

胡娅敏,覃志年,陈丽娟,等. 2013. 基于多时间尺度的回归集成预测模型. 气象, 39(9): 1182-1189.

# 基于多时间尺度的回归集成预测模型<sup>\* 1</sup>

胡娅敏<sup>1,2</sup> 覃志年<sup>3</sup> 陈丽娟<sup>4</sup> 罗晓玲<sup>2</sup>

1 广东省气候中心, 广州 510080

2 广州市气候与农业气象中心, 广州 510080

3 广西壮族自治区气候中心, 南宁 530022

4 国家气候中心中国气象局气候研究开放实验室, 北京 100081

**提 要:** 引入能够将非线性、非平稳过程的数据进行线性化和平稳化处理的 EMD 方法, 对广东降水的时间序列进行时间尺度分离, 从复杂的非平稳信号中提取相对简单以不同时间尺度振荡的准周期信号, 选取能较好描述降水周期特征的 IMF 分量作为建模备选因子, 然后以均生回归、均生相关、韵律拟合误差和拟合误差 4 种方法构建预测模型, 结果得到采用多尺度因子构建的 4 种单预测模型近 10 年  $P_s$  评分和降水距平符号同号率平均分在 68~73 分和 50%~58% 之间, 而采用 4 种模型构建的回归集成模型两种评分方法的平均分分别高达 79.8 和 68.8%, 较单一预测模型评分分别提高了近 10 分和 10% 以上。将具有降水指示信号的前冬赤道东太平洋海温因子耦合到回归集成预测模型, 其  $P_s$  评分结果与纯降水集成模型相当, 但同号率评分略高 3.1%。从而, 提取要素序列的多种时间尺度特征, 并采用多模型的集成预报, 均能有效提高短期气候预测水平。

**关键词:** 多时间尺度, 经验模态分解, 回归集成

**中图分类号:** P456

**文献标志码:** A

**doi:** 10.7519/j.issn.1000-0526.2013.09.014

## The Regression Ensemble Prediction Model Based on Multi-Time Scale

HU Yamin<sup>1,2</sup> QIN Zhinian<sup>3</sup> CHEN Lijuan<sup>4</sup> LUO Xiaoling<sup>2</sup>

1 Guangdong Climate Centre, Guangzhou 510080

2 Guangzhou Climate and Agricultural Meteorology Centre, Guangzhou 510080

3 Guangxi Climate Centre, Nanning 530022

4 Laboratory for Climate Studies, CMA, National Climate Centre, Beijing 100081

**Abstract:** The empirical mode decomposition (EMD) method has the advantage of dealing with the nonlinear and nonstationary data, making them linearized and stationary. So EMD is adopted to analyze the precipitation data based on the multi-time scale viewpoint, and the relatively simple semi-period signal with different oscillations are decomposed from the complex nonstationary and nonlinear signal. Then the characteristic intrinsic mode functions (IMFs) are chosen to construct the regression ensemble prediction model (REPM), which is based on the mean generation regression (MGR) method, the mean generation correlation (MGC) method, the rhythm fitting error (RFE) method and the fitting error (FE) method. The results show that the average score of the  $P_s$  and the same symbol ratio (SSR) are 68–73 and 50%–58%, respectively, among the four kinds of single models during rainy period in Guangdong for the recent 10 years. However, in the REPM, the average  $P_s$  and SSR scores have reached 79.8 and 68.8%, respectively, increasing 10 scores and 10% or so compared with one of the four kinds of single models. Meanwhile,

\* 国家自然科学基金青年基金项目(40905043)、广州市科技计划项目(2010Y1-C031)、广东省气象局课题(2011B04)、中国气象局气候预测创新团队、中英瑞适应气候变化项目(ACCC/2010527) 和中国气象局气候变化专项(CCSF2011-25)共同资助

2012 年 12 月 17 日收稿; 2013 年 3 月 6 日收修定稿

第一作者: 胡娅敏, 主要从事短期气候预测研究. Email: huyamin\_1978@sohu.com

if the SST signals in tropical East Pacific in the previous winter are coupled into the REPM, the  $P_s$  and SSR scores have improved, but the SSR scores, 3.1% higher than the former. Therefore, both the multi-time scale information extracting from the meteorological elements and the ensemble model construction can improve the accuracy of short-term climate prediction.

