

张舰齐,左瑞亭,王丽琼. 基于集合经验模态分解和傅里叶分析的风场预测试验[J]. 干旱气象, 2015, 33(5): 769-776, [ZHANG Jianqi, ZUO Ruiting, WANG Liqiong. Forecasting Experiments of Wind Based on Ensemble Empirical Mode Decomposition and Fourier Analysis[J]. Journal of Arid Meteorology, 2015, 33(5): 769-776], doi:10.11755/j.issn.1006-7639(2015)-05-0769

基于集合经验模态分解和傅里叶 分析的风场预测试验

张舰齐^{1,2}, 左瑞亭², 王丽琼²

(1. 中国人民解放军95871部队, 湖南 衡阳 421000; 2. 解放军理工大学气象海洋学院, 江苏 南京 211101)

摘要:采用集合经验模态分解(EEMD)和滑动傅里叶分析方法,建立了非线性气候序列的统计预测模型。针对气候要素距平场,对EOF分解得到的各模态时间系数进行EEMD分解,对得到的各IMF分量构建滑动傅里叶(Fourier)分析预报模型,提取出控制当前复杂气候信号的主要傅里叶频谱组合作为IMF分量的主要成分,即确定当前信号的主要波内频率,再将各个IMF分量和剩余项预测结果重构得到各模态时间序列的预测结果,最终通过时空重构得到预测场。将上述思想方法应用于新疆地区风场的预测试验,并采用距平相关系数(ACC),预报技巧(SS)和同号率(R)进行评估,结果表明对于区域性的风速预报,基于上述思想的算法模型能够较好地把握当前气候信号的主要变化频率,较为理想地预测了气候要素时间系数,对新疆地区风速变化的形态分布有较好的估计,使其预报时效在40候以内均拥有一定的预报技巧,平均SS在0.5以上,36候以内平均ACC达到0.4以上。

关键词: EOF分解; EEMD分解; 滑动傅里叶分析; 波内频率; 时空重构

文章编号: 1006-7639(2015)-05-0769-08 doi:10.11755/j.issn.1006-7639(2015)-05-0769

中图分类号: P457.5

文献标识码: A

引言

自20世纪60年代起,各种统计分析技术在气候预测中得到充分地应用,统计预测作为短期气候预测的重要手段之一,在当前甚至未来相当长的一段时间仍然占有重要地位^[1-2],特别是当前全球和区域气候模式对未来气候变化的预估仍然存在着诸多问题^[3],探索新的统计预测技术显得尤为重要。

各种要素的短期气候预测实质是对气候要素的距平场时间序列进行线性或者非线性外推。以往较常用的是直接对时间序列进行预测,如利用支持向量机对复杂的时间序列直接进行预测,然而气候中多数时间序列是非平稳非线性的,直接进行预测往往误差较大,效果并不好,于是一些学者开始运用隔离预测法对大气中的非线性非平稳的时间序列进行预测,取得了一些成果。所谓的隔离预测方法就是借助于某些分解手段,将时间序列分解为一系列具

有特定尺度的分量,进而再通过对每一特定的时间尺度分量进行预测,最后将这些预测结果合成作为最终的预测结果。如玄兆燕等^[4]运用经验模态分解法(Empirical Mode Decomposition, EMD)对复杂的大气时间序列进行了预测,预测精度比直接运用神经网络预测精度有较大的提高;刘科峰等^[5]运用小波技术将经验正交函数(Empirical Orthogonal Decomposition, EOF)分解的500 hPa位势高度场时间序列进行分解,得到了各个频段的高频信号和低频信号,再利用最小二乘支持向量机对各个频段进行了预测,最后将各个频段的预测结果重构合成得到了最终的时间序列预测结果,较好地预测了1~7 d的副热带高压活动;毕硕本等^[6]运用经验模态分解(EMD)将广西88个气象测站48 a的夏季降水距平百分率进行处理,再运用支持向量机(Support Vector Machine, SVM)对各个本征模态函数进行预测和不经EMD处理的反向传播(Back Propagation, BP)

收稿日期:2014-11-17; 改回日期:2015-01-11

基金项目:国家自然科学基金(1475071)资助

作者简介:张舰齐(1990-),男,助理讲师,主要从事气候变化研究. E-mail:15852933750@163.com

通讯作者:左瑞亭,男,博士,副教授, E-mail:ratinzuo@126.com

神经网络直接预测进行了对比验证,结果证明了经过 EMD 处理后的预测结果有较高的预测精度和较好的推广前景;文献^[7-10]均采用不同的隔离预测方法对时间序列进行预测,都取得了一定的效果。但是这些方法均有一个共同的特点,即无法克服大气信号中普遍存在波内频率调制现象^[11],特别是高频信号中波内频率调制现象更为严重,预测难度更大,因而导致预报结果有较大的误差,甚至预报结果与实际情况截然相反。因此,需要准确判断大气信号中波动频率,以达到较好的预报效果。本文正是基于该思想,尝试较为准确地捕捉气候信号中的波内频率,以此提升预测效果。当然,对特定气象要素的预报还可借助某些关键性的影响因子进行判断,例如厄尔尼诺对中长期预报的作用^[12-13],副热带高压或者南亚高压对降水的作用^[14],但是筛选因子是一个较为复杂的过程,一些因子还具有一定的时效性,过分夸大或降低因子的作用将会导致预报出现严重的偏差,因此文中暂不考虑因子的作用影响。

