

胡凤良, 王丽琼, 左瑞亭, 等. 2017. 几种统计预测方法对 1998 年南京降水的跨季节预测试验及分析 [J]. 气候与环境研究, 22 (1): 23–34. Hu Fengliang, Wang Liqiong, Zuo Ruiting, et al. 2017. Extra-seasonal predicting tests and analyses of several statistical forecasting methods on precipitation over Nanjing in 1998 [J]. Climatic and Environmental Research (in Chinese), 22 (1): 23–34, doi:10.3878/j.issn.1006-9585.2016.15251.

几种统计预测方法对 1998 年南京降水的跨季节预测试验及分析

胡凤良¹ 王丽琼¹ 左瑞亭¹ 张舰齐²

¹ 解放军理工大学气象海洋学院, 南京 211101

² 中国人民解放军 95871 部队, 衡阳 421002

摘要 对 1998 年南京降水分别设计并开展了求和自回归滑动平均 (Auto-Regressive Integrated Moving Average, ARIMA) 模型预测、经验模态分解 (Empirical Mode Decomposition, EMD) 预测和基于 Hilbert 变换 (Hilbert Transformation, HT) 的幅频分离预测等 3 种跨季节统计预测试验。结果表明: ARIMA 模型预测结果存在明显的系统性误差且对夏季的降水突变现象预测困难; EMD 分解预测的结果虽在降水演变趋势上有明显提高, 但仍未能预测出夏季的强降水突变现象, 究其原因可能是对高频分量预测效果不好所致; 而基于 Hilbert 变换的幅频分离预测方法能够对各模态分量的瞬时频率和瞬时振幅实施隔离预测, 消除两者的相互影响, 显著改善高频模态的预测效果, 使得最终预测结果最为理想, 不仅具有最高的趋势相关性和最小的偏差, 而且还较好地预测出了夏季两次强降水过程。不仅如此, 在对 2003 年的降水预测验证中, 基于 Hilbert 变换的幅频分离预测方法同样具有最好的预测效果, 表明该方法预测效果较为稳定, 为改进跨季节短期气候统计预测技术提供了一个新思路。

关键词 短期气候预测 求和自回归滑动平均 (ARIMA) 经验模态分解 (EMD) Hilbert 变换 (HT) 最小二乘支持向量机

文章编号 1006-9585 (2017) 01-0023-12

中图分类号 P468.0⁺24

文献标识码 A

doi:10.3878/j.issn.1006-9585.2016.15251

Extra-seasonal Predicting Tests and Analyses of Several Statistical Forecasting Methods on Precipitation over Nanjing in 1998

HU Fengliang¹, WANG Liqiong¹, ZUO Ruiting¹, and ZHANG Jianqi²

¹ Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101

² 95871 Troop, People's Liberation Army of China, Hengyang 421002

Abstract Three statistical forecasting methods, i.e. ARIMA (Auto-Regressive Integrated Moving Average) model prediction, EMD (Empirical Mode Decomposition) decomposition prediction, and isolated prediction of frequency and amplitude based on Hilbert transformation, are designed and employed to make extra-seasonal prediction tests on the precipitation over Nanjing in 1998. Results show that the ARIMA model exhibits severe system errors and is hard to reproduce the abrupt variation of precipitation. Although the EMD decomposition prediction makes an obvious improvement in the evolution trend of precipitation, it still fails in the depiction of precipitation catastrophes in the

收稿日期 2015-11-10; 网络预出版日期 2016-05-23

作者简介 胡凤良, 男, 1993 年出生, 硕士研究生, 主要从事气候数值模拟研究。E-mail: 15895904544@163.com

通讯作者 左瑞亭, E-mail: ratinzuo@126.com

资助项目 国家自然科学基金项目 41490642、41475071

Funded by National Natural Science Foundation of China (NSFC, Grants 41490642 and 41475071)

summer due to its incapability of predicting high frequency modes. The isolated prediction method improves the prediction of high frequency modes since it can separately predict the frequency and amplitude of each mode and their interactions are avoided. Thereby the isolated prediction method gives a pretty good final prediction with the highest trend correlation and the smallest deviation. The two precipitation catastrophes in the summer of 1998 are realistically predicted. Additionally, a further verification of the precipitation prediction for 2003 also indicates that the isolated prediction method performs best among the three methods proposed in this study. The above results suggest that the isolated prediction method may provide a new idea for the technological improvement on extra-seasonal short-term climate prediction.

Keywords Short-term climate prediction, Auto-Regressive Integrated Moving Average (ARIMA), Empirical mode decomposition, Hilbert transformation, Least square support vector machine

1 引言

1998 年夏季我国遭遇了显著的洪涝灾害, 长江、嫩江、松花江等流域均发生了不同程度的暴雨洪水, 尤其是长江流域遭受了自 1954 年以来的又一次全流域性大洪水 (孙继昌等, 2004), 普遍连降大到暴雨, 其中南京地区 6 月下旬至 7 月初和 7 月底出现两次持续时间较长的大到暴雨, 属于显著的降水突变现象。鉴于此, 跨季节的汛期预测, 尤其是气候突变的预测显得尤为重要。气候系统是一个高度非线性、多层次的系统, 虽然各种数值模式也在不断尝试短期气候预测的试验甚至是业务运用 (袁重光等, 1996; 曾庆存等, 1999; 丑纪范和徐明, 2001), 但是鉴于气候系统本身的高度非线性、非平稳性和多层次性 (穆穆等, 2002), 以及数值模式本身的不完备性和不确定性等因素 (廖洞贤, 1999), 统计预测仍然是气候预测的重要手段。相比而言, 经典的统计预测方法大多忽视了气候系统非线性、非平稳性的特点, 例如线性回归预测模型的预测精度依赖于自变量与因变量之间线性关系的显著性, 在处理非线性问题时, 往往拟合效果不好, 对未来的预测造成很大影响 (霍玉洪, 2009); 自回归滑动平均 (Auto-Regressive Moving Average, ARMA) 模型的研究对象为平稳序列, 对非平稳序列的预测效果不佳 (管鹏, 2007); 求和自回归滑动平均 (Auto-Regressive Integrated Moving Average, ARIMA) 模型虽然可通过差分运算将非平稳时间序列转换为平稳时间序列, 但过度的差分也会造成序列中有用信息的丢失 (张海勇等, 2002); 传统的灰色预测模型精度较高的仅是一两个数据的短期预测, 依靠任意存在的原始数据序列来预测未来的动态变化时, 常出现精度不高的情形