**Key words:** multi-time scale, empirical mode decomposition (EMD), regression ensemble prediction model

## 引言

目前,提高短期气候预测水平的常用方法有两种:一是对数值模式结果进行统计降尺度或动力降尺度应用,其预报效果的好坏依赖于数值模式预测结果的优劣。二是依靠改进统计方法,虽然统计方法存在着某些局限性和不稳定性,比如历史样本的有限性,统计方法无法对历史上没有出现过的气候异常强度和分布做出预测,历史资料得到的统计关系随着气候的长期变化也在不断的改变,甚至很多关系目前已经变得不如当初发现它们时显著。但是,由于数值预报水平发展仍有待提高,研究和发展的统计方法仍是提高省(市)短期气候预测水平的有效途径之一。

短期气候预测中引入的经典数学方法大多是针对线性和平稳时间序列进行分析的,而气象问题本质上都是非线性的,因而对于气象要素序列中很多非平稳和非线性过程不能较好地提取出有用的信息。经验模态分解(Empirical Mode Decomposition, EMD),即逐级进行平稳化处理,把不同周期的波动从原信号中分离出来,且该波动是平稳的,称该波动为本征模态函数(Intrinsic Mode Function, IMF),不同的 IMF 分量是平稳信号,具有非线性特征和缓变波包的特征。另外,EMD 方法依据数据自身的时间尺度特征来进行信号分解,无须预先设定任何基函数,这一点与建立在先验性的谐波基函数和小波基函数上的傅里叶分解与小波分解方法具有本质性的差别。因此,EMD 方法在处理非平稳及非线性数据上,具有非常明显的优势(张明阳等, 2007)。EMD 方法及相应的 Hilbert 变换正成为处理非线性、非平稳时间序列的有力手段,并已在生物、海洋、大气科学、天文学和工程技术等领域中得到了初步应用(林振山等, 2004; 郑祖光等, 2010)。许多气象学家开展了基于 EMD 方法对各种气象资料时间序列的分析工作,主要包括: MJO (Love et al, 2008)、降水 (McMahon et al, 2008)、气温(方仕

全等, 2005; 邹明玮, 2007; 玄兆燕等, 2008a; 2008b)、降水日数(毕硕本等, 2010)、副热带大气系统(侯威等, 2006)、成灾面积(刘莉红等, 2008)、海平面高度(刘莉红等, 2010b)、大气边界层高度(刘莉红等, 2010a)和海水温度(杨周等, 2010)。但是把这种方法初步应用到预测的研究主要有: 万仕全等(2005)和邹明玮(2007)以扬州 530 年(1470—1999 年)旱涝级别序列和北半球 1995 年(1—1995 年)树木年轮序列为例,采用 EMD 方法、均生函数和最优子集回归方法构建了一个新的预测模型,结果表明,特征 IMF 分量有较高的可预测性,它对原序列趋势的预测有重要指示意义。玄兆燕等(2008b)采用 EMD 和神经网络方法相结合对石家庄的气温和降水进行预测,结果得到 EMD 方法降低了被预测信号中的非平稳性,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。这些研究(万仕全等, 2005; 邹明玮, 2007; 玄兆燕等, 2008a; 2008b)采用实例数据在预测方面做了初步尝试,这为气候预测开辟一条新的有效途径。

另一方面,随着科技的发展,集成方法已经成为当前气候预测中的关键技术,尤其是气候模式发展的重要方向(陈法敬等, 2011; 尤凤春等, 2009; 狄靖月等, 2013; 纪永明等, 2011)。然而,目前大部分气象部门预测方案都是采用单一的数理方法来构建模型,由于使用的预测手段不同,考虑影响气象要素的物理因素不同,各种预报方法得到的预报结果也不尽相同或存在很大差异,但都能在一定程度上提供一些有用的信息。因此,若采用一种客观方法将各种预报结果加以集成,可提高对气象变量的短期气候预测准确率(魏凤英, 2007; 2011)。雷向杰(2011)基于多元回归预测法、月际持续性预测法、年际持续性预测法和基于 EOF 的 Downscaling 法共 4 种方法建立了集成预测模型,结果得到集成预测方法的效果明显优于单一预测方法。毕硕本等(2012)采用 EMD 对广西 2 月气温序列进行分解,然后对得到的 IMF 分量构建集成预报成员,用均生函数逐步回归法对各集成成员进行预测,结果表明加入 EMD 算

法和集成预报技术的方法比单一预测方法具有更好的预测能力。由此可见,集成方法的引入可有效改进短期气候预测效果。

本文引入 EMD 方法对广东降水时间序列进行多时间尺度分离,将复杂的非平稳信号简化为相对简单的不同时间尺度振荡的准周期信号,选取能较好描述降水周期特征的 IMF 分量作为预测模型的备选因子,然后分别采用均生回归、均生相关、韵律拟合误差和拟合误差 4 种方法对选取的备选因子构建预测模型。最后,参照雷向杰(2011)的研究方法,再以这 4 种预测模型为备选因子,采用多元线性回归方法构建集成预测模型。与毕硕本等(2012)研究不同的是,他们是对 IMF 分量构建集合序列,即考虑了初值的集合,本文是对预测方法进行集成,即考虑了预测模型的集成。本研究期望引入 EMD 方法和回归集成预测模型能提高广东降水的短期气候预测水平,为政府决策部门提前做出准确的指导。