1 资料与方法

1.1 资料来源及处理

研究资料为新疆地区 35 个测站 1995~1998 年的逐日平均地面风速场。将新疆地区 35 个测站的逐日资料做候平均,再处理为标准化距平场。以 1995~1997 年的候平均标准化距平场作为历史资料,以 1998 年候平均标准化距平场数据资料作为预报效果的最终检验。

1.2 方法

1.2.1 经验正交函数分解法(EOF)

此方法又称之为主分量分析,其基本思想是将气候变量场的观测资料表示为固定空间场与时间序列的乘积形式,此方法广泛应用于大气科学研究领域中,具体的算法与步骤不再赘述。本文将站点资料处理成为风速场的标准化距平场,再用 EOF 对标准化距平场进行分解。

1.2.2 集合经验模态分解法(EEMD)

经验模态分解法的基本思想是把一个非线性非平稳的序列分解为具有某种时间尺度的特征模态函数(Intrinsic Mode Function, IMF)分量和一个趋势项^[15-17],基本方法是通过用 3 次样条函数拟合局部极小值构成下包络线和拟合局部极大值构建上包络线,通过取 2 条包络线间的均值,原序列与均值之差构成第一分量,第一分量是否是 IMF 还需要看是否满足以下要求:第一,是否消除了骑行波,使得序列的曲线变成极大值—零均值—极小值—零均值—极

大值的简单波形;第二是局部的峰和谷基本对称,如果不满足条件还需要重复上一步过程,其详细的叙述请参见文献^[5]中对此方法的操作步骤。

经验模态分解(EMD)在应用中的一个难题就是存在模态的混杂现象,由于观测资料由 2 部分组成,即观测的真实值与噪声部分,一般噪声部分来源于观察时的外部环境间歇性的干扰和仪器的测量误差等等。由于噪声的干扰使得经验模态的分解结果具有不确定性和不稳定性。为了消除经验模态分解的模态混合现象,相关学者^[18-19]提出了集合经验模态分解,其思想是在同一个大尺度的低频信号上多次加入互不相同的小尺度白噪声,构成大同小异的多个混合资料,它们分别经过 EMD 后会产生互不相同、彼此无关的多组 IMF 分量,既含有信号 EMD 的成分,也含有噪声 EMD 的成分,把这多组相对应的 IMF 分量集合起来取平均就有可能把噪声 EMD 的成分相互抵消只剩下信号 EMD 的成分。集合经验模态分解的观念基于以下的观测事实:其一,白噪声的作用经过集合后在时空的集合平均中相互抵消,因此只有输入序列真正的 IMF 分量才会在最后的集合平均中保存下来;其二,有限但不是无限小的振幅的白噪声,是集合覆盖全部解答所必须的,有限小振幅的白噪声使得不同尺度的信号保留在双重筛选后对应的 IMF 分量中,使集合平均更加具有物理意义;其三,资料的 EEMD 的结果不是来自原始没有加噪声的资料,而是来自大量的加进噪声的混合资料,最后得到的 IMF 分量是大量混合资料的 EMD 试验的集合平均。相关学者^[18-19]验证了添加白噪声对原始序列影响服从一定的统计关系。

1.2.3 预测评价指标

这里主要采用短期气候预测中国际与国内通用的评估指标^[20-22],即距平相关系数(ACC)、技巧评分(SS)、同号率(R)和均方误差。

(1) 距平相关系数(ACC)

使用距平百分率和平均距平计算距平相关系数,用下式表示:

$$ACC = \frac{\sum_{i=1}^N (\Delta R_f - \overline{\Delta R_f})(\Delta R_0 - \overline{\Delta R_0})}{\sqrt{\sum_{i=1}^N (\Delta R_f - \overline{\Delta R_f})^2 \sum_{i=1}^N (\Delta R_0 - \overline{\Delta R_0})^2}} \quad (1)$$

其中, ΔR_f 为距平的预测值, $\overline{\Delta R_f}$ 为距平预测的平均值。 ΔR_0 为实际距平值, $\overline{\Delta R_0}$ 为实际距平的平均值, N 为参与评分的空间点总数目。距平相关系数原应用于评估动力模式对位势高度距平场的预报,

在评估趋势的同时,主要反映距平量级预报的水平。

(2) 同号率(R)

具体计算公式为:

$$R = \frac{N_t}{N} \quad (2)$$

其中 N_t 和 N 分别为预测与实况距平符号相同格点或测站数目和参与评分的总的空间格点数目或测站数目,这一指标用来反映气象要素异常的趋势。

(3) 技巧评分(SS)