(张军, 2008); 神经网络模型则存在着学习速度很慢、网络训练失败的可能性较大等问题 (管鹏, 2007)。

近年来, 不少新的统计技术相继出现并在气象领域得到广泛应用。如, Huang et al. (1998) 提出了一种新的时间序列分解技术, 即经验模态分解 (Empirical Mode Decomposition, EMD), 该方法能将一个非平稳序列分解为若干平稳序列, 从而提高其可预测性。后来, 基于该方法又出现了能够消除模态混杂现象的集合经验模态分解 (Wu and Huang, 2009) 和二维 EMD 的图像分解技术 (Linderhed, 2005)。EMD 的提出为短期气候的统计预测提供了新的思路 (万仕全等, 2005; 张明阳等, 2005; 刘莉红等, 2008), 这些预测方法往往在将原始序列分解为多个平稳 EMD 分量后, 直接对其进行回归或神经网络预测, 由于平稳性的提高, 使得可预报性得以提高, 预测效果有了一定改善。但是, Huang et al. (2001) 提出的 Hilbert 变换 (Hilbert Transformation, HT) 技术, 可以进一步提取出序列的瞬时振幅和瞬时频率, 进而将两者的相互影响降到最低。而在统计预测方面, 最小二乘支持向量机 (Least Square Support Vector Machine, LS-SVM) 能够将非线性序列映射到高维空间进行回归建模, 从而实现对非线性序列的预测 (Suykens and Vandewalle, 1999), 相对于传统的支持向量机和神经网络等预测技术具有独特的预测特性 (李丽娟, 2008), 也在气象领域有着相当广泛的应用 (刘科峰等, 2009; 纪玲玲等, 2009)。基于这些新的统计技术, 本文针对 1998 年夏季南京降水的短期气候预测问题展开预测试验, 将经典统计预测与新统计技术预测进行对比分析, 以考察新技术对统计预测的改进效果, 为短期气候序列的统计预测提供新的思路和技术借鉴。

2 资料与方法

2.1 资料

本文所用资料为中国气象数据网提供的“中国地面国际交换站气候资料日值数据集 V3.0” (<http://data.cma.gov.cn/data/detail/dataCode/A.0029.0001.html> [2015-08-03])，该数据集包含了中国 194 个站点 1981 年 1 月以来本站气压、气温、降水量、蒸发量等地面要素的日值数据，本文选用南京站降水量（单位：mm）作为研究对象，截取了 1994 年 3 月至 1998 年 9 月的降水量数据用于建模和回报试验，选取 1994 年 3 月 1 日至 1998 年 2 月 28 日的降水量作为历史样本，设置预测时段为 1998 年 3 月 1 日至 9 月 30 日，该时段的实际历史资料将作为检验对比的实况。

2.2 方法

2.2.1 求和自回归滑动平均模型预测方法

求和自回归滑动平均模型 ARIMA 是由 Box and Jenkins (1970) 提出的时间序列预测方法，该方法实质是差分运算与自回归滑动平均模型的组合。该模型可简称为 ARIMA (p, d, q)，其中， p 为自回归阶数， q 为滑动平均阶数， d 为差分次数。当非平稳序列在 d 阶差分后平稳，就可以用 ARMA 模型对平稳序列进行拟合和预测，将预测结果进行反差分运算便可得到非平稳序列的预测结果。而 ARMA 实质是自回归模型 (Auto-Regressive, AR) 与滑动平均模型 (Moving Average, MA) 的组合，该模型认为时间序列的当前值既由历史状态决定，又受历史白噪声即扰动影响，即

$$x_t = u + \sum_{i=1}^p \alpha_i x_{t-i} + \varepsilon_t - \sum_{j=1}^q \beta_j \varepsilon_{t-j}, \quad t = p+1, p+2, \dots, n \quad (1)$$

其中， x_t 为资料序列， u 、 α_i 、 β_j 为模型参数，这些未知参数需要利用样本资料序列进行估计， $\{\varepsilon_t\}$ 为独立白噪声序列， t 为时间， n 为需要预测的资料序列的长度， p 和 q 可通过最终预测误差 (Final Prediction Error, FPE) 准则、Akaike 信息准则 (Akaike Information Criterion, AIC)、贝叶斯信息准则 (Bayesian Information Criterion, BIC) 等 (胡广书, 2003) 来确定。

2.2.2 经验模态分解

经验模态分解技术 (EMD) 是美国国家宇航局

戈达德空间飞行中心提出的一种时间序列分解技术。该方法首先确定原序列 $x(t)$ 的局部极大和极小值点，然后利用 3 次样条插值得到上、下包络曲线，对上、下包络取平均得到均值序列，再将原序列与均值序列 $x(t)$ 作差得到差值序列，考察差值序列是否满足特征模态函数 (Intrinsic Mode Function, IMF) 的两个条件：其一是极大、极小和过零点的数目必须相等或至多相差一点，其二是任意时刻上下包络的均值为 0。若满足，则判定该差值序列为一个特征模态函数分量 $F_{Mi}(t)$ ；否则，继续循环求该差值序列的上下包络均值及新的差值序列并作判据检验，直到满足 IMF 条件为止。然后，用原序列减去求解出的 IMF 分量，并对剩余序列重复上述步骤，直到两个相邻特征模态函数的标准差达到一个较小值时分解结束。EMD 分解的实质是从原始资料序列中多次、逐步扣除由上、下极值包络构成的平均趋势，以此得到去除了不同尺度趋势项后多个平稳的单分量信号，即具有不同尺度的若干特征模态函数分量，从而将多尺度非平稳、多分量序列分解为如下有限个单分量平稳序列

$$x(t) = \sum_{i=1}^k F_{Mi}(t) + r_k(t), \quad (2)$$

其中， $F_{Mi}(t)$ 是第 i 个 IMF 分量， $r_k(t)$ 为最终趋势项 (residual, RES)， k 为最终分解的次数，反映了分解的收敛速度。EMD 分解得到的各 IMF 分量具有时变的瞬时频率和振幅，克服了传统傅里叶分解仅能分解为固定频率标准谐波的不足，可以更精确地刻画原序列的时变特征。

