

基于 EMD 方法的观测数据信息提取与预测研究^{*}

万仕全

(江苏省扬州市气象局,扬州,225000; 国家气候中心气候研究开放实验室,北京,100081)

封国林

(国家气候中心气候研究开放实验室,北京,100081)

周国华 王冰梅 秦铭荣 许遐桢

(江苏省气象局,南京,210009)

摘 要

用统计方法作月、季尺度的短期气候乃至年际尺度的长期气候预测是当前气候预测业务的主要依据,在短时期内这种情况仍然不可能彻底改变。虽然数值预报模式的预测能力达到了 7 d 的时效,不过要积分到月、季尺度并实现短期气候预测还面临着重重困难。其根本原因是气候系统的混沌分量和非线性/非平稳性等因素在起作用。而现有气候预测的统计方法(主要包括经验统计、数理统计和物理统计等方法)的数学基础却忽略了这些特点,这是因为以现有的科学水平人们不得不假设时间序列是线性和平稳的。实际气候观测序列普遍具有层次性、非线性和非平稳性,这给建立预测方法带来了极大困难。文中构建了一个新的预测模型,即首先利用经验模态分解(empirical mode decomposition, EMD)方法将气候序列作平稳化处理,得到一系列平稳分量-本征模函数(intrinsic mode function, IMF);其次,利用均生函数(mean generate function, MGF)模型获得各分量的初次预测值;最后,在最优子集回归(optimal subset regression, OSR)模型的基础上,通过直接或逐步拟合一部分预测值,构建两种预测方案达到提高预测能力的目的。典型气候序列的预测试验结果表明,具有平稳化的 IMF 分量,尤其是特征 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。大力开展气候系统机理和气候层次的研究,并建立相应的气候模式是未来发展趋势。该文是这方面的一个初步尝试,相信该模型能为气候预测(评估)开辟一条新的有效途径。

关键词: 经验模态分解,非线性/非平稳时间序列,层次理论,气候预测。

1 引 言

众所周知,气候系统是非线性复杂系统,其固有的非线性决定了气候序列具有明显的非线性/非平稳性。迄今为止,人们还没有建立任何关于非平稳过程的预测理论,现有的气候预测方法无一例外地构筑在平稳性假设基础之上,如均生函数(mean generate function, MGF)预测模型、最优子集回归(optimal subset regression, OSR)模型等,其数学基础决定了它们不可能有效处理非线性/非平稳问题。研究表明,气候系统除了具有非线性/非平稳性,还

有层次性,许多大小不一的时、空尺度构成了多层次结构。不同层次的气候具有不同的可预报性和稳定性^[1~4],各层次间呈高度自组织。事实上,气候系统的多层次结构正是导致其非平稳行为的重要原因之一。一个气候态可以由低层次向高层次转化,这个状态的改变通常称为“突变”,这是非平稳性的集中体现。一般而言,通过非平稳数据建立精确有效的拟合预测方程是相当困难的^[5~9],因此,近年来有的学者开展了一些针对气候层次的预测研究^[10],方法之一是根据小波系数的准周期性作周期延拓而达到预测目的,某种意义上部分实现了气候的“多层次预

^{*} 初稿时间:2005 年 1 月 8 日;修改稿时间:2005 年 4 月 6 日。

资助课题:国家自然科学基金(90411008 和 40325015)及国家重点发展基础研究项目(2006CB400503)。

作者简介:万仕全,男,1976 年生,工程师、硕士,从事气候预测、气候变化研究;E-mail: wan_sq@163.com

测”。

集平稳化和层次化处理能力于一体的 EMD 方法,为探索上述问题开辟了一条新的有效途径。Huang 等^[11,12]针对非线性/非平稳时间序列的 Hilbert 谱分析而提出了一种新的处理方法——经验模态分解,此后又对其作了改进。EMD 方法可将非平稳时间序列平稳化,得到一系列不同频率的分量-本征模函数。也就是说,非平稳/非线性气候资料可以分解成不同时间尺度的平稳信号,这一过程称为“数据筛选”过程。筛出的最初 IMF 分量代表原始序列的高频组分(小尺度或低层次),随着分解的深入,相应 IMF 的频率变小、周期增大。这些 IMF 可作为原序列的一组完全或几乎正交的展开基,且它们与一个平均趋势项之和可以代表原序列。EMD 分解的主要目的是进一步作 Hilbert 变换得到 Hilbert 谱,它能反映原动力系统的真实物理过程。近年来,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、地球科学、天文学和工程技术等领域中得到了初步应用^[13~19]。

在 EMD 和现有预测方法的基础上,本文构建了一个新的预测模型,其核心是从观测序列中提取那些对气候预测有重要贡献、可预报性大且稳定的气候层次,并定义了一个特征层次,由它们组成模型的主要预测因子。新模型包含了 EMD、MGF 模型和 OSR 模型,其中 MGF 和 OSR 已发展成熟,应用较广泛^[20,21]。所使用的资料为扬州 530 a (1470~1999 年)旱涝级别(DWI)序列^[22~24]和北半球 1995 a (1~1995 年)树木年轮(treering)序列^[25],均为典型气候序列,它们的质量好,可靠性高,相关研究及结论多,是理想的试验资料。

2 旱涝序列的本征模函数

EMD 分解主要步骤如下(详细原理及分解过程可参阅文献[1]):

假如一时间序列 $X(t)$ 的极大值(或极小值数目)比过零点数目多 2 个以上(包括 2 个),该时间序列就是非平稳的,则需进行平稳化处理。方法是:找出 $X(t)$ 所有极大值点并将其用三次样条函数拟合成原始数据的上包络线;找出 $X(t)$ 所有极小值点拟合成下包络线;上、下包络线的均值为原序列的平均包络线 $m_1(t)$;将原序列 $X(t)$ 减去该平均包络线;即得到一个去掉低频的新序列 $h_1(t)$

