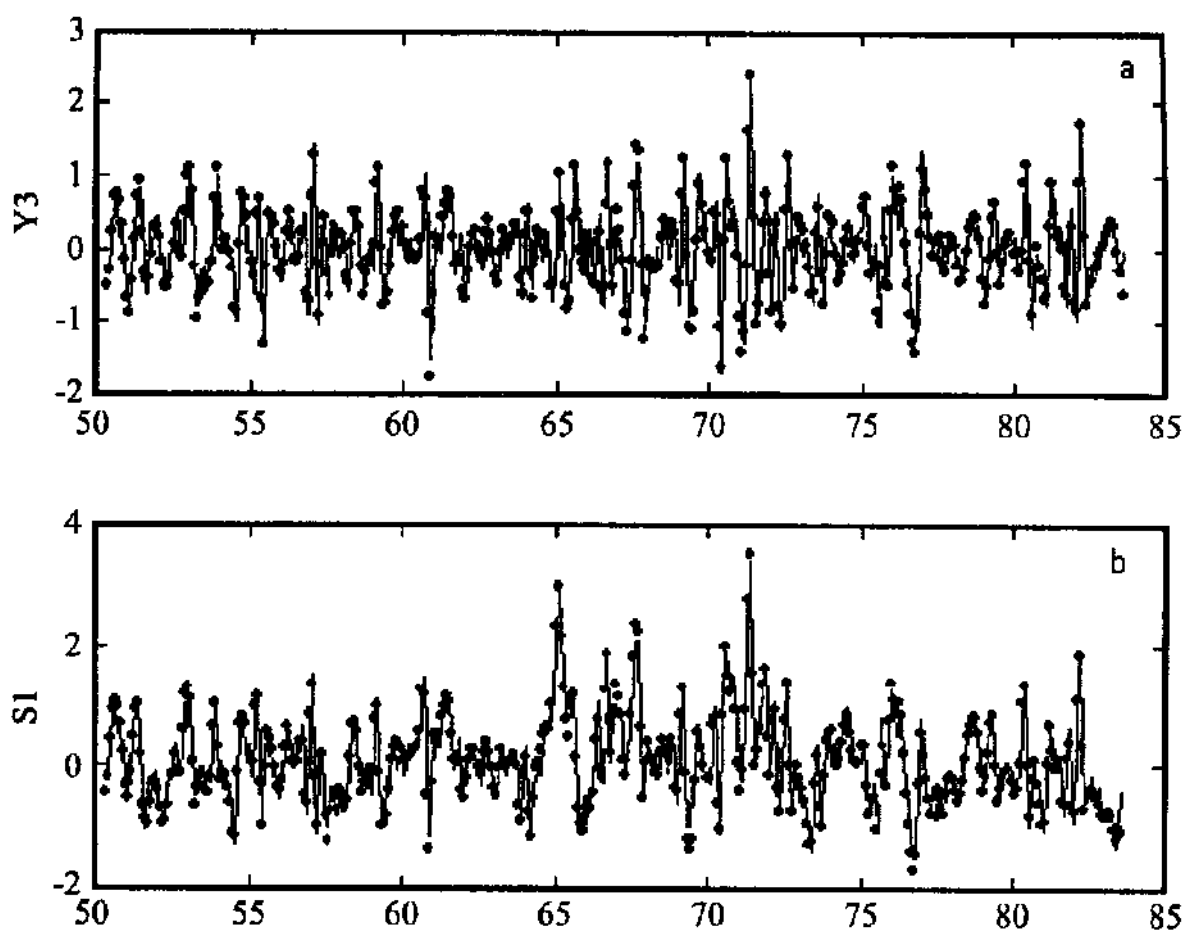


图 4 西太平洋热带气旋生成频数分解及合成拟合结果曲线

(黑点为实际值, 实线为拟合值、a. $Y3$ 序列、b. 合成序列 $S1$)

基于上述预测模型的设计思想, 对 $Y1$ 预报序列, 预报因子我们主要选择自身的前期因子, 预报时效为 1 个月。预报方法采用多元线性回归方法。值得一提的是, 我们这里没有采用 BP 神经网络方法, 主要是通过试验发现, 对于变化较为平缓的时变序列,