2.2.3 希尔伯特变换

希尔伯特变换 (HT) 是数学和信号处理的重要方法，是一种从时域到时域转换的转换，其实质是将连续函数 $f(t)$ 与 $h(t) = 1/\pi t$ 进行卷积，变换后函数 $\hat{f}(t)$ 为

$$\hat{f}(t) = f(t) * h(t) = \int_{-\infty}^{+\infty} f(\tau) \cdot h(t-\tau) d\tau = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{f(\tau)}{t-\tau} d\tau, \quad (3)$$

HT 变换相当于 90° 移相器，变换后函数 $\hat{f}(t)$ 与原函数 $f(t)$ 呈正交状态。上式积分取柯西主值，以剔除 $t = \tau$ 和 $t \rightarrow \pm\infty$ 时的奇点问题。

由 HT 变换可进一步求解信号的时变特征，首先构造解析信号 $Z(t)$

$$Z(t) = f(t) + i\hat{f}(t) = a(t)e^{i\theta(t)}, \quad (4)$$

其中振幅 $a(t)$ 和幅角 $\theta(t)$ 分别为

$$a(t) = \sqrt{\hat{f}^2(t) + \hat{f}'^2(t)}, \quad (5)$$

$$\theta(t) = \arctan[\hat{f}'(t) / \hat{f}(t)], \quad (6)$$

瞬时频率 $\omega(t)$ 定义为

$$\omega(t) = d\theta / dt. \quad (7)$$

2.2.4 最小二乘支持向量机

最小二乘支持向量机 (LS-SVM) 是在 Cortes and Vapnik (1995) 提出的支持向量机 (Support Vector Machine, SVM) 的基础上改进而来。该方法针对样本的非线性回归问题, 首先通过非线性映射, 把低维度非线性样本映射到一个高维乃至无穷维的特征空间, 然后应用线性学习机方法建立高维空间的线性回归模型, 以实现实际样本空间的非线性回归。特别是在建立非线性映射关系时, 采用了核函数代替标准 SVM 中的内积运算, 并引入误差平方和损失函数代替标准 SVM 中损失函数, 因此实现以等式约束代替了标准 SVM 中的不等式约束, 将二次规划问题转换为极值约束方程组的求解, 算法简单易实现, 降低了计算复杂度, 提高了实时性, 同时需确定的模型参数少, 具有更好的扩展性、灵活性和泛化能力。

由 LS-SVM 确定的非线性回归模型决策函数为

$$f(t) = \sum_{i=1}^n \lambda_i K(x_i, x) + b, \quad (8)$$

其中, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)^T$ 为 Lagrange 乘子, $K(x_i, x)$ 为核函数, b 为偏置。最小二乘支持向量机在建模过程中需要确定两个重要参数, 即正则化参数 γ 和高斯径向基核函数的宽度系数 σ , 这两个参数决定模型的性能。正则化参数 γ 也称为平衡因子, 其作用是决定模型的复杂度和对于回归偏差的惩罚力度, γ 偏小表示模型简单, 回归偏差可能过大, 易导致欠学习; γ 偏大表示模型复杂, 回归偏差惩罚力度大, 易导致过学习, 模型不易泛化。而核函数宽度系数 σ 反映了样本数据的分布特点, 精确定义了高维空间的结构, 较小的 σ 可能导致核函数过于局部化, 回归函数可能导致过学习, 模型泛化能力差, 反之回归函数对损失函数不敏感, 易造成欠学习, 导致回归预测精度差。实际应用中参数的确定可通过基于交叉检验、基于网格搜索或基于进化类算法实现。

3 预测试验及结果分析

3.1 预测试验设计

在 1998 年夏季的洪涝灾害中, 南京地区自 6 月下旬开始出现了多次大到暴雨 (如图 1a), 6 月 25 日至 7 月 5 日是降水的第一个集中期, 11 d 中有 5 d 超过大雨的标准, 其中 6 月 27 日、30 日、7 月 1 日为暴雨, 7 月 17 日降大雨, 7 月 29~31 日是降水的第二个集中期, 其中 29 日降水达到创纪录的 70.1 mm, 8 月 17 日、27 日又降大雨。图 1b 给出了对应的候降水距平, 可见 6 月底和 7 月底有两次降水异常突变, 分别对应上述两个降水集中期。

考虑到气候预测的统计平均特点, 本文的预测以候降水距平为单位, 历史样本资料为 1994 年 3 月至 1998 年 2 月共 288 候, 预测时段为 1998 年 3~9 月共 42 候。

为与经典回归预测进行对比, 本文设计了 3 个预测试验, 分别是回归预测、EMD 分解预测和基于瞬时频率和振幅的 Hilbert 预测, 具体如下:

试验 1: 回归预测试验。利用 ARIMA 模型拟合降水距平序列的历史样本资料 288 个候, 将拟合序列向前延伸 42 个候进行未来时段的预测。

试验 2: EMD 分解预测试验。先对 1994 年 3 月至 1998 年 2 月共 288 候降水距平序列进行 EMD 分解, 得到各特征模态 IMF 分量, 对各分量进行 LS-SVM 预测, 由预测的 IMF 分量重构得到降水距平的候预测结果。

试验 3: 基于 Hilbert 变换的幅频分离预测试验。该方案前期操作类似试验 2, 但对 EMD 分解得到的各 IMF 分量不进行直接预测, 而是进行 HT 变换, 提取各 IMF 分量的瞬时振幅和瞬时频率, 然后对瞬时振幅和瞬时频率进行 LS-SVM 预测, 由预测的瞬时振幅和瞬时频率重构 IMF 分量, 再由重构的 IMF 分量再次重构得到降水距平的候预测结果。

3.2 预测建模与分析

3.2.1 回归试验预测分析

经检验, 样本资料的一阶差分序列符合平稳性的要求, 可用 ARMA 模型对差分序列进行建模预测。根据 AIC 信息准则, 当 $p=10$ 、 $q=1$ 时, AIC 最小。因此最终确立的拟合模型为 ARIMA(10,1,1), 模型参数 $\alpha_p = (1.0000, 0.2044, 0.0046, 0.0004, -0.0486, -0.0589, -0.0049, -0.0538, -0.0694, -0.0935)^T$ 、