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

通常, $h_1(t)$ 仍然不是平稳的,需要重复上述过程 k 次,直到所得平均包络线趋于 0 为止,这样就得到了第 1 个 IMF 分量 $C_1(t)$

$$C_1(t) = h_{1k}(t) - m_{1k}(t) \quad (2)$$

第 1 个 IMF 代表原序列的高频成分,将原始序列 $X(t)$ 减去第 1 个 IMF 分量 $C_1(t)$,可以得到第 1 个去掉高频成分的差值序列 $r_1(t)$ 。对 $r_1(t)$ 进行上述平稳过程就可以得到第 2 个 IMF 分量,直到不能分解为止,那么 $r_n(t)$ 就代表原序列的平均趋势

$$\begin{aligned} r_2(t) &= r_1(t) - C_2(t) \\ &\vdots \\ r_n(t) &= r_{n-1}(t) - C_n(t) \end{aligned} \quad (3)$$

而原始序列就可以由这些 IMF 分量及一个平均趋势序列之和表示

$$X(t) = \sum_{j=1}^{n-1} C_j(t) + r_n(t) \quad (4)$$

Huang 指出,每个 IMF 分量可能对应某一物理背景,IMF 通过 Hilbert 变换得到 Hilbert 谱,谱结构的特征就可进一步揭示隐藏在时间序列中特定的物理过程。因此,找出气候序列中的关键 IMF 分量有重要的气候预测意义,本文定义该分量为特征 IMF 分量或特征气候层次。根据已有的气候诊断^[26,27],本文用 EMD 方法将扬州 1470~1999 年(图 1a)旱涝级别序列分成 6 个层次(图 1b~g)。IMF 的小波分析表明,该旱涝级别序列的层次包含了 10 a 以下周期(IMF1)、10 a 左右周期(IMF2)、20~40 a 周期(IMF3)、70 a 左右周期(IMF4)、100 a 周期(IMF5)及 100 a 以上周期(IMF6)。如图 2 中 IMF3 的小波系数清楚地揭示了准 30 a 周期振荡,该层次是一个关键的预测因子。

3 基于本征模函数的气候预测模型构建

该预测模型的样本序列主要由 IMF 分量组成。首先利用 EMD 对气候序列进行平稳化处理,得到一组平稳分量 IMF;然后利用 MGF 模型分别对各 IMF 分量作多步预测,最后以部分预测值作为新样本作最优子集回归 OSR 拟合建立预测方程,实现预测,具体流程如图 3 所示。

分别截取两段旱涝级别序列及相应的 IMF 作为试验对象(共 7 个序列),第 1 组取法为:1770~1869 年(长度为 100 a)作为 MGF 初次预测的样本,1870~1899 年保留为被预测区间(长度 30 a,用于