## 1 资料和方法

### 1.1 资料处理

降水和气温数据是由广东省气候中心提供的 1961—2011 年广东 86 个台站的逐月降水和气温观测资料,以 1981—2010 年作为气候平均值。

用于降水和气温检验的方法为 2010 年中国气象局国家气候中心《短期气候预测质量分级检验办法》中的 Ps 六级评分方法。

### 1.2 研究方法

本文采用相关分析、经验正交函数(EOF)、EMD、均生回归、均生相关、韵律拟合误差、拟合误差、多元线性回归和回归集成等方法对资料进行处理、多时间尺度分离和模型构建。其中,均值生成函数(魏凤英等,1990)(Mean Generating Function, MGF),提出了视 MGF 为原序列生成的、具有周期性的基函数的新构思,设一时间序列  $x(t)$ :

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

式中,  $n$  为样本量。

$x(t)$  的均值为:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x(i) \quad (2)$$

对于式(1)定义均值生成函数:

$$\bar{x}_l(i) = \frac{1}{n_l} \sum_{j=0}^{n_l-1} x(i+jl) \quad (i = 1, \dots, l, 1 \leq l \leq m) \quad (3)$$

式中  $n_l = \text{INT}(n/l)$ ,  $m = \text{INT}(n/2)$  或  $\text{INT}(n/3)$ , INT 表示取整数。根据式(3),可以得到  $m$  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,构造均生函数延拓矩阵。而均生回归法,即先把降水序列作均生延拓,对得到的所有均值生成函数因子,采用多元线性回归方法(魏凤英,2007;吴诚鸥等,2007;黄嘉佑,2004)建立预测模型。另外,均生相关法,即先把降水序列作均生延拓,再对延拓结果与原序列计算相关,取 2~7、7~15 和 15~ $N/2$  年间( $N$  为序列长度)的延拓序列高相关的 3 个因子,即找到长、中、短 3 个不同尺度因子,采用多元线性回归方法建立预测模型。韵律拟合误差法参见魏淑秋(1985)工作。拟合误差法参见谢小康(1994)工作。本文主要采用 EMD(郑祖光等,2010)方法对各月、季降水序列进行多时间尺度分离,分别采用均生回归、均生相关、韵律拟合误差和拟合误差 4 种方法对基于 EMD 得到的 IMF 分量构建月、季降水预测模型。

## 2 基于 EMD 算法的多时间尺度信息提取及集成模型构建

### 2.1 基于 EMD 算法的多时间尺度信息提取

长时间的降水序列本身蕴含了其多重周期演变特性,但由于影响降水系统的复杂性,使该序列包含了许多非平稳和非线性信息,如何从中提取出具有预测指示意义的周期特征,是众多研究者关注的重点之一,EMD 工作在尺度分离工作中的优势已被一些学者证明。

对广东各月或季降水序列进行标准化,然后用 EMD 方法分解,通常能得到 5~6 个 IMF 分量和 1 个趋势项,计算各 IMF 分量与原序列的相关系数,以及 IMF 分量所占方差贡献,选取相关系数高、方差贡献大的 IMF 因子作为 2.2 节建模的备选因子。

选取广东汛期(4—9 月)降水为例进行 EMD 分析。首先,对 1961—2011 年广东汛期降水进行 EMD,得到 4 个 IMF 分量和 1 个趋势项(图 1)。在构成广东汛期降水量变化的 4 个不同时间尺度的波频中,IMF1 与原序列相关系数为 0.78,其周期分别

表现为 4 a, 而 IMF2、IMF3 和 IMF4 的相关系数分别为 0.25、0.22 和 0.16, 周期分别为 7、12 和 25 a, 最后一项 IMF5 为趋势项, 与原序列相关系数为 0.07, 表现为自 20 世纪 60 年代中期以来广东汛期降水一直呈现上升趋势, 但近 5 年增加趋势不明显。