线性回归方法预报效果优于 BP 神经网络方法的结果。对 Y2 预报序列, 预报因子我们除了选择 Y2 自身的前期因子外, 还选择了前期海温因子, 因为在 1~5 年尺度上海温的变化与热带气旋生成频数的变化关联较好, 在预报方法上仍采用多元线性回归方法。对 Y3 预报序列, 预报因子我们仍选择 Y3 自身的前期因子, 但预报方法采用 BP 神经网络方法。预报试验研究发现, 对于变化较为剧烈的 Y3 序列, BP 神经网络方法显示出了极大的优越性, 其预报效果和拟合效果均远远高于多元线性回归方法(可提高 10%)。

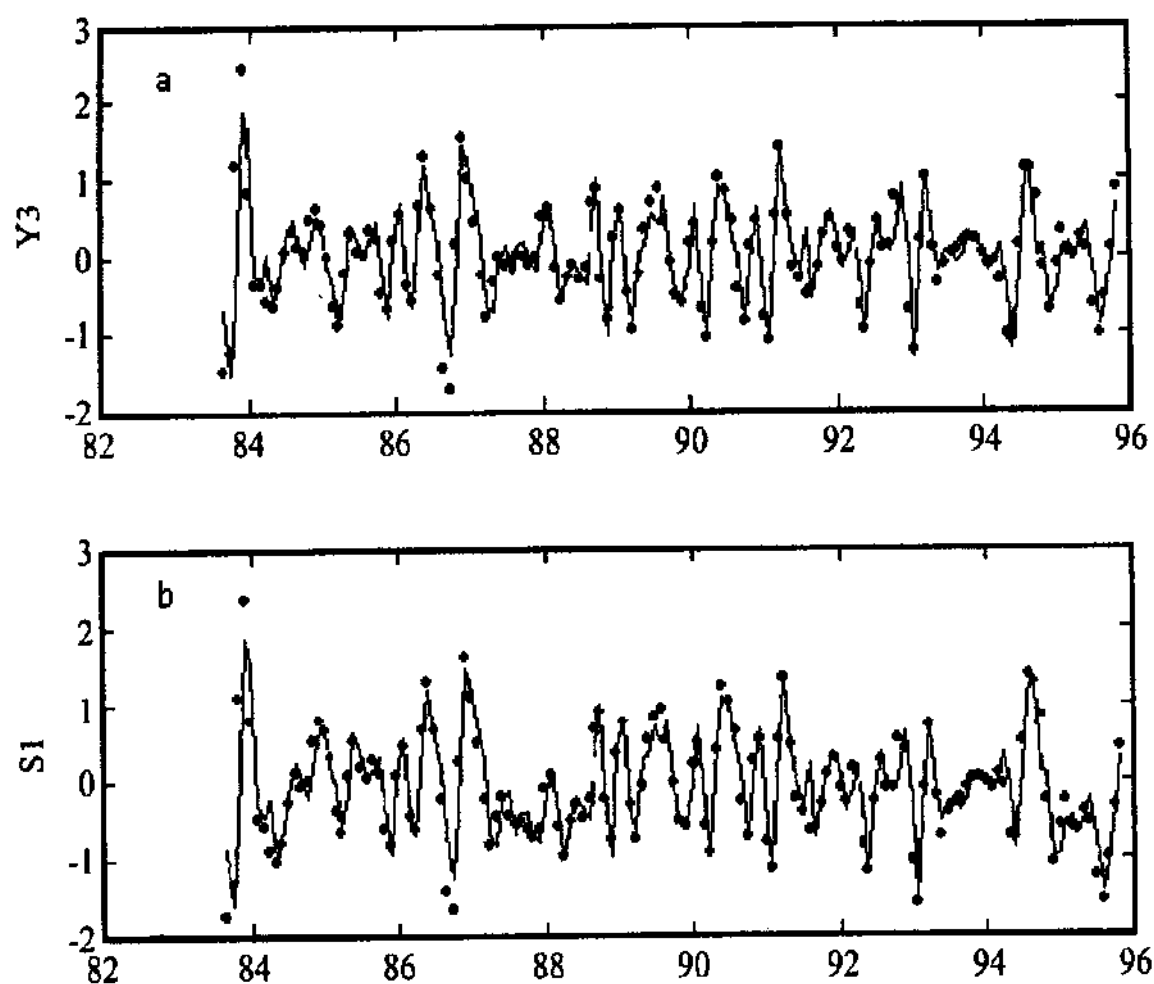


图 5 西太平洋热带气旋生成频数分解及合成预报结果曲线
(黑点为实际值, 实线为拟合值、a. Y3 序列。b. 合成序列 S1)