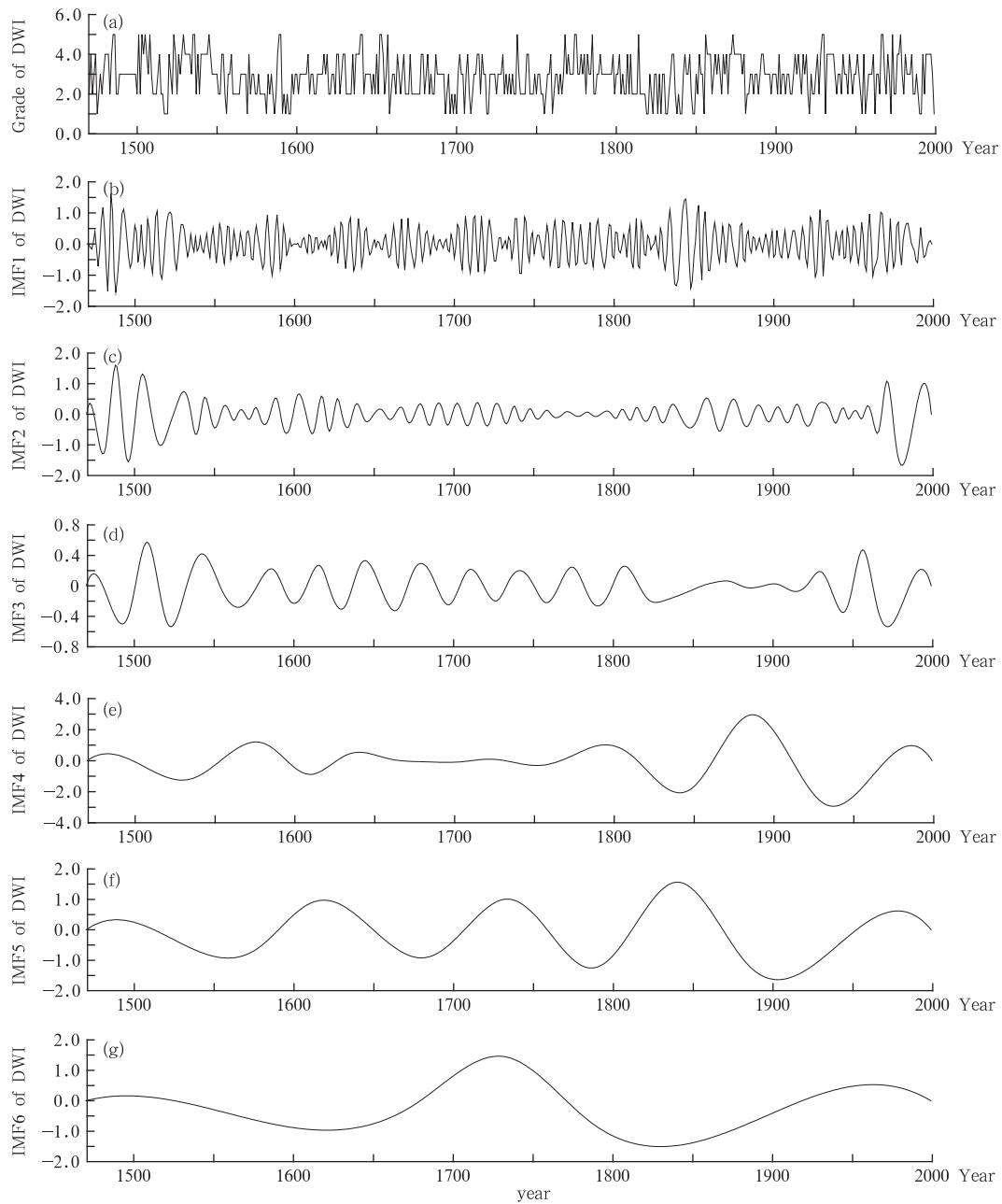


图 1 扬州 530 a 旱涝级别(a)及其 6 个 IMF(b~g)
 Fig. 1 DWI of Yangzhou (a) and its IMF1—IMF6 (b—g)

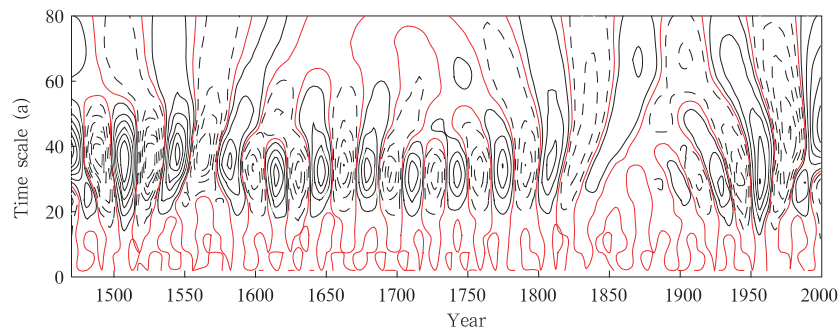


图 2 IMF3 的小波系数
 Fig. 2 The wavelet coefficient of DWI's IMF3

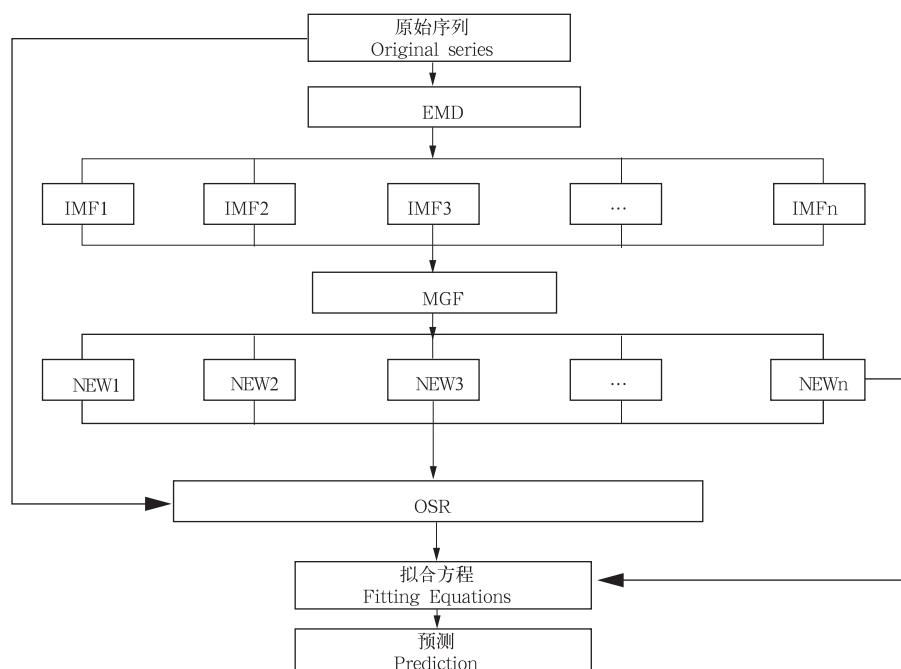


图3 基于 EMD 方法的非线性/非平稳时间序列预测模型

Fig. 3 The structure of prediction model based on nonlinear/non-stationary time series and EMD method

获取 OSR 拟合预测所需的样本);第 2 组相应地取为 1870~1959 年(长度为 90 a)和 1960~1999 年(长度 40 a)。用 MGF 模型对每组样本作初次预测,获得长度为 30 和 40 的两组新序列,以它们作为 OSR 模型的拟合因子和预测因子,最后预测可通过两种方案实现。

3.1 方案 1:直接拟合与预测

这一方案的预测长度为 5 步,样本长度分别为 25 和 35。以第 1 组样本为例,方案 1 是以前 25 个值拟合原始序列,得到一个固定的预测方程,后 5 个值代入方程实现真实预测。由于 OSR 模型能作因子最优拟合,因此可在以下 3 种情况下进行预测试验:(1)单个 IMF 分量作 OSR 拟合预测(用 IMF1—IMF6 表示);(2)所有 IMF 分量联合作 OSR 拟合预测(以 IMF_s 表示);(3)原始序列(用 DWI 表示)作 OSR 拟合预测。最后以拟合原序列的均方根误差最小、相关系数最大这一 IMF 分量作为旱涝级别序列的特征层次(特征 IMF 分量)。

表 1 列出了旱涝级别序列的两组新样本在方案 1 中的拟合和预测结果。根据表 1 第 1 组的拟合效果,可确定 IMF3 为特征层次,它是周期为 20~40 a 的平稳分量。作为单个预测因子,它拟合原序列的均方根误差最小(RMSE 为 0.81),相关系数最大(R