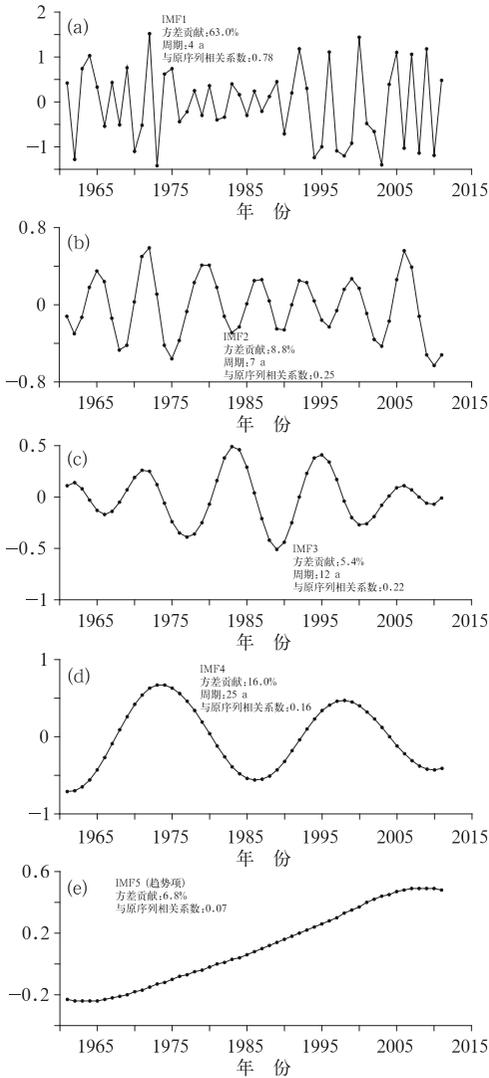


图 1 1961—2011 年广东汛期降水量的 IMF1~IMF5 分量

Fig. 1 The IMF1 to IMF5 components of Guangdong precipitation during rainy period from 1961 to 2011

计算各 IMF 分量的方差贡献得到, IMF1 占方差的贡献达 63.0%, 比较 IMF1 分量与原序列 (图 2a), 可以看出 IMF1 基本能拟合出原序列, 说明汛期降水主要以 4 a 振荡为主。IMF1 和 IMF4 这 2 个分量累积方差贡献达 79%, 取这两个 IMF 分量合成与原序列曲线的对比 (图 2b), 可以看出合成

曲线基本包含了原序列信息, 与原序列相关系数高达 0.88, 即在 IMF1 分量中加入 25 a 长周期的 IMF4 分量, 其合成效果较图 2a 更好。若取 IMF1~IMF4, 则累积方差贡献达 93.2%, 与原序列相关系数高达 0.97, 其重构值与原值非常接近, 可见 EMD 方法分离出多时间尺度信息基本能重构原信息的特征。

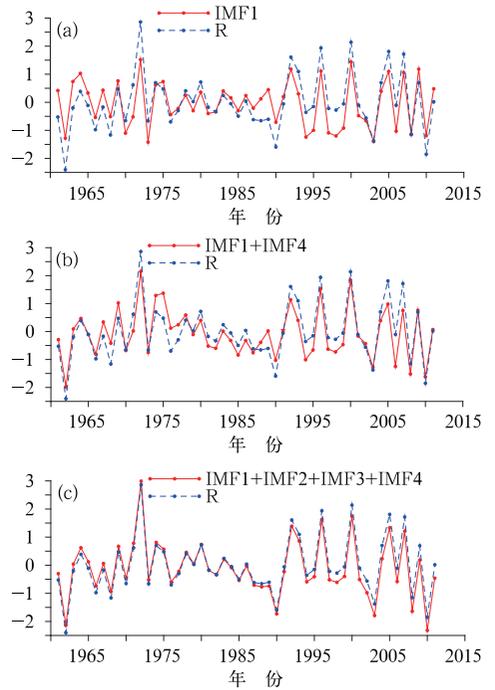


图 2 汛期降水量与 IMF1 分量 (a)、IMF1+IMF4 (b) 和 IMF1+IMF2+IMF3+IMF4 重构序列 (c) 的对比 (虚线: 原序列, 实线: IMF 分量合成)

Fig. 2 The comparison between the precipitation and the first component (a), the composition from the first and the fourth components (b) and the composition from the first, the second, the third, the fourth components (c) during rainy period in Guangdong (The dashed line is the precipitation and the solid line is the IMF components)

## 2.2 基于多尺度信息的统计模型构建

对基于 EMD 方法得到具有较高相关系数和较大方差贡献的 IMF 分量, 分别采用均生相关法、均生回归法、韵律误差法和拟合误差法构建降水的预测模型, 建模年份为 1961 年至预测前一年, 回报检

验时采用逐年向前滚动检验法。其预报结果如表 1 和表 2, 可以看出 4 种预测方法对近 10 年广东汛期降水的  $P_s$  评分平均值在 68~73 分之间, 而降水距平同号率评分的平均值在 50%~58%。与原降水序列构建模型相比(表略), 多时间尺度各单一预测模型均有不同程度的提高,  $P_s$  评分增加了 3 分, 同号率评分增加了 5%, 可以看出若剔除了原序列中的噪音, 在一定程度上能有效改进预报效果。

表 1 基于 EMD 多时间尺度的预测模型广东汛期降水  $P_s$  评分

Table 1 The  $P_s$  score of Guangdong precipitation from April to September during 2002—2011 based on the multi-time scale prediction model from EMD