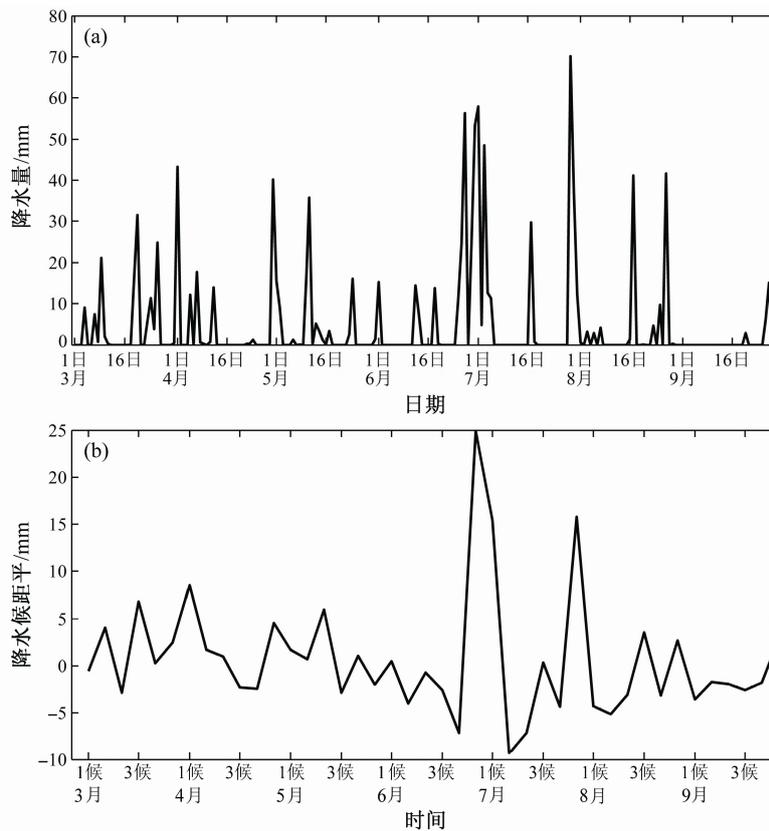


图 1 1998 年 3~9 月南京地区 (a) 逐日降水量和对应的 (b) 降水候距平
Fig. 1 (a) Daily precipitation and (b) pentad-mean precipitation anomalies in Nanjing from Mar to Sep in 1998

$\beta = -0.2415$ ，拟合结果如图 2 所示，试验 1 的拟合序列能够较好反映出实际序列的变化趋势，两者相关性为 0.12，均方根误差为 6.54。利用该模型向后延伸 42 候，便可得到预测时段结果（如图 3），试验 1 基本预测出降水候距平的趋势，两者的相关系数为 0.11，但预测的距平强度普遍偏小，大部分呈现负距平，即未预测出偏涝情况，尤其是对 6 月底和 7 月底的两次距平突变未预测出来，除系统性偏差外，这也反映出经典统计预测对突变预测能力的不足。

3.2.2 EMD 分解预测试验分析

为得到历史样本和预测时段 EMD 分解的主要特征模态的近似真值，将 1994 年 3 月至 1998 年 9 月（即历史样本与预测时段的总和）共 330 候的数据作为全序列进行 EMD 分解，可得到 7 个 IMF 模态分量和 1 个趋势项（结果如图 4）。IMF 特征模态的频率随序号依次递减，平均而言，高频分量模态的振幅较大，表示其所占的分解比重也较大，低频分量则反之。表 1 列出了各模态分量的方差贡献率，其中，前 4 个模态分量的方差贡献率达到 95.3%，

表明原始数据的波动主要与前 4 个模态分量有关，在后续的重构预测中，高频分量，尤其是 IMF1 和 IMF2 是预测的主要部分，也是影响最终预测效果的关键。

表 1 各模态分量的方差贡献率和累积贡献
Table 1 Variance contribution rates of each component and cumulative contribution

	方差贡献率	累积贡献
IMF1	43.9%	43.9%
IMF2	34.9%	78.8%
IMF3	9.5%	88.3%
IMF4	7.0%	95.3%
IMF5	1.4%	96.7%
IMF6	2.3%	99.0%
IMF7	0.1%	99.1%
RES	0.9%	100.0%

但是，由于端点效应（韩建平等，2010），这样得到的各特征模态，其前 288 候序列与实际预测时仅采用历史样本 288 候分解得到的对应时段序列，会存在一定差异（图略）。根据 EMD 分解原理，构造局部极值点的包络时，边界端点的影响会随着

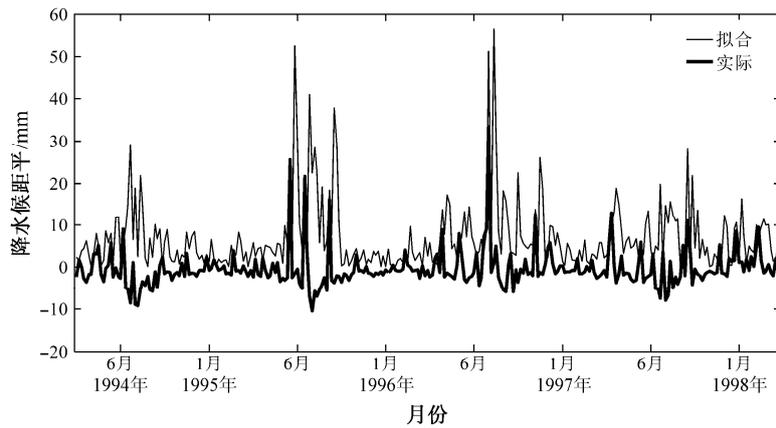


图2 ARIMA 模型对历史训练序列的拟合

Fig. 2 Fitting results of the ARIMA model to the historical training series

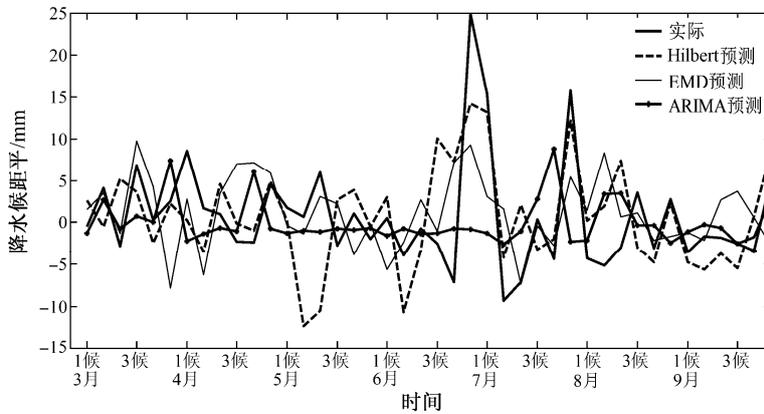


图3 1998年3~9月各预测试验降水候距平结果对比

Fig. 3 Comparison of the pentad-mean precipitation anomalies from the prediction tests and the real data for the period of Mar to Sep of 1998