$=0.51$),并通过了 0.01 的置信度检验。因此 IMF3 是旱涝级别序列 1770~1899 年最稳定的气候层次和关键预测因子,可表征该时期旱涝演变趋势;作为多因子的 6 个 IMF 的联合拟合(IMF_s)效果显著提高,其均方根误差达到了最小(RMSE 为 0.66),相关系数最高($R=0.71$),通过了 0.001 的置信度检验。同时还可以看出,原序列预测(DWI)的均方根误差 RMSE 为 0.94,相关系数只有 0.05,后者没有通过置信度检验;特征分量 IMF3 的预测(图 4a)或 IMF 的联合预测(图 4a 中 IMF_s)均比原序列的预测(图 4a 中 DWI)优越,前两者所预测的趋势与实际旱涝演化趋势(1895~1900 年)较一致,而后的拟合和预测均趋于一个常数,基本没有拟合能力,预测值则更不能反映旱涝发展的真实演化特征。可见该模型相对传统方法而言,其拟合、预测能力均有明显提高。

第 2 组与第 1 组略有不同,它所包含的 IMF 分量取自 EMD 分解值的末端,Huang 指出这部分 IMF 会有一定程度的失真。但无论如何,事实上它对新预测模型有更实际的意义。为了降低边界效应,所取样本长度已减少了 10 a,并增加 10 a 的初次预测步数,前 35 个值拟合原序列得到预测方程,后 5 个值代入方程实现预测,计算结果列在表 1 第

表 1 方案 1 对旱涝序列的直接拟合结果

Table 1 The direct fitting results of DWI based on the Scheme 1

变量名	变量个数	IMF 周期 (a)	第 1 组(长度为 25,1870~1894 年)			第 2 组(长度为 35,1960~1994 年)		
			RMSE	相关系数	<i>t</i> -检验	RMSE	相关系数	<i>t</i> -检验
IMF1	1	<10	0.94	0.04		1.18	0.05	
IMF2	1	10	0.89	0.33		1.10	0.35	0.05
IMF3	1	30~40	0.81	0.51	0.010	1.17	0.16	
IMF4	1	70	0.85	0.43	0.050	1.18	0.05	
IMF5	1	100	0.92	0.17		1.18	0.11	
IMF6	1	>100	0.88	0.33		1.17	0.14	
IMFs*	6		0.66	0.71	0.001	1.04	0.49	0.01
DWI	1		0.94	0.05		1.17	0.16	

* 表示多个 IMF 的联合拟合; ** 空白处表示相关系数未通过信度检验。

2 组。与第 1 组相比,该组各分量拟合能力虽有所下降,但分量间的拟合预测能力的差异是类似的,如 IMF 的均方根误差为 1.04,相关系数达到了 0.49,并通过 0.01 的置信度检验,所预测 1995~1999 年的趋势(图 4b 中 IMF 的)与实况较吻合;IMF2 类似第 1 组的 IMF3,代表 1960~1999 年的主要趋势和

振荡周期,相关系数通过了 0.05 的置信度检验,可作为该时段的特征层次;而具有 30~40 a 周期的 IMF3 可充当次特征分量,也基本预测出了相同时期的旱涝趋势(图略);同样,DWI 根本未能有效拟合原序列,预测已无任何意义。

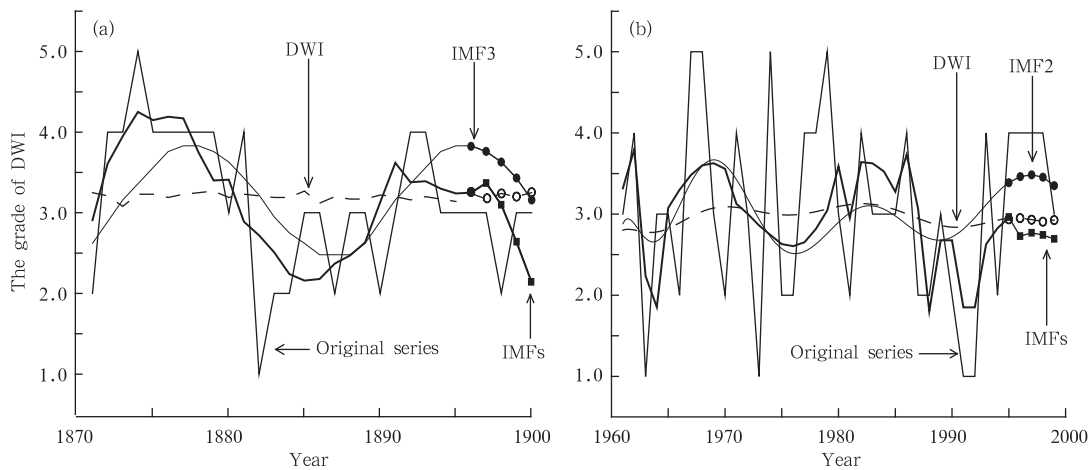


图 4 旱涝级别序列在方案 1 下的直接拟合与 5 步预测

(1870~1894,1960~1994 年为拟合曲线,1895~1899,1995~1999 年为预测曲线)

Fig. 4 The curves of direct fitting (1870—1894, 1960—1994) and 5-step prediction (1895—1899, 1995—1999) of DWI based on Scheme 1

3.2 方案 2:逐步拟合与预测

第 1 种预测方案的有效预测步数很有限,方案 1 仅预测了 5 步。因此考虑第 2 种方案,预测长度增加 1 倍(10 a)。相应地,分别取 20 个值和 30 个值作为初始拟合样本,以第 1 组为例,以前 20 个值拟合原始序列,建立方程后只预测一步,得到 1 个预测值,假设它为真实值添加到被拟合的原序列后部,长度延长到 21,再次建立新方程,继续预测下一个值,添加到原序列中,依次类推,直到完成所有预测。由于每次只预测一步,原则上能够提高预测长度。

这里重点分析 IMF2,IMF3,IMFs 和 DWI 的拟合及预测情况(图 5)。由均方根误差和相关系数可见,由于在被拟合样本中假设了一部分预测值为真实值,因此随着被拟合序列长度的增大,逐步拟合能力相应下降,最明显的是相关系数逐渐变小,均方根误差作小幅度增长。就第 1 组的特征分量 IMF3 而言,RMSE 略有增长,相关系数则由 0.55 逐步降到了 0.40,但仍通过了 0.05 的置信度检验;相应 IMFs 的相关系数由 0.73 降到了 0.62,通过了 0.01 的置信度检验。尽管如此,它们预测 10a 的效果是