用公式表示:

$$SS = \frac{N_a - N'}{N - N'} \quad (3)$$

式中, N_a 、 N 分别为预报准确的站数和参加评分的总站数或空间格点数目,这里预报准确是指距平符号预报是否准确。 N' 为基于某种无技巧预期能预报准确(这里仅指距平符号报对;预报和实况出现零距平均视为正距平)的站数, N' 由 $N' = F_N$ 得到, F_N 即为随机预报或气候预报的预报准确率。当 $N_a = N'$ 时,技巧评分 SS 为 0,当 $N_a = N$ 时,技巧评分 SS 为 100%,当 $N_a < N'$ 时,技巧评分 SS 为负值。这里无技巧的对比预报采用随机预报和气候预报。

(4) 均方根误差

$$e_n = \sqrt{\frac{\sum_{i=1}^n (x_i - x_{oi})^2}{n}} \quad (4)$$

其中 x_{oi} 为实际值, x_i 为预测值, n 为参与计算的数据总量。

(5) 2 个序列的相关系数(r)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

其中 x 与 y 分别为两段序列, \bar{x} 为 x 序列的平均值, \bar{y} 为 y 序列的平均值, n 为参与计算的序列长度。

对于各个模态的时间系数主要采用均方根误差(e_n)、同号率(R)和相关系数(r)来评估,对区域性的预测场结果采用距平相关系数(ACC)和同号率(R)以及技巧评分(SS)评估。

2 预测建模

对于新疆地区的地面风速距平形势预测问题,拟先采用 EOF 分解技术将风速距平场进行时空分离,得到并选定需要预测的空间模态和对应的时间系数序列;对各个模态的时间系数序列进行 EEMD

分解,得到各个 IMF 分量。利用滑动傅里叶分析技术,对各个 IMF 分量进行统计预测;对预测得到的 IMF 分量进行 EEMD 重构和 EOF 重构,即可得到原地面风速距平的预测场。具体步骤如下:

(1) 对地面风速距平序列进行 EOF 分解

对新疆区域 35 个测站 219 个标准化候距平风速场进行 EOF 分解:

$$W_m(t) = \sum_{k=1}^n P_{m,k} \cdot T_k(t) \quad (6)$$

其中,测站序数 $m \leq 35$, 候序数 $t \leq 219$, $W_m(t)$ 为第 m 个测站第 t 候的风速, $P_{m,k}$ 为分解得到的第 k 个特征模态中第 m 个测站的值, $T_k(t)$ 为对应的第 k 个模态第 t 候的时间系数。为尽可能多的保留资料特征,以保证预测效果,取前 13 个模态进行后续预测。如表 1 所示,前 13 个模态的累积方差贡献已达到 90%,在实际应用中还可根据预测精度需要,增加模态数量,以提高累积方差贡献标准,保留更多资料特征。

表 1 新疆地区 35 个测站 1995 ~ 1997 年风速候距平场 EOF 分解结果

Tab. 1 The results of EOF analysis of pentad - mean wind velocity field in 35 stations in Xinjiang area from 1995 to 1997

序号	特征值	方差贡献/%	累计方差贡献/%
1	3 728. 00342	31. 091	31. 091
2	3 597. 41162	30. 002	61. 093
3	810. 49664	6. 759	67. 852
4	629. 51215	5. 250	73. 103
5	473. 00262	3. 945	77. 047
6	327. 17136	2. 729	79. 776
7	265. 56979	2. 215	81. 991
8	222. 97641	1. 860	83. 850
9	196. 94899	1. 643	85. 493
10	175. 60326	1. 465	86. 957
11	142. 76399	1. 191	88. 148
12	134. 23674	1. 120	89. 267
13	121. 63831	1. 014	90. 282

(2) 对各 EOF 模态时间系数序列 $T_k(t)$ 进行 EEMD 分解

从上述步骤得到的各模态时间系数序列中多次叠加入等长度的正态分布的白噪声 $n_i(t)$, $T'_{k,j}(t)$ 为第 k 个 EOF 模态第 j 次加入白噪声后的时间序列:

$$T'_{k,j}(t) = T_k(t) + n_j(t) \quad (7)$$

对(7)式所得到的序列进行 EMD 分解,得到第 i 个 IMF 分量 $F'_{k,i,j}(t)$ 和剩余项 $r'_{k,j}(t)$,再将各个 IMF 分

量 $F'_{k,i,j}(t)$ 进行整体平均最终得到 EEMD 分解的结果为:

$$F_{k,i}(t) = \frac{1}{N} \sum_{j=1}^N F'_{k,i,j}(t) \quad (8)$$

$$r_k(t) = \frac{1}{N} \sum_{j=1}^N r'_{k,j}(t) \quad (9)$$

N 为添加白噪声序列数目,文献^[14]证明,添加的白噪声对原始序列的影响遵循统计规律:

$$\varepsilon_n = \varepsilon/N \quad (10)$$

ε 是添加噪声序列的幅值, ε_n 为最终标准离差,即输入信号与相应的 IMF 重构结果的偏离。文献^[14]亦证明添加噪声幅值的大小对分解的结果影响并不是很大。这里在对经 EOF 分解后的时间系数进行 EEMD 分解时加入白噪声与原始信号的信噪比为 0.3,集合数目取 1 000。