逐次分解而向内区延伸，这将造成端点误差随分解层级向内扩展，导致低频分量的端点附近误差和受影响的数据点数显著大于高频分量。表2列出了两种分解结果之间的相关系数 r 、相对均方根误差 R_{RMSE} (Relative Root-Mean-Square Error, 即均方根误差与全序列分解结果之比) 和差值大于 10% 的数据点数占历史样本的比例 k_c 。

表2 全序列与历史样本序列 EMD 分解的差异对比

Table 2 Differences between the EMD decomposition results of the whole sequences and historical sample sequences

	r	R_{RMSE}	k_c
IMF1	0.9994	0.0350	1.0%
IMF2	0.9955	0.0955	5.6%
IMF3	0.9929	0.1193	4.5%
IMF4	0.9958	0.0928	8.3%
IMF5	0.9626	0.2746	17.0%
IMF6	0.9160	0.4444	26.4%
IMF7	0.8234	0.8321	99.0%
RES	0.7182	0.6410	89.2%

由表2可见，两种分解的差异随 IMF 分量的频率降低而增大，误差点从序列的右端向内区逐渐传播，各分量的相关系数均较高，平均约为 0.9255，反映出误差传播对模态分量的变化趋势影响较小，但对模态分量的偏差量影响显著，最显著的低频模态的相对均方根误差可达到 83%。总体看，前4个中高频分量受显著影响的数据点比例较低，差值大于 10% 的数据点不超过 9%，且其相关系数较高和相对均方差较小，可以近似认为两种序列分解的前4个中高频模态较为相似。基于 330 候全序列分解的该4个分量，其后续 42 候可作为基于 288 候历史样本预测的近似真值，而其他低频模态包括趋势项的分解结果仍具有部分参考价值，尤其是其合成结果参与最后重构的贡献仍然比较大，故后续分析中，这些低频分量的单独预测对比从略，其贡献通过最终的预测结果体现出来。试验2对历史样本序列 288 候降水距平进行分解，并对得到的 IMF 分量

进行 LS-SVM 预测，预测模型的正则化参数 γ 和核函数宽度参数 σ 采用 10 折交叉检验得到，根据 10 次迭代后得到的均方误差 (Mean Square Error, MSE) 平均值来估计期望泛化误差，选择一组为最优参数，预测模型各模态分量的 γ 和 σ 如表 3 所示。预测结果的前 4 个分量与其近似真值的比较如图 5 所示。

表 3 试验 2 预测模型各模态分量的 γ 和 σ 值
Table 3 γ and σ of each component of prediction model in experiment 2

	$\gamma(\times 10^5)$	σ
IMF1	1.6	130
IMF2	1.0	4000
IMF3	0.3	500
IMF4	4.5	24
IMF5	1.4	2100
IMF6	1.0	20
IMF7	1.4	2
RES	16.0	0.8

由图 5 的近似真值分量可见，前 2 个高频模态平均强度较强，约在 4~5 mm，且对应有 6 月中下旬和 7 月中下旬两次明显的强突变，而后两个模态分别反映的是 45 d 和 90 d 左右的均匀波动分量，整个降水距平的突变主要由前 2 个高频分量控制。试验 2 预测的前 4 个中高频模态能够基本刻画出各模态的演变趋势，但对强度和位相的预测存在一定偏差，特别是对模态序列中强度较大的突变现象预测不够准确，如 IMF1 和 IMF2 中的突变强度预测明显偏弱。将试验 2 所得各模态预测结果进行重构可得到降水距平的预测结果（具体结果见图 3）。

表 4 给出了各预测试验的相关系数和均方根误差(Root-Mean-Square Error, RMSE)对比。对试验 2 而言，除 IMF2 的相关系数为 0.20 外，其他几个高频模态的相关系数平均在 0.3 左右，整个剩余分量的相关系数为 0.63，比高频分量的预测精度要高，反映出低频分量的波动平缓，预测较容易，但由于低频分量的方差贡献小，对最终的预测结果没有决定性影响。各模态的位相和演变趋势预测基本正确，但准确性尚不足，总体预测的相关系数仅为 0.29。对预测误差的分布而言，随着频率的降低，均方根误差减小，这主要是因为低频模态的强度小，这同样也反映出，预测误差主要产生于中高频分量，前 4 个模态的预测精度决定了整个降水距平序列的预测效果。

表 4 高频分量模态预测结果与 EMD 分解结果之间的相关系数和均方根误差

Table 4 Correlation coefficients and RMSEs of predicted high frequency components and EMD decomposition results

	r			RMSE/mm		
	试验 1	试验 2	试验 3	试验 1	试验 2	试验 3
IMF1	—	0.29	0.42	—	4.73	4.51
IMF2	—	0.20	0.45	—	3.66	3.54
IMF3	—	0.41	0.35	—	1.84	1.81
IMF4	—	0.33	0.76	—	1.77	1.31
剩余分量	—	0.63	-0.22	—	1.75	0.93
合成	0.11	0.29	0.47	7.32	6.51	6.26

由图 3 中试验 2 结果还可看出，虽然试验 2 结果要明显好于试验 1，消除了经典 ARIMA 统计预测的系统性误差问题，但对于 6 月下旬和 7 月下旬的两次降水距平的强突变现象均未预测出来，而这两次突变主要来源于前 2 个高频模态的贡献，这在图 5 中有清晰的表现，表 4 也反映出试验 2 对高频分量预测效果的不理想，这可能是对突变预测失败的主要原因。

3.2.3 基于 Hilbert 变换的幅频分离预测试验分析

按照 EMD 的分解原理，得到的 IMF 模态分量中扣除了趋势项，可视为平稳的单分量信号序列。因此，在对 288 候历史样本序列进行 EMD 分解后，试验 3 运用 Hilbert 变换求解所得模态分量的瞬时振幅和频率，求解结果如图 6 所示。各模态的振幅序列强度随频率降低而明显减弱，振幅序列的波动要明显弱于频率的起伏。对比图 2 可以发现，历史上 1995 年和 1996 年夏季也对应有明显强降水突变，这种突变与中高频分量的振幅有很好的对应，如图 6 中前 4 个模态的振幅均在 1995 年和 1996 年的夏季发生显著的突变增大，这表明中高频分量模态振幅的预测准确性，可能对突变现象的预测有着至关重要的影响。