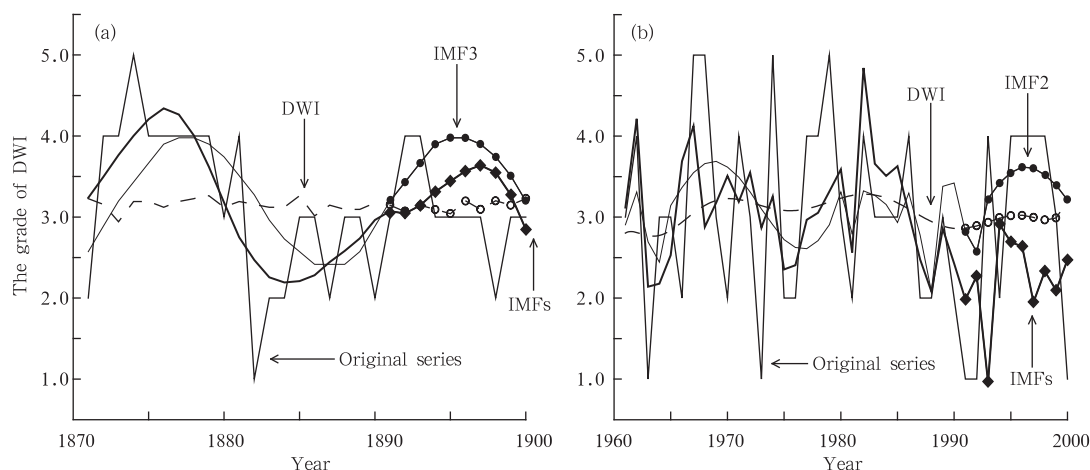


图 5 旱涝级别序列在方案 2 中的逐步拟合及 10 步预测

Fig. 5 The curves of stepwise fitting and 10-step prediction of DWI based on Scheme 2

理想的(图 5a),尤其 IMF3 的预测效果并没受到步数延长的显著影响,1891~1900 年先升后降的旱涝趋势在 IMF3 的预测值中得到了较好体现。

第 2 组的特征分量 IMF2 的拟合能力有所增强,相关系数由 0.27 上升到了 0.38,通过了 0.05 的信度检验,RMSE 变化不大;IMFs 与第 1 组情况类似,相关系数由 0.43 下降到了 0.31,通过了 0.1 的置信度检验。从 10 步预测结果(图 5b)可见,特征层次 IMF2 和多因子的 IMFs 预测的效果较理想。另外,两组中代表原始序列预测的 DWI 均没有理想的拟合及预测效果。

可见,就有效预测步数而言,方案 2 比方案 1 要略胜一筹,就有效拟合精度而言,后者略占优势。两种方案各有优点,均可提高气候原序列的预测能力。

3.3 北半球树木年轮序列的预测试验

仅仅一个气候序列的试验也许还不足以证明该

模型的稳定性,因此,下面以北半球树木年轮序列(图略)为例作类似分析预测,进一步验证模型的可靠性。相应两组样本的取法为:用 MFG 模型,(1)以 1591~1690 年为样本获得 1691~1730 年(长度为 40 a)的初次预测值作为第 1 组新样本;(2)以 1846~1945 年为样本获得 1946~1995 年(长度为 50 a)的初次预测值作为第 2 组新样本。类似地也在两种方案的 3 种情况下进行拟合和预测试验。

方案 1 的两组试验结果见表 2 和图 6,其中 treering 代表原始序列预测(下同)。根据相同原则,由表 2 确定了北半球树木年轮序列第 1 区间特征分量为 IMF3,其周期为 210 a 左右;第 2 区间特征分量为 IMF4,周期 260 a 左右。特征层次 IMF3 和 IMF4 拟合原序列的相关系数分别达到了 0.61 和 0.44,通过了 0.001 和 0.01 的置信度检验,比旱涝序列的特征层次 IMF2,IMF3 略好,其原因可能是

表 2 方案 1 对树木年轮序列的直接拟合结果

Table 2 The direct fitting results of treering based on the Scheme 1

变量名	变量个数	IMF 周期 (a)	第 1 组(长度为 35, 1691~1725)			第 2 组(长度为 45, 1946~1990)		
			RMSE	相关系数	t-检验	RMSE	相关系数	t-检验
IMF1	1	100	0.41	0.21		0.25	0.30	0.050
IMF2	1	170	0.36	0.50	0.010	0.25	0.26	0.10
IMF3	1	210	0.33	0.60	0.001	0.24	0.38	0.010
IMF4	1	260	0.36	0.49	0.010	0.23	0.44	0.010
IMF5	1	340	0.39	0.40	0.020	0.26	0.20	
IMF6	1	450	0.41	0.12		0.25	0.25	0.10
IMF7	1	800	0.36	0.50	0.010	0.26	0.21	
IMFs*	7		0.19	0.92	0.001	0.16	0.80	0.001
treering	1		0.41	0.23		0.25	0.31	0.100