年份	均生相关	均生回归	韵律	拟合误差	回归集成
2002	69.2	72.3	68.3	69.8	80.2
2003	72.8	74.4	68.0	76.6	85.3
2004	70.0	74.1	71.0	67.8	78.3
2005	74.4	72.8	70.2	70.5	81.3
2006	72.1	72.2	62.2	76.7	78.4
2007	70.9	73.8	70.3	71.4	80.7
2008	74.5	71.9	67.0	72.4	77.9
2009	57.4	67.1	69.5	64.7	77.9
2010	66.4	73.8	72.2	70.9	78.1
2011	74.1	76.7	64.4	70.9	80.3
平均	70.2	72.9	68.3	71.2	79.8
大于 60 的年份	9	10	10	10	10

表 2 基于 EMD 多时间尺度的预测模型广东汛期降水距平符号的同号率评分(单位: %)

Table 2 The anomaly sign consistency rate between the hindcast and the observed precipitation in Guangdong from April to September during 2002—2011 based on the multi-time scale prediction model from EMD (unit: %)

年份	均生相关	均生回归	韵律	拟合误差	回归集成
2002	46.5	50.0	47.7	45.3	67.4
2003	55.8	55.8	44.2	61.6	70.9
2004	52.3	59.3	55.8	47.7	65.1
2005	59.3	53.5	51.2	53.5	69.8
2006	66.3	66.3	46.5	68.6	72.1
2007	51.2	52.3	48.8	50.0	66.3
2008	68.6	61.6	52.3	67.4	69.8
2009	38.4	50.0	52.3	44.2	66.3
2010	45.3	53.5	53.5	51.2	61.6
2011	69.8	74.4	52.3	60.5	79.1
平均	55.3	57.7	50.5	55	68.8
大于 50 的年份	7	10	6	7	10

## 2.3 基于回归方法的集成预测模型构建

集成预报的基本含义是将两个以上模型的预报

结果用统计方法集成为单一的预报结果。集成预报的关键是如何确定权重系数。通常采用简单的算术平均或根据各种方法事先人为设定历史预报技巧或用回归系数给各种预报方法不同的权重。在预报样本量不是足够大的情况下, 算术平均通常不能得到最优集成预报。在有限样本情况下, 回归系数可以保证在最小方差意义下得到最优集成拟合(魏凤英, 2007), 因此本文选取回归集成法(朱伯承, 1981)构建预测模型。回归集成法将  $n$  种原始预报模型  $y_1, y_2, \dots, y_n$  作为新的预报因子, 求预报量实况值  $y$  的回归方程:

$$p = a_0 + a_1 y_1 + a_2 y_2 + a_3 y_3 + a_4 y_4 \quad (4)$$

其系数满足如下线性方程组

$$\begin{cases} s_{11} a_1 + s_{12} a_2 + \dots + s_{1n} a_n = s_{1y} \\ s_{21} a_1 + s_{22} a_2 + \dots + s_{2n} a_n = s_{2y} \\ \vdots \\ s_{m1} a_1 + s_{m2} a_2 + \dots + s_{mn} a_n = s_{my} \\ a_0 = \bar{y} - (a_1 \bar{y}_1 + a_2 \bar{y}_2 + \dots + a_n \bar{y}_n) \end{cases} \quad (5)$$

$$s_{ij} = \sum_{k=1}^M (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j)$$

其中:

$$s_{iy} = \sum_{k=1}^M (y_{ik} - \bar{y}_i)(y_k - \bar{y}_j)$$

式中,  $M$  为预报次数,  $y_{ik}$  为第  $i$  种原始预报方法所作的第  $k$  次预报值,  $y_k$  为第  $k$  次实况值。将式(5)的解代入式(4)便得回归集成预报方程。

本文选取均生相关  $y_1$ 、均生回归  $y_2$ 、韵律误差  $y_3$  和拟合误差  $y_4$  共 4 种预报模型作为集成成员的备选因子, 采用多元线性回归方法构建如式(4)的集成预测模型。取 4 种单预测模型预测年份前 10 年的回报值代入式(5), 得到  $a_0, a_1, a_2, a_3$  和  $a_4$  5 个系数的解, 代入式(4)即得到回归集成预测模型。该集成预测模型是动态的, 各台站的回归系数随预测对

表 3 以汛期降水 5 个台站为例得到  
的集成模型的回归系数

Table 3 The ensemble model regression coefficients for precipitation at five stations of Guangdong from April to September

台站	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$
阳山	-1.7	0.41	0.0	0.98	-0.65
广州	16.5	-1.71	2.57	0.58	-1.06
饶平	-22.1	0.28	-2.74	0.71	1.19
台山	-0.2	0.0	1.00	0.55	-0.80
中山	-5.2	0.18	0.0	-0.47	-0.42