对各模态的瞬时振幅和频率分别进行 LS-SVM 预测，模型的参数同样用 10-折交叉检验确定。将基于 330 候分解得到的前四个 IMF 分量的后 42 个候的振幅和频率作为近似真值，预测结果与近似真值之间的相关系数和均方根误差如表 5 所示。其中振幅的相关系数要明显高于频率，尤其是 IMF4 的相关高达 0.78，这或将显著改善对降水距平突变的预测效果。相对而言，高频分量的瞬时频率变化剧烈，预测难度大，其趋势预测效果不如振幅，但频率序列的强度小，预测偏差要比振幅小得多。无论振幅还是频率序列，其强度都要比 IMF 分量小，预

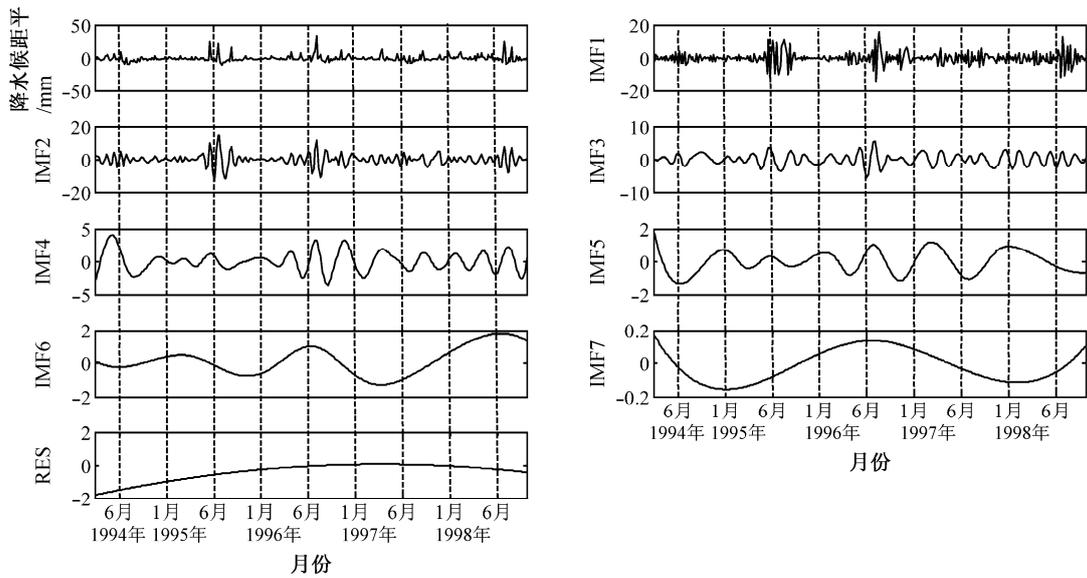


图 4 1994 年 3 月至 1998 年 9 月共 330 候降水距平 EMD 分解结果

Fig. 4 EMD decomposition results of 330 pentad-mean precipitation anomalies from Mar 1994 to Sep 1998

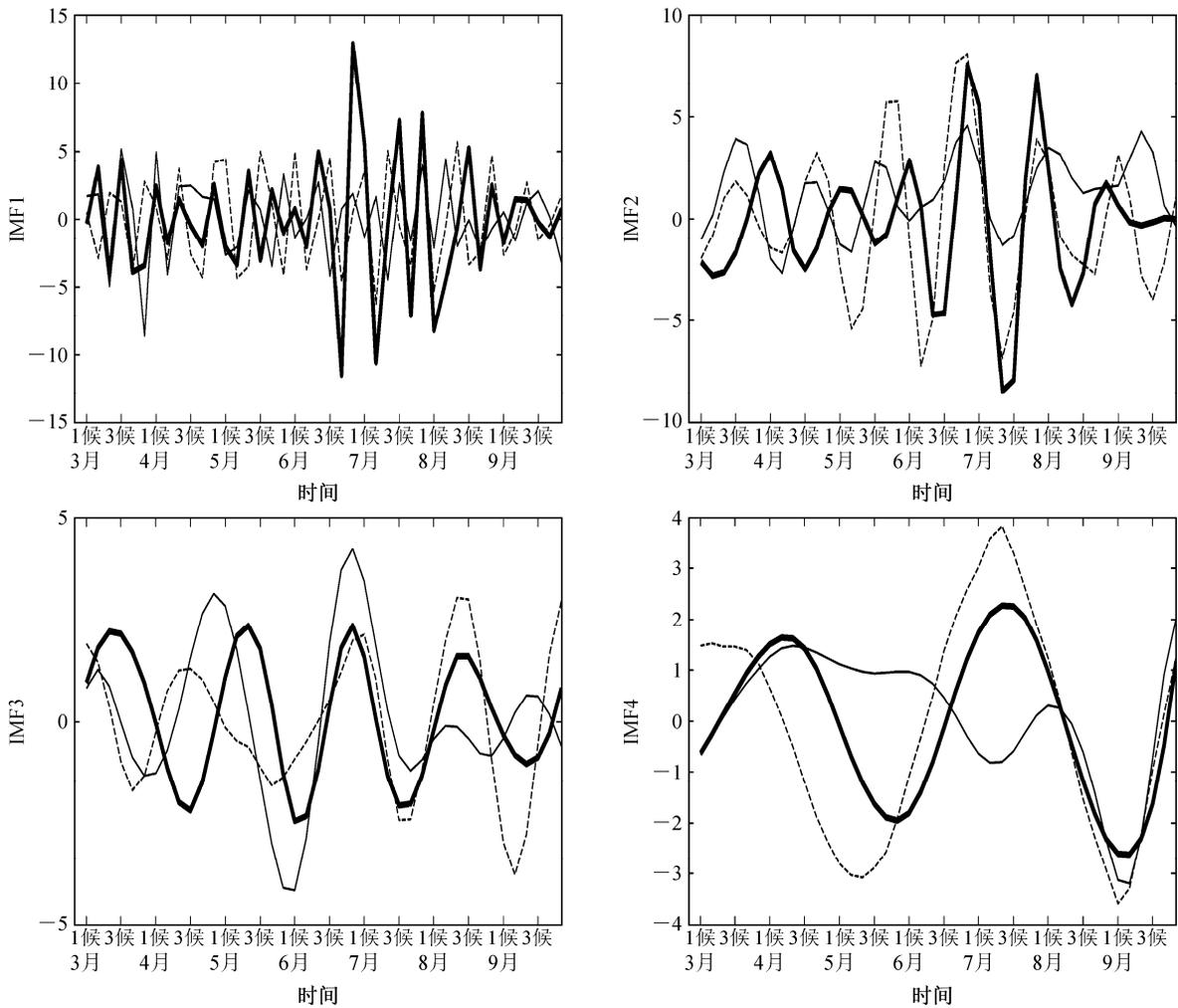


图 5 1998 年 3~9 月降水候距平前 4 个 IMF 高频分量的预测结果 (粗实线表示近似真值, 细实线表示试验 2 结果, 虚线表示试验 3 结果)