* 表示多个 IMF 的联合拟合; ** 空白处表示相关系数未通过信度检验。

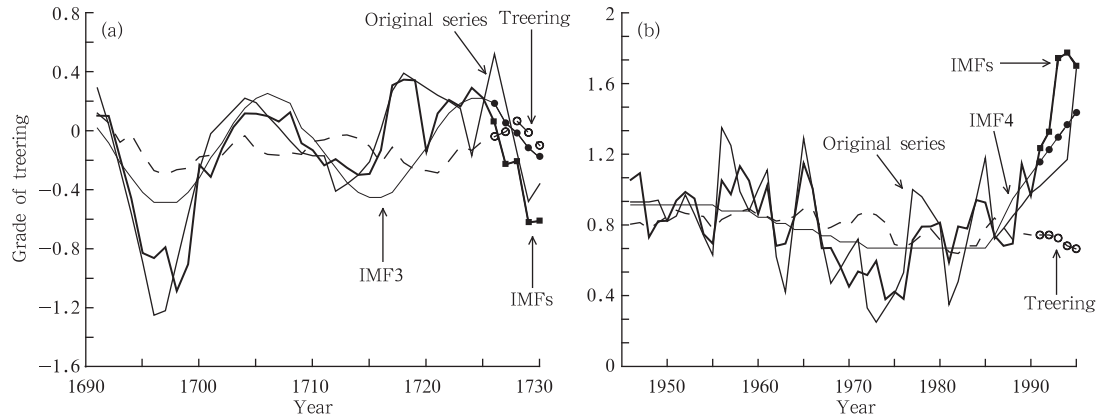


图 6 树木年轮序列在方案 1 中的直接拟合与 5 步预测
(1691~1725, 1946~1990 年为拟合曲线, 1726~1730, 1991~1995 年为预测曲线)
Fig. 6 The curves of direct fitting (1691—1725, 1946—1990) and 5-step prediction (1726—1930, 1991—1995) of treering based on Scheme 1

旱涝级别受到人为定级的影响; 同样, 两组 IMF 的拟合能力也要高于旱涝级别序列的相应情况, 相关系数达到了 0.92 和 0.80, 均通过了 0.001 的置信度检验, RMSE 只有 0.16~0.19, 拟合程度非常高; 由预测结果 (图 6a, b) 易见, 不管是特征 IMF3, IMF4, 还是 IMF5 的预测均与实况变化 (1721~1730 年的下降或 1986~1995 年的上升) 趋势相当吻合; 而代表原始序列的 treering 拟合能力与前两者相去甚远, 基本失去了预测能力。

方案 2 的两组试验结果 (图 7) 与方案 1 相比,

虽然特征层次 IMF3, IMF4 的有效预测步数延长到了 10 步, 但拟合和预测能力不但没有显著下降, 而且还略有提高, 其相关系数由 0.58 上升到了 0.61 和由 0.44 上升到了 0.51, 均通过了 0.001 的置信度检验, 再次证明特征层次的稳定性和重要性; 而作为多因子联合的 IMF5 拟合能力是逐步减小的, 相关系数由 0.93 下降到了 0.81 和由 0.81 下降到 0.56, 仍都通过了 0.001 的置信度检验, 保持了相当理想的拟合效果和预测能力。

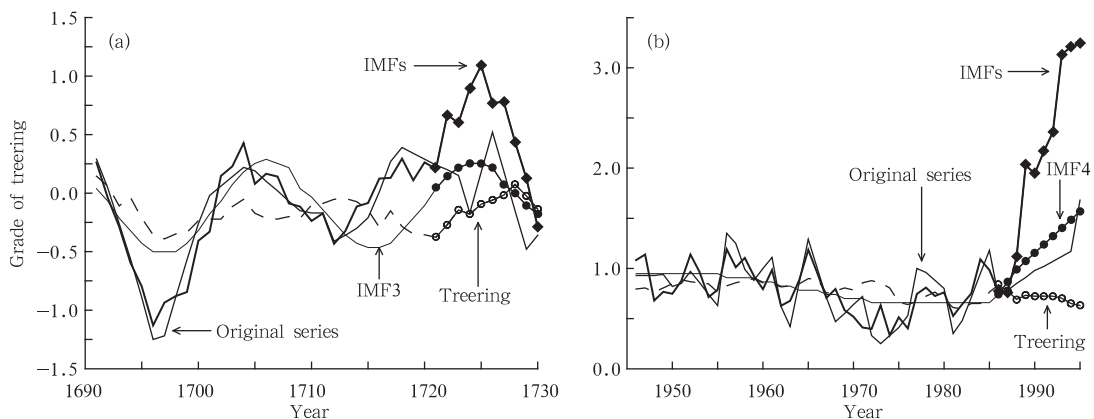


图 7 树木年轮序列在方案 2 中的逐步拟合与 10 步预测
Fig. 7 The curves of the stepwise fitting and 10-step prediction of treering based on Scheme 2

3.4 气候序列的特征层次

以上试验表明, 对于包含多层次的气候序列, 利用其特征 IMF 分量作趋势预测是行之有效的。一般, 高频组分即低层次 (如 IMF1) 的拟合能力较低,

低频分量即高层次 (IMF5, IMF6) 的拟合效果较好, 而中间层次 (IMF2~IMF4) 的拟合效果最佳, 也就是说, 特征层次服从正态分布, 即高层次代表长周期和大尺度变化, 低层次代表较短期的气候甚至天气

尺度的变化。特征层次主要集中在正态分布的大概率区,其周期一般不处于极大或极小值区。因此,特征 IMF 分量在不同时期略有差异,是由于它在一定层次范围内(即类似于正态分布曲线的峰值区)调整并适应系统发展的结果,这种动态变化符合气候系统是复杂、非线性发展的特点。就扬州 530 a 旱涝级别序列而言,其特征层次是振荡周期为 20~40 a 的 IMF 分量,即在 IMF2 和 IMF3 之间变动,两者的周期相毗邻,预测能力相当,能表征气候系统发生旱、涝的一般趋势。另外,每个序列的第 2 组样本与第 1 组样本的预测效果存在一定差距,其根本原因是 EMD 在分解原序列的末端时有部分误差,导致 IMF 分量的精度有所下降。随着 EMD 方法的不断改进,由分解误差所致的预测能力下降问题将得以很好解决。