象(降水或气温等)、预测年和预测时段(季、月及旬等)的变化而动态改变。以汛期降水为例,取2002—2011年86个台站近10年4种单预测模型的回报值代入式(5),分别计算得到各台站的5个回归系数,表3列出任意5个台站的 $a_0, a_1, a_2, a_3$ 和 $a_4$ 回归系数,可以看出在不同的台站预测过程中各种预测模型发挥着不同的作用,效果各有优劣。

对回归集成方法近10年广东汛期进行回报(表1和表2),可以看出回归集成的效果明显优于各单独预测模型, $P_s$ 评分的平均值上升到79.8分,较单一模型中最优的均生回归法增加了6.9分,降水距平符号同号率平均值为68.8%,较单预测模型增加了10%以上。较原降水序列回归集成模型(表略),其 $P_s$ 评分和同号率评分分别增加了10分和10%以上,说明开展物理模型的集成预测,能有效提高短期气候预测水平。

进一步分析各预测模型的空间预测效果,以2012年汛期为例,对广东86站降水进行EMD分解后,选取特征IMF分量,分别采用均生相关法、均生回归法、韵律拟合误差法和拟合误差法构建模型进行预测,得到如图3所示结果,实况的分布为广东省大部分降水偏少,各单预测模型也均预测出降水偏少的形势,其 $P_s$ 评分在67.4~74.2分之间。对这4种预测方法进行回归集成,得到的预测结果与实况最为接近,其 $P_s$ 评分在77.2分,可以看出,回归集成较其他单一预测模型在空间分布和量级的预测上较单一预测模型更优。

### 3 前冬海温对广东汛期降水的影响及其与序列方法的集成

虽然降水序列本身包括了很多重要信息,但降

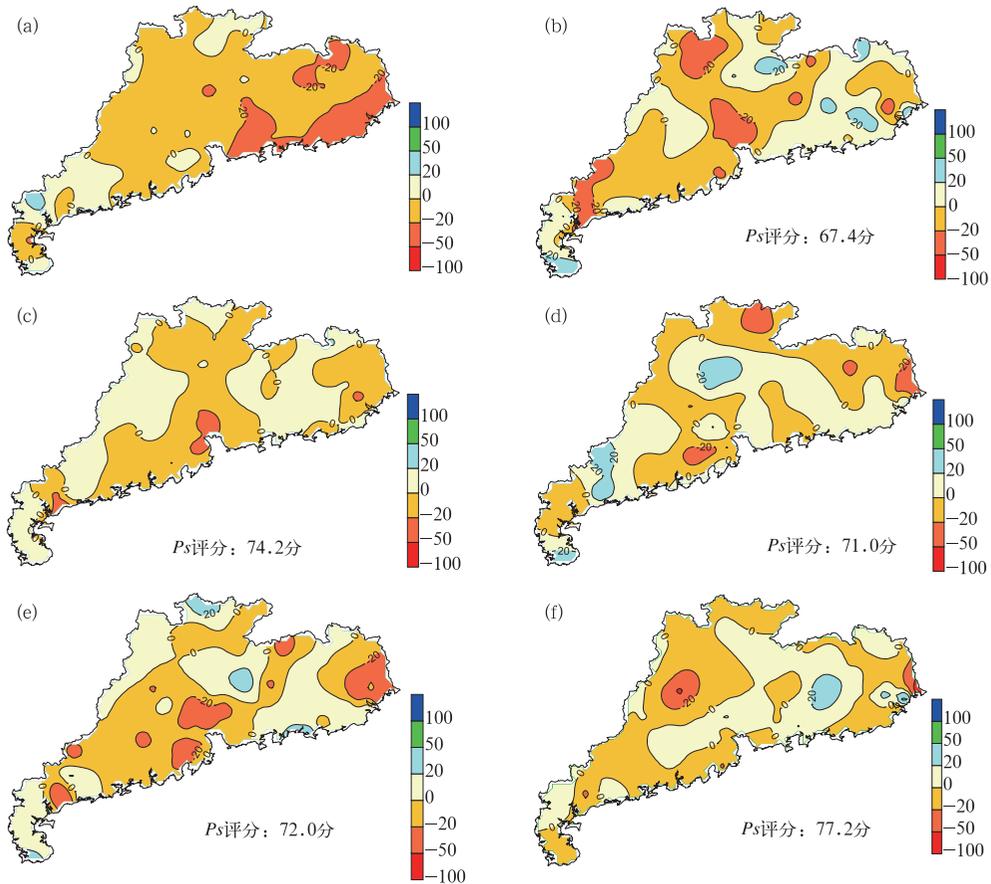


图3 2012年汛期广东降水预测实况及各种预测结果

(a)实况,(b)均生相关,(c)均生回归,(d)韵律拟合误差法,(e)拟合误差,(f)回归集成

Fig. 3 The precipitation observation (a) and prediction results from the methods of mean generation correlation (b), mean generation regression (c), rhythm fitting error (d), fitting error (e) and regression ensemble (f)