(3) 对各 IMF 分量 $F_{k,i}(t)$ 进行预测

经 EEMD 分解得到的各 IMF 分量可分别视为由一组信号波叠加而成。其中,低频 IMF 分量的构成相对稳定,可认为是由持续时间较长且相对稳定的长波构成,预测难度低,预测准确性高;而高频 IMF 分量可以看成由多个短波分量叠加而成,容易出现波内频率调制现象,预测难度和误差较大。这里暂且认为,IMF 分量的各构成信号瞬时频率在 Δt 时间段内基本保持不变。则未来时间点瞬时频率可由其上一时间点的瞬时频率决定,即未来时段信号频率可近似由其紧邻的历史时段资料确定,而历史资料的显著频率可由傅里叶分析得到。如此进行傅里叶滑动分析,即可得到整个预报时效内各 IMF 分量的构成信号。具体操作说明如下:

① 对 IMF 分量序列 $F_{k,i}(t) = (f_1, f_2, f_3, \dots, f_n)$ 截取一段连续分量 $F_{k,i}|_{t=1:n-m} = [f_1, f_2, \dots, f_{n-m}]$, 截取后剩余 m 长度的序列,即剩余分量为 $F_{k,i}|_{t=n-m+1:n} = [f_{n-m+1}, f_{n-m+2}, \dots, f_n]$, 其应至少有一个极大值或者极小值以确定瞬时频率;根据截取的分量,依据信号的傅里叶表达式,使用最小二乘法计算出含有未知量 k 和 ω 的傅里叶表达式,其中 k 的选择范围为 $[1, 10]$ 。在拟合出的 10 个傅里叶表达式中通过对剩余分量的预测与对比试验进行筛选,以挑选出最佳的 k 和 ω 值,即通过以下原则筛选:拟合的公式对剩余部分的预测与剩余部分相关达到 0.6 以上;预测结果与剩余部分的均方根误差最小。通过以上两条原则所筛选出的 IMF 分量的傅里叶表达式应该较好地表现了各 IMF 分量的特征时间尺度,反映了控制当前气候信号中的主要频率。得到的 IMF 分量的傅里叶预报公式如下:

$$f_i(t) = c_0 + \sum_{n=1}^k a_n \cos(\omega_n t) + b_n \sin(\omega_n t) \quad (11)$$

② 预报下一个时刻后,截取分量向后滑动变换为 $F_{k,i}|_{t=2:n-m+1} = [f_2, f_3, \dots, f_{n-m+1}]$, 剩余分量则变为 $F_{k,i}|_{t=n-m+2:n+1} = [f_{n-m+2}, f_{n-m+3}, \dots, f_{n+1}]$, 依据步骤 1,再预报出下一个时刻。

③ 依据上述 2 个分析步骤,预测了 p 步,得到最后预测的 IMF 分量序列为 $F_{k,i}|_{t=n+1:n+p} = [f_{n+1}, f_{n+2}, \dots, f_{n+p}]$ 。

④ 对剩余趋势项 $r_k(t)$ 的预测则采用线性回归的方法预测,同样预测了 p 步,得到预测的趋势项序列为 $r_{k(n+1,n+p)}(t)$ 。

(4) 重构预测场

对预测场的时空重构分为 3 步:

① 重构 EOF 各模态的时间系数序列;依据 IMF 分量和剩余趋势项的预测结果,对时间系数进行重构,如下式所示:

$$T_{k(n+1,n+p)}(t) = \sum_{i=1}^M F_{k,i(n+1,n+p)}(t) + r_{k(n+1,n+p)}(t) \quad (12)$$

其中 $T_{k(n+1,n+p)}(t)$ 即为重构得到的预测了 p 步的时间系数。

② 重构预测场的时空分布。根据 EOF 分解原理,依据(6)式,由预测的各模态时间系数与对应的空间模态重构得到预测序列。

$$W_{m(n+1,n+p)}(t) = \sum_{k=1}^M P_{m,k} \cdot T_{k(n+1,n+p)}(t) \quad (13)$$

其中, $W_{m(n+1,n+p)}(t)$ 为第 m 个测站第 $n+1 \sim n+p$ 候的预测风速。

3 预测结果分析

3.1 时间系数的预测分析

将经过 EOF 分解得到的各个模态的时间系数做集合经验模态函数分解(EEMD),得到一系列的 IMF 分量。通过建立各个 IMF 分量的预测模型得到预报量,再通过预报量的合成得到最终的时间系数的预测。这里建立 3 种评价指标:预测曲线与实际曲线的同号率、二者之间的均方根误差、相关系数。前三模态的累积方差贡献为 67%,代表了新疆地区平均风速距平场的主要模态变化。各模态的时间系数预测效果参见图 1 和表 2。对比预测结果,首先是对主要模态的时间系数的预测达到了较好的效果,相关性较好,离散度较小,同号率比较理想。特别是对占有方差贡献率为 67% 的前 3 个模态同号