4 模型的预测性能分析

在新预测模型的两种预测方案的基础上,作了两个典型气候序列的 8 组预测试验,结果表明该模型预测能力相当稳定且有较高的预报技巧。

模型的第 1 种方案延续了传统直接代入固定拟合方程实现预测的方式,由于方程相对稳定,当预测的步数较少时,结果是可靠的。第 2 种方案突破了方程是固定的限制,以预测值作为其将来值的被拟合因子,动态地建立预测方程,实现了“逐步逼近”的预测,并在保证预测稳定性的前提下,一定程度上提高了预测的有效长度,尤其是气候特征层次,其固有的“不变”性更适合在第 2 种方案中发挥预测作用。

5 结 论

针对气候系统的非线性/非平稳性及现有预测方法的缺陷,本文提出了一个新的预测模型,其物理基础是气候层次理论,数学基础是非线性/非平稳序列的处理手段 EMD 方法。该预测模型首先利用 EMD 对气候序列进行平稳化处理,得到一组平稳分量 IMF;然后用 MGF 模型分别对各 IMF 分量作多步预测,最后以部分预测值作为新样本进行 OSR 拟合,并通过两种方案即直接拟合和逐步拟合建立预测方程,两者在预测的精度和长度方面各有所长。两条典型气候序列(即 530 a 旱涝级别序列和 1995 a 北半球树木年轮序列)利用该模型的预测试验表明,经平稳化所得的特征 IMF 分量,即气候序列的特征层次比原序列有更稳定、更高的可预报性;当拟合方

程中的预测因子(IMF 分量)越多时,模型的预测能力越好。从某种程度上来说,特征层次有更多的预报技巧,由于其固有的“不变”性,使得它在较长期气候预测方面有独特的优势,可代表原序列作气候趋势预测或评估。该模型能有效提高现有气候预测方法的预测能力,丰富了气候系统的层次性理论,并可在其他相关研究领域得到推广及应用。

参考文献

- [1] 穆穆, 李建平, 丑纪范等. 气候系统可预报性理论研究. 气候与环境研究, 2002, 7(2): 227~235
Mu Mu, Li Jianping, Chou Jifan, et al. Theoretical research on the predictability of climate system. Climatic Environmental Res (in Chinese), 2002, 7(2): 227~235
- [2] 林振山. 气候层次理论及其应用. 北京大学学报, 1990, 26: 355~360
Lin Zhenshan. Climatic hierarchal theory and its applications. Acta Scientiarum Naturalium Universitatis Pekinensis (in Chinese), 1990, 26: 355~360
- [3] 李建平, 丑纪范. 气候系统全局分析理论及应用. 科学通报, 2003, 48(7): 703~707
Li Jianping, Chou Jifan. The global analysis theory of climate system and its applications. Chinese Sci Bull, 2003, 48(10): 1034~1039
- [4] 刘式达, 荣平平, 陈炯. 气候序列的层次结构. 气象学报, 2000, 58(1): 110~114
Liu Shida, Rong Pingping, Chen Jiong. The hierarchical structure of climate series. Acta Meteor Sinica (in Chinese), 2000, 58(1): 110~114
- [5] 刘式达, 傅遵涛, 刘式适等. 非线性波动方程的 Jacobi 椭圆函数包络周期解. 物理学报, 2002, 51(4): 718~722
Liu Shida, Fu Zuntao, Liu Shikuo, et al. The envelope periodic solutions to nonlinear wave equations with Jacobi Elliptic function. Acta Physica Sinica (in Chinese), 2002, 51(4): 718~722
- [6] 封国林, 戴新刚, 王爱慧等. 混沌系统中可预报性的研究. 物理学报, 2001, 50(4): 606~611
Feng Guolin, Dai Xingang, Wang Aihui, et al. On numerical predictability in the chaos system. Acta Phys Sinica (in Chinese), 2001, 50(4): 606~611
- [7] 刘式适, 陈华, 付遵涛等. Lamé 函数和非线性演化方程的多级准确解的不变性. 物理学报, 2003, 52(8): 1842~1847
Liu Shikuo, Chen Hua, Fu Zuntao, et al. Lamé function and invariants of multi-order exact solutions among nonlinear evolution equations. Acta Phys Sinica (in Chinese), 2003, 52(8): 1842~1847.
- [8] 莫嘉琪, 林万涛. ENSO 非线性模型的摄动解. 物理学报, 2004, 53(4): 996~998