图 4a、b 分别为 Y3 和合成序列 S1 的拟合效果图示。图中, 实线为实际值、虚线为拟合值。对尺度较大的 Y1 和 Y2 预报序列, 多元线性回归的拟合效果极佳(相关系数分别为 1.000 和 0.997), 基本与原序列重合(图略); 而由图 4a 则可看出, 对 Y3 预报序列, BP 神经网络方法拟合率很高, 相关系数达 0.924(多元线性回归方法仅为 0.812); 图 4b 是对 S1 整体的合成拟合结果, 可以看出, 拟合效果非常好, 拟合相关系数为 0.955。

图 5a、b 则分别为 Y3 和合成序列 S1 的预报结果图示, 图中, 黑点为实际值、实线为预报值。同样对 Y1 和 Y2 预报序列, 预报效果也非常好(预报相关系数分别为 1.000 和 0.996); 即使是 Y3 预报序列, 预报结果也很好, 预报相关系数达到 0.885(多元线性回归方法为 0.809); 图 5b 是对 S1 整体的合成预报结果, 预报效果非常好, 合成预报相关系数达 0.905。

由以上结果可以发现, 对去掉短时间尺度变化后的 S1 时间序列的拟合和预报效果都非常好, 相关系数达到 0.9 以上。造成如此高的拟合和预报效果的主要原因是因为我

们的拟合或预报对象 - 合成序列 S1 完全去年去掉了难以预测的高频变化部份 b1，从而突出了较有规律的信号所致。但是 S1 时间序列与原始时间序列 S 是有差别的，图 6a、b 分别给出了合成拟合和预报效果与原始时间序列 S 的对比曲线，可以看出，合成预报报出了原始时间序列的主要变化趋势，与原始时间序列 S 的拟合与预报相关系数分别是 0.719 和 0.690。由于预测序列 S1 与原始时间序列 S 的相关系数仅为 0.749，分解合成预报效果可以说是比较令人满意的。

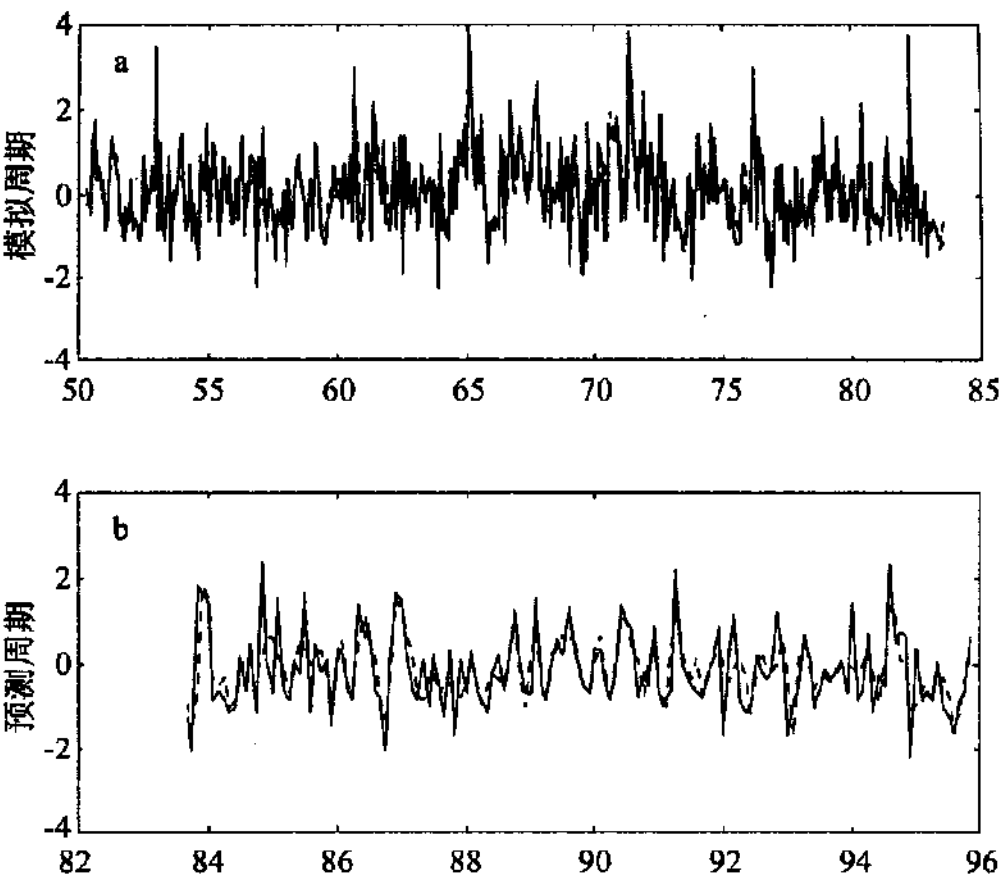


图 6 西太平洋热带气旋生成频数分解合成拟合和预报效果
(a. 拟合结果、b. 预报结果、实线为观测值，虚线为预报值)