Fig. 5 The first four IMFs of pentad-mean precipitation anomalies from Mar to Sep of 1998 (the bold solid line indicates the real data, the thin solid line indicates the results of test 2, and the dash line indicates the results of test 3)

测偏差要显著小于对 IMF 分量的直接预测，这或将改善 Hilbert 法预测的各模态分量的偏差程度。

表 5 预测结果与近似真值的相关系数和均方根误差
Table 5 Correlation coefficients and RMSEs between the predicted results and the approximate real values

	<i>r</i>		RMSE	
	振幅	频率	振幅	频率
IMF1	0.52	0.14	3.31	0.12
IMF2	0.48	0.32	2.60	0.04
IMF3	0.28	-0.40	1.32	0.03
IMF4	0.78	0.03	0.54	0.02

将各分量模态的振幅和频率进行重构可得到各模态的预测结果，其中高频分量见图 5，其对应

的评估参数见表 4。从图 5 看，试验 3 得到的高频模态对突变的预测水平显著提高，尤其是 IMF2 对 1998 年 6 月下旬和 7 月下旬的 2 次突变强度和位相预测都较为准确，相比而言，对 IMF1 的预测略显不足，主要反映在对 6 月下旬突变强度的预测偏弱，但位相预测基本准确。从表 2 的相关系数看，试验 3 中 IMF1、IMF2、IMF4 的相关系数要显著高于试验 2，这表明，通过抑制振幅和频率的相互影响，进一步降低被预测序列的非平稳性，能够提高对复杂模态的演变趋势的预测能力。从均方差看，由于减小了被预测序列的平均强度，加强了对演变趋势的模拟能力，IMF1、IMF2、IMF4 的均方差相对试验 2 都有了较大程度的减小。

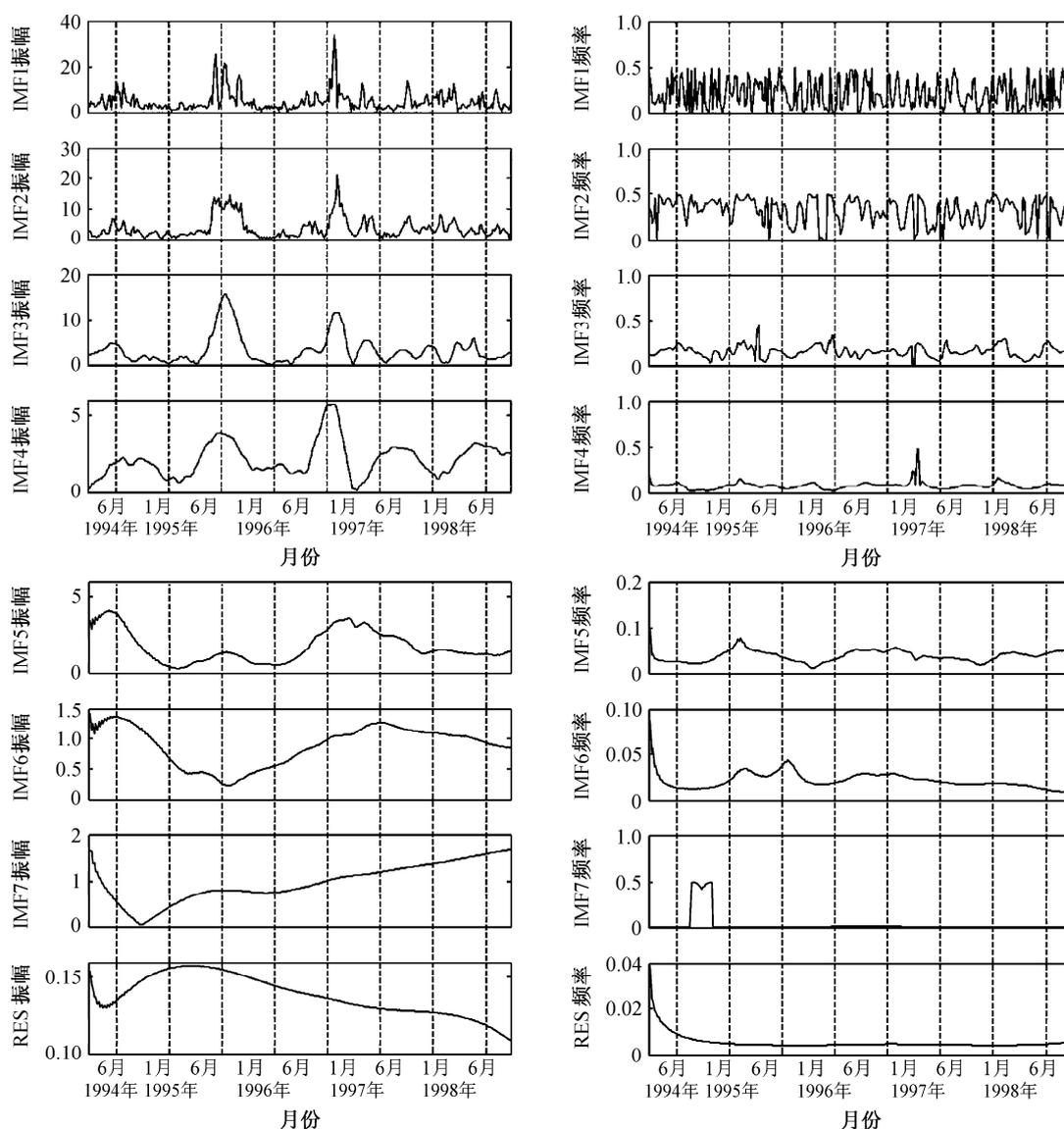


图 6 1994 年 3 月至 1998 年 2 月降水候距平序列 EMD 分解各特征模态的振幅和频率

Fig. 6 Instantaneous amplitudes and frequencies of IMFs for pentad-mean precipitation anomalies from Mar 1993 to Feb 1998