- Mo Jiaqi, Lin Wantao. Perturbed solution for the ENSO non-linear model. *Acta Phys Sinica* (in Chinese), 2004, 53(4): 996—998
- [9] Feng Shida, Ren Rongcai, Cui Xiaopeng, et al. Corrections to collision term in BGK Boltzmann equation. *Chinese Physics*, 2001, 10(12): 1106
- [10] Huang N E, Shen Z. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society London*, 1998, A, 454: 903—995
- [11] Huang N E, Shen Z, Long S R. A new view of nonlinear water waves: the Hilbert spectrum. *Ann Rev Fluid Mech*, 1999, 31: 417—457
- [12] Wei Huang, Zheng Shen, Huang N E, et al. Engineering analysis of biological variables: An example of blood pressure over 1 day. *Proc Natl Acad Sci*, 1998, 95(9): 4816—4821
- [13] Coughlin K T, Tung K K. 11-year solar cycle in the stratosphere extracted by the empirical mode decomposition method. *Advances in Space Res*, 2004, 34: 323—329
- [14] 林振山, 汪曙光. 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*, 2004, 20(1): 90—96
Lin Zhenshan, Wang Shuguang. EMD analysis of northern hemisphere temperature variability during last 4 centuries. *J Tropical Meteor* (in Chinese), 2004, 20(1): 90—96
- [15] 杨宇, 于德介, 程军圣等. 经验模态分解(EMD)在滚动轴承故障诊断中的应用. *湖南大学学报(自然科学版)*, 2003, 30(5): 25—31
Yang Yu, Yu Dejie, Cheng Junsheng, et al. Application of empirical mode decomposition (EMD) in roller bearing fault diagnosis. *J Hunan University (Natural Sciences)* (in Chinese), 2003, 30(5): 25—31
- [16] 赵犁丰, 张晓亮, 宋洁. 利用 EMD 方法和小波变换进行信号奇异性检测. *青岛海洋大学学报(自然科学版)*, 2003, 33(5): 759—763
Zhao Lifeng, Zhang Xiaoliang, Song Jie. Application of EMD and wavelet transform to the detection of signal singularity. *Journal of Ocean University of Qingdao(Natural Sciences)* (in Chinese), 2003, 33(5): 759—763
- [17] 胡劲松, 杨世锡, 周方洁等. 旋转机械振动信号基于 EMD 的 HT 和 Winger 分布时频分析比较. *汽轮机技术*, 2003, 45(5): 308—314
Hu Jinsong, Yang Shixi, Zhong Fangjie, et al. The comparison of vibration signals time-frequency analysis between EMD based HT and winger distribution method in rotation machinery. *Turbine Technology* (in Chinese), 2003, 45(5): 308—314
- [18] Zhao H W, Huang N E. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings: Mathematical, Physical and Engineering Sciences*, 2004, 460(6): 1597—1611
- [19] 林振山, 鲍名. 气候多层次的趋势预报. *热带气象学报*, 2001, 17(2): 188—192
Lin Zhenshan, Bao Ming. Climatic trend prediction at multiple hierarchy. *J Tropical Meteor* (in Chinese), 2001, 17(2): 188—192
- [20] 魏凤英, 曹鸿兴. 长期预测的数学模型及其应用. 北京: 气象出版社, 1990. 175pp
Wei Fengying, Cao Hongxing. *Mathematical Model and Its Application for Long-term Forecast*. Beijing: China Meteorological Press, 1990. 175pp
- [21] 封国林, 曹鸿兴, 魏凤英等. 长江三角洲汛期预报模式的研究及其初步应用. *气象学报*, 2001, 59(2): 206—212
Feng Guolin, Cao Hongxing, Wei Fengying, et al. On area rainfall ensemble prediction and its application. *Acta Meteor Sinica* (in Chinese), 2001, 59(2): 206—212
- [22] 中央气象局气象科学研究所. 中国近五百年旱涝分布图集. 北京: 地图出版社, 1981. 332pp
Meteorology Institute of Central Meteorological Bureau. *Yearly Charts of Dryness/Wetness in China for the Last 500-year Period*. Beijing: Cartographic Publishing House, 1981. 332pp
- [23] 张德二, 刘传志. 《中国近五百年旱涝分布图集》续补(1980—1992 年). *气象*, 1993, 19(11): 41—46
Zhang Deer, Liu Chuanzhi. Continuation(1980—1992) of the “Yearly Charts of Dryness/Wetness in China for the last 500 Years Period”. *Meteorological Monthly*, 1993, 19(11): 41—46
- [24] 张德二, 李小泉, 梁有叶. 《中国近五百年旱涝分布图集》的再续补(1993—2000 年). *应用气象学报*, 2003, 14(3): 379—382
Zhang Deer, Li Xiaoquan, Liang Youye. Re-continuation (1993—2000) of the “Yearly Charts of Dryness/Wetness in China for the Last 500 Years Period”. *Journal of Applied Meteorological Science* (in Chinese), 2003, 14(3): 379—382
- [25] Keith R, Briffa. Annual climate variability in the Holocene: Interpreting the message of ancient trees. *Quat Sci Rev*, 2000, 19(1—5): 87—105
- [26] Zhu Jinghong, Wang Saowu. 80a-oscillation of summer rainfall over the east part of China and East-Asian summer monsoon. *Adv Atmos Sci*, 2001, 18(5): 1043—1051
- [27] Zhu Yimin, Yang Xiuqun. Joint propagating patterns of SST and SLP anomalies in the North Pacific on bidecadal and pentadecadal timescales. *Adv Atmos Sci*, 2003, 20(5): 694—710

EXTRACTING USEFUL INFORMATION FROM THE OBSERVATIONS FOR THE PREDICTION BASED ON EMD METHOD

Wan Shiquan

*(Yangzhou Meteorological Office, Yangzhou 225000; Laboratory for Climate Studies,
National Climate Center, China Meteorological Administration, Beijing 100081)*

Feng Guolin

(Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100081)

Zhou Guohua Wan Bingmei Qin Mingrong Xu Xiazhen

(Jiangsu Meteorological Bureau, Nanjing 225000)

Abstract

At present, the main proach to the climate prediction in practice such as the short-term and long-term climate prediction is using the statistical method to predict the climate in month scale, season scale and annual scale, respectively. Now the numerical model is capable to forecast weather up to 7 days, but there are numerous difficulties in realizing the short-climate prediction by numerical integration due to the nonlinear/non-stationary effects of the climate system. In fact, most of the present statistical climate prediction methods (mainly includes empirical, mathematical and physical statistics methods) are based on the hypothesis that the system is stationary. However, the observations, in particular for the climate data, are often nonlinear/non-stationary and multi-hierarchical, which makes the prediction very difficult. Aiming at this problem, a new prediction model is introduced, in which, firstly, using the empirical mode decomposition the observation sequence are stationarized and a variety of intrinsic mode functions (IMF) are obtained; secondly the IMFs are predicted by the mean generating function model separately; finally with the optimal subset regression model the part of predictions are used as new samples to fit the original series directly or step by step and a system of prediction equations are set up. The climate sequences prediction research shows that the individual IMF, especially the eigen-IMF, has more stable predictability than that of its sources. The trend of development in climate prediction lies in researching the mechanism and hierarchy of the climate system, constructing the corresponding climate prediction model. An attempt has been accomplished in this paper. It is believed that the model proposed can open up a new effective way for the climate prediction or evaluation.

Key words: Empirical mode decomposition, Nonlinear/non-stationary time series, Hierarchy theory, Climate prediction.