5 结语

利用小波分解重构技术，通过去掉原始预报序列中变化复杂、难以预测的高频变化部分，使一个原本无法预报的序列转化成一个容易预报的序列，并且报出了所关心的时变序列的大致变化趋势，这就为短期气候预测提供了一种新的思路和方法。其次，利用小波的多尺度分辨特性，可以根据不同尺度上影响因素的差异选择不同的预报因子建立不同尺度的分解预报模型，最后进行合成。这种方法在物理解释上是合理的，也使预报效果有所提高。它比单纯整体预报的预报能力(0.631)提高了大约 8% 以上，显示了应用小波分析技术与短期气候预测的优势。当然，本文所建立的西太平洋热带气旋生成频数的模型可能过于粗糙，特别是对快变部分的预报因子仅选择了自身的前期因子，是否通过寻找到影响热带气旋快变的外界因子从而可进一步提高预报效果是一个值得研究的问题。另外，从本文的研究还可看出，由于去掉了短时间尺度变化后的预测序列 S1 与原始序列 S 的相关不太高，这就限制了预报能力的进一步提高。是否存在一种新的途径，一方面使预报成为可行，另一方面又能提高预报能力仍是值得进一步研究的。

参考文献：

- [1] 叶笃正, 曾庆存, 郭裕福. 当代气候研究. 北京: 气象出版社, 1991, 164~177.
- [2] 金 龙, 陈 宁, 林振山. 基于人工神经网络的集成预报方法研究和比较. 气象学报, 1999, 57(2): 198~207.
- [3] 金 龙, 秦伟良, 姚华栋. 多步预测的小波神经网络预报模型. 大气科学, 2000, 24(1): 79~86.
- [4] 张 韧, 蒋国荣, 余志豪, 蒋全荣. 利用神经网络计算方法建立太平洋副高活动的预报模型. 应用气象学报, 2000, 11(4): 474~483.
- [5] 林振山, 邓自旺. 子波气候诊断技术的研究. 北京: 气象出版社, 1999.
- [6] 胡昌华, 等. 基于 Matlab 的系统分析与设计-小波分析. 西安: 西安电子科技大学出版社, 2000, 1~20.
- [7] 潘怡航. 赤道东太平洋热力状况对西太平洋台风发生频率的影响. 气象学报, 1982, 40(1): 24~33.
- [8] 董克勤, 齐树芬. 赤道东太平洋海温与西太平洋台风频数年际变化的关系. 海洋学报, 1990, 12(4): 505~510.
- [9] 蒋国荣, 沙文钰. 用赤道东太平洋海温预报西太平洋热带气旋年际变化的探讨. 热带气象学报, 1993, 9(3): 221~228.
- [10] 蒋国荣, 张 韧, 沙文钰. 用 EOF 展开和人工神经网络方法预测 ENSO 的研究. 海洋预报, 2001, 18(3): 1~11.

THE PREDICT MODEL OF TROPICAL CYCLONE OCCURRENCE FREQUENCY NUMBER IN WEST PACIFIC OCEAN BASING ON THE RESOLVING AND RECONSTRUCTING TECHNIQUE OF WAVELET

Wang ju^{1,2}, Jiang Guorong^{1,2}, Li Jiandong², Zhang Ren²

(1. Nanjing Institute of Meteorology, Nanjing 210044, China ;

2. Institute of Meteorology, PLA University of Science and Technology, Nanjing 211101, China)

Abstract : In this paper, using the multi-scale resolution feature and resolving technique of wavelet, at first, we can change the unpredictable problem into predictable problem by discarding high frequency change ; Second, we can construct resolving predict model according to the difference of the time-scale and chosing the difference predict method; Finally, we make synthetical prediction by reconstructing technique of wavelet. It is found that we can use this method both to predict complex climate change and to advance predict effects.

Key word : statictical prediction ; resolving and reconstructing of wavelet ; neural network ; tropical cyclone