率在前 40 候都能达到 0.6 以上,第一模态时间系数预测相关系数平均达到了 0.74,第二、三模态的时间系数预测相关性在前 20 候都能达到 0.6,均方根误差也相对较小。其次是随着模态序号的递增,离散度加大,相关性出现了下降,同号率稍好。第 4 候的时间系数预测在前 10 候的效果较好,相关性达到了 0.6 以上,同号率较高,均方根误差也相对较小。而第 5、6 候系数的预测整体效果较差。再次是时间系数的预测效果中的同号率要好

于相关性,而相关性又好于均方根误差。其对时间系数趋势预测较为准确,强度稍弱。究其原因,第 1、2 模态的时间系数演变代表了新疆地区候平均风速距平场主要的演变特征,其包含的物理信息最多,因此模型可以很容易地抓住其波内主要的变化频率。而其它模态的时间系数波内频率演变较为复杂,控制其当前时间演变的频率变化较快,因而模型不容易获得其主要变化频率,预测起来难度较大,预测的时效较短。

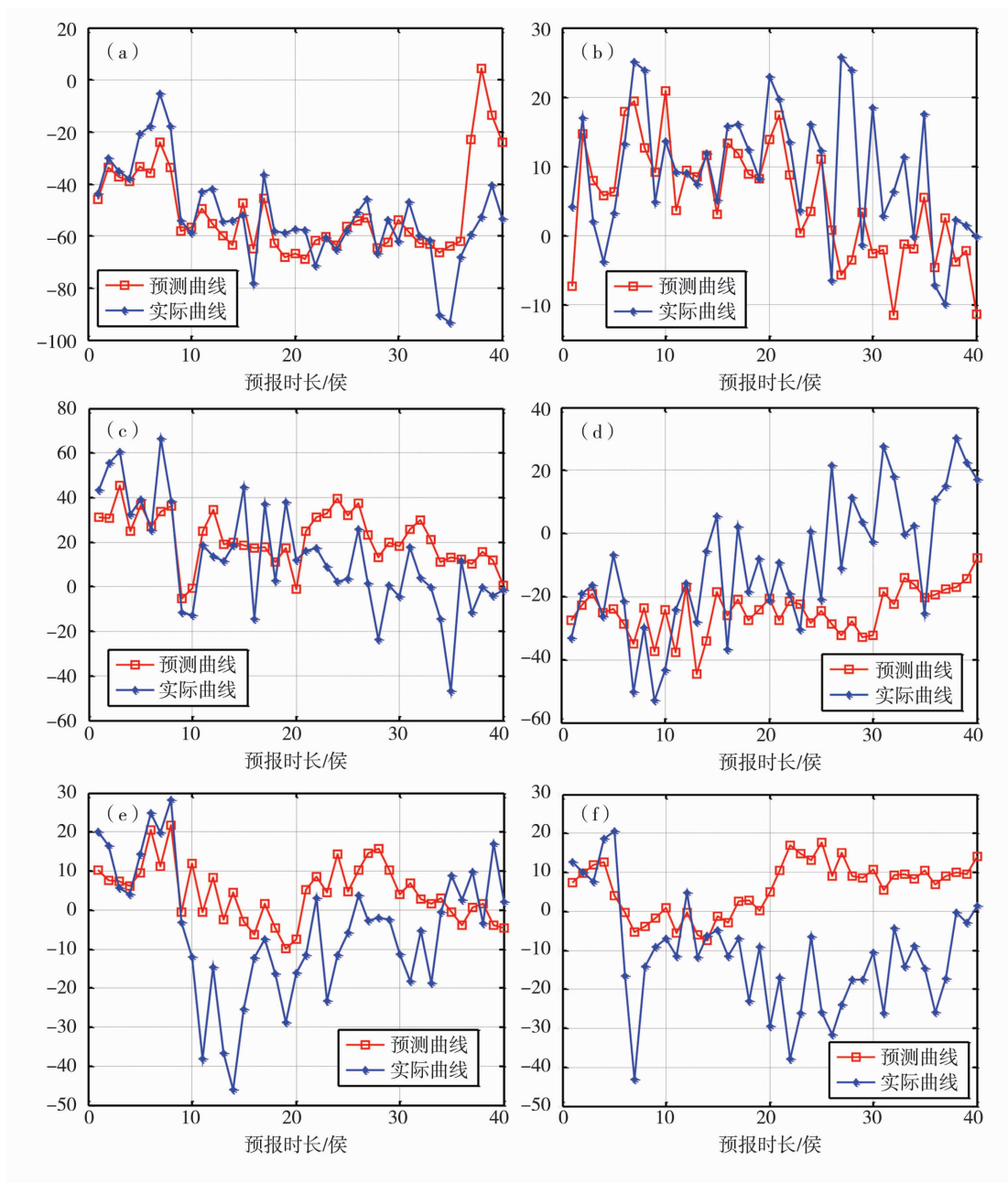


图 1 各 EOF 模态时间系数的预测与实际对比
a ~ f 分别对应第 1 ~ 6 模态

Fig. 1 The prediction of EOF components's time series and practical contrast
a - f corresponding to the each mode 1 - 6