水本身只是一个事后信息,更有效的提高短期气候预测的方法是分析影响降水的物理系统,由于影响降水系统的复杂性,这种信息具有非线性和非平稳性。众所周知,降水的成因很复杂,因而降水预测是目前短期气候预测的难点和重点。降水不仅受自身变化规律的影响,同时受到外强迫和大气环流影响。随着全球气候变暖变“乱”,仅用降水时间序列本身预测一方面不稳定,另一方面预测不出异常等级。因此,有必要考虑对降水预测具有指示意义的外强迫信号和环流因子。然而,考虑到环流因子是一个相对的快变量,很容易遗忘前期信号对降水的指示,另一方面大气的混沌性会限制季节尺度的预报性。因此,下面开展海温外强迫因子与降水自身规律结合构建预测系统。

将广东 86 站汛期降水量进行 EOF 展开,取能表述其平均特征的第一特征向量对应的时间系数与前期不同时期全球海温(SST)计算相关,结果得到前冬(12 月到次年 2 月)赤道东太平洋地区有一正高相关中心(图略),对该区域海温序列进行 EMD 展开,取前 3 个 IMF 分量作为预测备选因子构建逐步回归方程。对 2002—2011 年海温模型进行回报检验(表 4),近 10 年平均  $P_s$  和同号率分别为 68.3 分和 52.8%,与降水单预测模型结果相当。进一步将海温因子耦合到回归集成预测模型,其  $P_s$  评分结果与纯降水集成模型相当,但同号率评分略高 3.1%。海温因子对降水的可能影响:当冬季赤道东太平洋海温偏高,有利西太平洋副热带高压偏南,易出

表 4 采用前冬 SST 信号预测及其与降水序列预测的集成结果对比

Table 4 The comparison between the prediction from SST signal of previons winter and the results with the SST coupled into the precipitation ensemble prediction model

年份	前冬 SST 模型		序列+SST 集成	
	同号率/%	$P_s$	同号率/%	$P_s$
2002	32.6	61.0	70.9	82.3
2003	70.9	79.5	66.3	79.3
2004	80.2	79.0	62.8	74.0
2005	27.9	60.7	70.9	79.7
2006	30.2	54.3	83.7	84.8
2007	50.0	70.0	75.6	82.4
2008	10.5	38.8	83.7	80.7
2009	68.6	77.0	70.9	80.5
2010	69.8	80.1	65.1	80.3
2011	87.2	82.1	68.8	75.1
平均	52.8	68.3	71.9	79.9
大于 50 的年份	6	9	10	10

现南方类雨型(陈兴芳等,2003)。

## 4 结论和讨论

本文引入能够将非线性、非平稳过程的数据进行线性化和平稳化处理的 EMD 方法,对广东降水和影响降水的海温因子的时间序列进行时间尺度分离,从复杂的非平稳信号中提取出相对简单的不同时间尺度振荡的准周期信号,选取能较好描述降水周期特征的关键备选因子,然后以均生回归、均生相关、韵律拟合误差和拟合误差 4 种方法构建集成预测模型。结果得到:

(1) EMD 方法能提取降水序列不同尺度的周期特征,广东汛期降水主要呈现为 4、7、12 和 25 a 周期,其中以 4 a 周期为主,占总方差贡献的 63.0%,其次是 25 a 长周期,它和 4 a 短周期占方差贡献的 79%,与原序列相关高达 0.88,4 和 25 a 周期的重构值基本能包含汛期降水的绝大部分信息。

(2) 采用均生相关、均生回归、韵律拟合误差法和拟合误差预测方法对近 10 年广东汛期进行回报,结果得到各单一预测模型的  $P_s$  评分的平均分为 68~73 分,同号率评分的平均分为 50%~58%,而回归集成  $P_s$  评分和同号率的平均值分别达到 79.8 分和 68.8%,较单一预测模型的评分分别偏高了 10 分和 10%以上。同时,回归集成也较好地模拟出降水的偏多空间分布型。从而开展物理模型的集成预测,能有效提高短期气候预测水平。

(3) 将具有降水指示信号的前冬赤道东太平洋海温因子耦合到回归集成预测模型,其  $P_s$  评分结果与纯降水集成模型相当,但同号率评分略高 3.1%。可见,寻找对降水预测具有指示意义的外强迫信号和环流因子可一定程度上提高降水预报的准确率。