表2 各 EOF 模态(对应序号 1~6) 时间系数的同号率(R)、相关系数(r)与均方根误差(e_i)的阶段分布

Tab.2 The distributing of anomaly sign score (R), correlation (r) and root - mean - square error (e_i) of each mode's (serial number 1 -6) time coefficient in different process

序号	候	1~5 候	6~10 候	11~20 候	21~30 候	31~40 候	平均
1	R	1	1	1	1	0.90	0.98
	r	0.89	0.99	0.73	0.49	0.60	0.74
	e_i	0.28	1.64	0.18	0.11	0.50	0.54
2	R	0.60	1	1	0.50	0.40	0.70
	r	0.55	0.39	0.81	-0.08	0.20	0.37
	e_i	1.15	0.53	0.29	1.40	37.29	8.13
3	R	1	1	0.80	0.80	0.30	0.78
	r	0.69	0.92	0.12	0.71	0.34	0.56
	e_i	0.28	0.54	1.47	15.25	84.88	20.48
4	R	1	1	0.80	0.60	0.20	0.72
	r	0.60	0.65	0.44	-0.58	0.17	0.26
	e_i	1.16	0.33	4.03	13.98	15.13	6.92
5	R	1	0.80	0.70	0.20	0.10	0.56
	r	-0.23	0.74	-0.33	-0.42	0.66	0.08
	e_i	0.46	0.99	0.96	4.17	2.79	0.87
6	R	1	0.80	0.50	0	0.10	0.48
	r	-0.23	0.74	-0.33	-0.42	0.66	0.08
	e_i	0.49	0.92	0.90	0.79	12.21	3.06

3.2 风场距平的时空重构

对时间系数的预测结果进行时空重构。用占有方差贡献率为近 80% 的前 6 项模态的空间场与预

测的时间序列进行重构,得到风速距平场的预测结果。由于预测的结果图较多,这里仅选取第五候和第十五候的预测与实际对比,结果如图 2 所示。

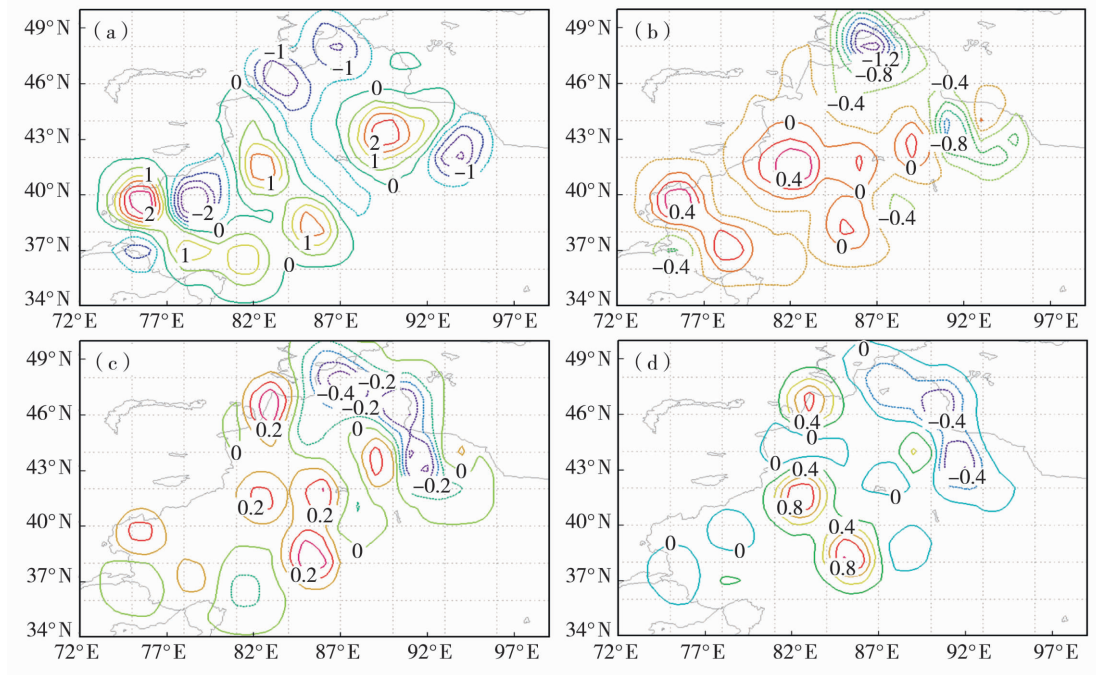


图2 新疆地区 1998 年候平均风速距平场预测与实际对比

(a) 第五候预测场, (b) 第五候实况值, (c) 第十五候预测场, (d) 第十五候实况值

Fig.2 The forecast of pentad - mean wind velocity anomalies vs its actual in Xinjiang of 1998

(a) the forecast of the fifth pentad, (b) the actual of the fifth pentad,

(c) the forecast of the fifteenth pentad, (d) the actual of the fifteenth pentad

从所选的两次预测值与实况值对比来看,预测场能够较好地预测出新疆地区风速距平分布形势变化,整体分布较为相似,对风速场距平的极值中心描述较为准确,但是强度稍有差异。对预测场的评估采用了 ACC , SS 和 R , 结论如图 3 所示。由此可以发现,随着预测时效的增加,三者均出现了下降趋势,但是 36 侯以内的 ACC 平均能达到 0.4 以上,具有较好的准确性,而且在强度上也具有一定的可信度。前 40 侯以内绝大多数的预测均有一定的预报技巧,40 侯以内的同号率能达到 0.5 以上,说明 40 侯以内的距平趋势预报较为准确。由此说明,该模型能够抓住新疆侯平均风速距平场形态变化的主要波动特征频率,进而能够较好地对其预测。

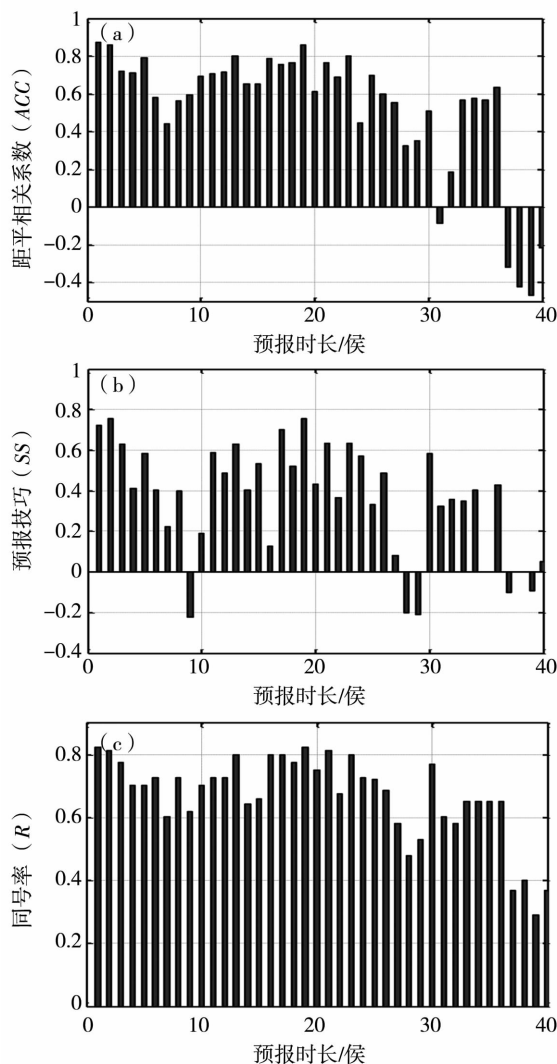


图3 对预测场的评估

(a) 距平相关系数(ACC), (b) 预报技巧(SS), (c) 同号率(R)

Fig. 3 The assessment of results

(a) the temporal anomaly correlation coefficients (ACC), (b) the skill score (SS), (c) the anomaly sign score (R)

4 结论

本文通过采用 EOF 对风速距平场进行时空分离,得到风速距平场各个模态的时间系数,对时间系数采用 EEMD 分解得到各个 IMF 分量,对各项 IMF 分量构建了滑动傅里叶分析预测模型。模型能够较好地抓住控制当前 IMF 分量的主要波动频率,并且能够达到较好的预测目的。将各个 IMF 分量的预测结果重构得到时间系数的预测结果,通过时间系数与空间场进行重构得到了风速距平场的预测结果。预测与实际对比结果显示,预测的效果良好,主要体现在 ACC 值较为理想,40 侯以内平均达到了 0.4 以上。本文主要有以下结论:

(1) 对于主要模态的时间系数(第一、二、三)的预测结果较好,反映在预报时效较长,在预报 40 侯以内时仍有较高的同号率、较高的相关性和相对较低的均方根误差。对主要模态的时间系数进行预测达到了较好的预测效果,相关性较好,离散度较小、同号率比较理想。随着模态序号的递增,离散度加大,相关性出现了下降,同号率稍好。时间系数的预测效果同号率要好于相关性,而相关性又好于均方根误差。对时间系数趋势预测较为准确,强度稍弱。究其原因在于主要模态的时间系数其频率的变化较为稳定。模型能够抓住波内主要的特征频率,进而能够达到较好的预测效果。

(2) 随着模态数目的增加,其相应时间系数的预测效果也呈现出下降趋势,反映在较低的相关性、较低的同号率和较高的均方根误差。究其原因在于,随着模态数目的增加,其相应的时间系数变化较为复杂,其时间系数的波内频率变化复杂多变,使得模型不容易抓住其频率变化特征,因而预测效果较差。

(3) 对距平场的时空重构结果发现,区域性风速场的侯平均距平预报在 40 侯以内有一定的预报技巧,虽个别侯预报效果较差,但平均来看预报技巧评分为正。对前 36 侯还有较高的 ACC , 对于前 40 侯均拥有较高的同号率。此模型对区域性风速形态有较好的估计。

由于本模型方法依靠历史资料的自身演变规律进行预报,并没有考虑其它因子,而风速的变化受诸多因素影响,因此还需要考虑更多因子对新疆风速场变化的影响,以进一步提高预报的准确性。此外,如果预测开始的节点上恰巧有频率的突变则可能使得模型错误地估计风速的变化频率,从而造成预报效果较差,这也是传统统计学预测方法面临的困难。