本文采用 EMD 方法提取能较好描述降水周期特征的关键备选因子,并分别采用均生回归、均生相关、韵律拟合误差和拟合误差 4 种方法构建预测模型,接着采用回归方法对 4 种模型进行集成,研究得到引入 EMD 和回归集成方法能有效提高短期气候预测效果。同时,将具有降水指示信号的前冬赤道东太平洋海温因子耦合到回归集成预测模型,其  $P_s$  评分结果与纯降水集成模型相当,但同号率评分略高 3.1%。寻找对降水预测具有指示意义的外强迫信号和环流因子可一定程度上提高降水预报的准确

率。但是由于影响降水因子的复杂性,且外强迫因子和环流因子均是和降水序列作前期相关来寻找关键区,而这种相关关系并不十分稳定,且这种信息易受非线性作用的干扰,因而如何提取影响降水系统中可识别的具有可预报性的物理因子,是今后将进一步深入开展的工作。

## 参考文献

- 毕硕本,陈讚,覃志年,等. 2012. 基于 EMD 和集合预报技术的气候预测方法. 热带气象学报, 28(2): 283-288.
- 狄靖月,赵琳娜,张国平,等. 2013. 降水集合预报集成方法研究. 气象, 39(6): 691-698.
- 陈法敬,矫梅艳,陈静. 2011. 一种温度集合预报产品释用方法的初步研究. 气象, 37(1): 14-20.
- 陈兴芳,赵振国. 2003. 中国汛期降水预测研究及应用. 北京:气象出版社.
- 侯威,封国林,董文杰,等. 2006. 利用排列熵检测近 40 年华北地区气温突变的研究. 物理学报, 55(5): 2663-2268.
- 黄嘉佑. 2004. 气象统计分析与预报方法(第 3 版). 北京:气象出版社, 58-70.
- 雷向杰. 2011. 月平均气温集成预测方法研究. 气象, 37(12): 1560-1565.
- 纪永明,陈静,矫梅燕,等. 2011. 基于多中心 TIGGE 资料的区域 GRAPES 集合预报初步试验. 气象, 37(4): 392-402.
- 林振山,汪曙光. 2004. 近四百年北半球气温变化的分析:EMD 方法的应用. 热带气象学报, 20(1): 90-96.
- 刘莉红,翟盘茂,郑祖光. 2008. 中国北方夏半年最长连续无降水日数的变化特征. 气象学报, 66(3): 474-477.
- 刘莉红,翟盘茂,郑祖光. 2010a. 中国北方夏半年极端干期的时空变化特征. 高原气象, 29(2): 403-411.
- 刘莉红,郑祖光,琚建华,等. 2010b. 夏季副热带大气系统的多尺度振荡分析. 高原气象, 29(1): 115-127.
- 万仕全,封国林,周国华,等. 2005. 基于 EMD 方法的观测数据信息提取与预测研究. 气象学报, 63(4): 516-525.
- 魏凤英. 2007. 现代气候统计诊断与预测技术(第 2 版). 北京:气象出版社, 182, 222.
- 魏凤英. 2011. 我国短期气候预测的物理基础及其预测思路. 应用气象学报, 22(1): 1-11.
- 魏凤英,曹鸿兴. 1990. 建立长期预测模型的新方案及其应用. 科学通报, 35(10): 777-780.
- 魏淑秋. 1985. 农业气象统计. 福建:科学技术出版社.
- 吴诚鸥,秦伟良. 2007. 近代实用多元统计分析. 北京:气象出版社.
- 谢小康. 1994. 用拟合误差分析寻找周期作七月雨量预报. 贵州气象, 18(3): 13-14.
- 玄兆燕,杨公训. 2008a. 经验模态分解法在大气时间序列预测中的应用. 自动化学报, 34(1): 97-101.
- 玄兆燕,杨公训. 2008b. EMD 在地面气温预测中的应用. 微计算机信息, 24(7): 516-525.
- 尤凤春,魏东,王雨. 2009. 北京奥运期间多模式降水检验及集成试验. 气象, 35(11): 3-8.
- 张明阳,王克林,刘会玉,等. 2007. 基于 EMD 的洪涝灾害成灾面积波动的多时间尺度分析——以湖南省为例. 中国生态农业学报, 15(1): 131-134.
- 郑祖光,刘莉红. 2010. 经验模态分析与小波分析及其应用. 北京:气象出版社, 32.
- 朱伯承. 1981. 统计天气预报. 上海:科技出版社, 373.
- 邹明玮. 2007. 基于非线性时间序列分析方法的非线性系统特性研究. 扬州:扬州大学.
- Love B S, Matthews A J, Janacek G J. 2008. Real-time extraction of the Madden-Julian oscillation using empirical mode decomposition and statistical forecasting with a VARMA model. J Cli, 21(20): 5318-5335.
- McMahon T A, Kiem A S, Peel M C, et al. 2008. A new approach to stochastically generating six-monthly rainfall sequence based on empirical mode decomposition. J Hydrometeoro, 9(6): 1377-1389.