参考文献:

- [1] 魏凤英. 气候统计诊断与预测方法研究进展——纪念中国气象科学研究院成立 50 周年[J]. 应用气象学报, 2006, 17(6): 736-742.
- [2] 贾小龙, 陈丽娟, 高辉, 等. 我国短期气候预测技术进展[J]. 应用气象学报, 2013, 24(6): 641-655.
- [3] 翟颖佳, 李耀辉, 陈玉华. 全球及中国区域气候变化预估研究主要进展简述[J]. 干旱气象, 2013, 31(4): 803-813.
- [4] 玄兆燕, 杨公训. 经验模态分解法在大气时间序列预测中的应用[J]. 自动化学报, 2008, 34(1): 97-101.
- [5] 刘科峰, 张韧, 洪梅, 等. 基于最小二乘支持向量机的副热带高压预测模型[J]. 应用气象学报, 2009, 20(3): 354-359.
- [6] 毕硕本, 徐寅, 覃志年, 等. EMD 在广西季节降水预报中的应用[J]. 应用气象学报, 2010, 21(3): 366-371.
- [11] 黄大吉, 赵进平, 苏纪兰. 希尔伯特-黄变换的端点延拓[J]. 海洋学报(中文版), 2003(1): 1-11.
- [12] 郑崇伟, 潘静, 黎鑫, 等. 1988-2009 年中国海大浪频率对厄尔尼诺的响应[J]. 海洋通报, 2014, 33(2): 140-147.
- [13] 郑崇伟, 刘铁军, 钱粤海. 1988-2009 年中国海海面风场、海浪场对 El Nino 的响应[J]. 云南大学学报(自然科学版), 2014, 36(2): 214-223.
- [14] 张宇, 李耀辉, 魏林波, 等. 南亚高压与西太平洋副热带高压对我国西南地区夏季降水异常的影响[J]. 干旱气象, 2013, 31(3): 464-470.
- [15] Huang N E, Chen X, Lo M T, et al. On Hilbert spectral representation: A true time-frequency representation for nonlinear and nonstationary data[J]. Advances in Adaptive Data Analysis, 2011, 3(1/2): 63-93.
- [16] Huang N E, Wu M L, Qu W, et al. Applications of Hilbert-Huang transform to non-stationary financial time series analysis[J]. Applied Stochastic Models in Business and Industry, 2003, 19(3): 245-268.
- [17] Wu Z, Huang N E, Wallace J M, et al. On the time-varying trend in global-mean surface temperature[J]. Climate dynamics, 2011, 37(3/4): 759-773.
- [18] Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.
- [19] Huang N E, Wu Z, Pinzón J E, et al. Reductions of noise and uncertainty in annual global surface temperature anomaly data[J]. Advances in Adaptive Data Analysis, 2009, 1(3): 447-460.
- [20] 雷向杰. 短期气候预测质量评估方法与业务考核办法[J]. 陕西气象, 2008(6): 25-28.
- [21] 陈桂英, 赵振国. 短期气候预测评估方法和业务初估[J]. 应用气象学报, 1998, 9(2): 178-185.
- [22] 赵彦, 林朝晖, 李旭, 等. IAP PSSCA 两组预测试验的评估及比较 I. 降水部分[J]. 大气科学, 2000(2): 215-222.

Forecasting Experiments of Wind Based on Ensemble Empirical Mode Decomposition and Fourier Analysis

ZHANG Jianqi^{1,2}, ZUO Ruiting², WANG Liqiong²

(1. 95871th army of PLA, Hengyang 421000, China; 2. Institute of Meteorology and Oceanology, PLA University of Science and Technology, Nanjing 211101, China)

Abstract: A statistical forecast model suitable for nonlinear climatic series is established adopting the Ensemble Empirical Mode Decomposition (EEMD) and Sliding Fourier Analysis (SDFA). Empirical Orthogonal Decomposition (EOF) is firstly employed on climatic pentad anomaly to obtain their time series, and then EEMD is employed on these time series to get their Intrinsic Mode Functions (IMF), the primary Fourier spectrum signals which composed the major component of each IMF are extracted through the model building with SDFA, which equivalently indicates the achieving of primary frequencies of climatic signals. With the reconstruction of IMFs and the residual, the predicted time series can be obtained and be used for further rebuilding of spatial and temporal fields to accomplish the final forecasting. Such ideas is then applied in the wind velocity forecasting in Xinjiang. With the assessment of anomaly correlation coefficients (ACC), the skill score (SS) and anomaly sign score (R), the results show that the proposed algorithm model in ideas mentioned above can capture the primary frequencies of climate variation and give a good prediction on time coefficient of velocity series, and ultimately, a fairly good spatial and temporal wind distributions in Xinjiang are achieved successfully. During the whole prediction, the method shows a good skill, the average SS for the former 40 pentads exceeds 0.5 and the average ACC within 36 pentads surpass 0.4.

Key words: EOF analysis; EEMD method; sliding Fourier analysis; intrawave frequency; temporal reconstruction