将试验 3 得到的各模态进行重构, 可得到降水候距平的预测值 (见图 3)。试验 3 较好地预测出 6 月下旬和 7 月下旬的 2 次强降水突变过程, 只是 6 月下旬的突变过程的降水强度预测略弱。整个夏季降水候距平的变化趋势预测有了显著提高, 其对应的相关系数高达 0.47, 远高于试验 1 和试验 2。每个模态的均方差均在试验 2 的基础上得到了改善, 整体的均方差更是有了较大幅度的提升。

由以上预测试验结果对比可以发现, ARIMA 模型的预测精度依赖于样本资料的平稳性, 尽管试验 1 通过差分等手段, 将非平稳序列转换为平稳序列, 但也因此流失了部分有用信息, 导致计算误差增大, 而且 ARIMA 模型的阶数以及参数的确定缺乏对样本资料物理特性的考虑, 导致虽能一定程度上反映出样本的变化趋势, 但造成预测结果系统性偏弱、对气候突变预测困难等问题。基于 EMD 分解的试验 2 能把对复杂降水距平序列的预测转换为对平稳 IMF 分量的预测, 可预测性显著提高, 克服了试验 1 预测系统性偏低的问题, 而且时间演变趋势预测水平显著提高, 但是由于对高频分量的突变现象预测困难, 导致最终对 2 次强降水异常突变的预测失败。相比之下, 基于 Hilbert 变换的试验 3 将 IMF 分量的瞬时振幅和频率分解隔离预测, 能够有效抑制两者的相互影响, 对高频分量振幅序列的成功预测较好地提高了对最终降水距平的异常突变预测水平, 且降水距平的演变趋势预测和偏差强度预测都得到较好的改进。

4 结果与讨论

本文基于长序列降水资料, 对 1998 年夏季南

京地区降水距平的预测问题, 设计了 ARIMA 模型预测、EMD 分解预测和基于瞬时振幅和频率的 Hilbert 分解预测等 3 种统计预测方法, 并进行了 7 个月的跨季节回报试验, 结果表明:

(1) 经典的 ARIMA 回归预测由于模型参数确定和优选难度大, 虽能一定程度上预测出未来降水距平的演变趋势, 但可能存在系统性误差和降水突变预测困难等问题。

(2) 基于 EMD 分解的预测方法通过对平稳化分量的预测能够显著提高可预测性, 预测的降水距平时间演变趋势明显提高, 但由于高频分量是 EMD 分解预测的关键, 而对高频分量的预测难度大, 尤其是对降水序列高频分量的突变信号预测不好, 导致对 1998 年夏季 2 次强降水预测失败。

(3) 基于 EMD 分解和 Hilbert 变换的预测方法, 通过提取出模态分量的瞬时振幅和频率序列并进行独立预测, 能够有效抑制两者的相互影响, 提高分量模态的预测准确性, 这种做法对高频分量的预测尤为有效, 能够有效预测出高频分量中的突变部分, 因此成功预测出夏季 2 次强降水突变异常, 并且降水距平的趋势预测和整体偏差预测都是最优的。

为进一步验证上述试验方法的通用性, 本文又用同样的方法对南京 2003 年 7 月初的暴雨进行了预测试验, 数据来源与处理方法与上文相同, 选取 1999 年 3 月 1 日至 2003 年 2 月 28 日的降水量作为历史样本, 设置预测时段为 2003 年 3 月 1 日至 9 月 30 日, 得到的结果如图 7 所示, 其所对应的相关系数和均方根误差如表 6 所示。

由预测结果的相关系数看, Hilbert 预测法得到

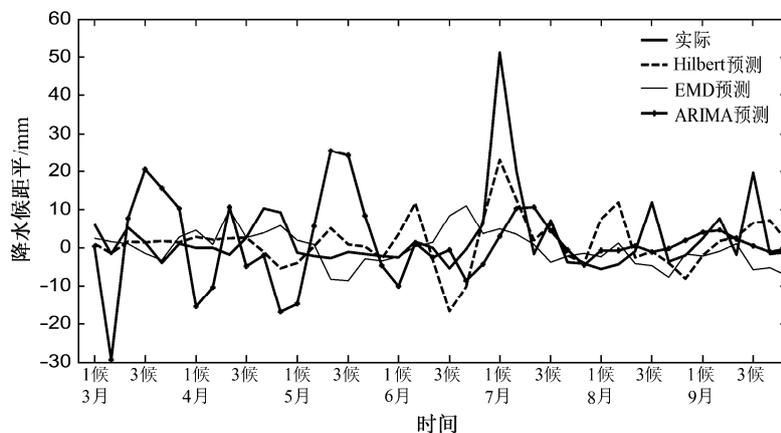


图 7 2003 年 3~9 月各预测试验降水候距平结果对比

Fig. 7 Comparison of pentad-mean precipitation anomalies of prediction tests and the real data from Mar to Sep of 2003

表 6 2003 年 3~9 月各预测试验降水候距平与观测数据之间的相关系数和均方差

Table 6 Correlation coefficients and RMSEs between the predicted pentad-mean precipitations anomalies and observed data from Mar to Sep of 2003

	r	RMSE/mm
ARIMA 试验	0.03	14.06
EMD 分解预测试验	0.18	10.24
Hilbert 预测试验	0.42	9.32

的降水候距平序列与实际序列之间的相关系数达到 0.42, 远高于 EMD 方法的 0.18 和 ARIMA 模型的 0.03, 表明 Hilbert 预测法的演变趋势预测是 3 个方法中最优的。均方差结果也表明 Hilbert 预测法的总体偏差只有 9.32, 也是各试验中最小的。此外, 图 7 也表明只有 Hilbert 方法预测出了 7 月的降水突变, 只是预测强度略偏小。由此可见, 本文提出的基于 Hilbert 变换的幅频分离预测方法具有较好的效果, 尤其是对气候突变具有较强的预测能力, 并且预测效果具有一定的普适应和稳定性, 这为跨季节尺度气候统计预测提供了新思路。

参考文献 (References)

- Box G E P, Jenkins G M. 1970. Time Series Analysis: Forecasting and Control (1st ed) [M]. San Francisco, CA: Holden-Day, 691pp.
- 丑纪范, 徐明. 2001. 短期气候数值预测的进展和前景 [J]. 科学通报, 46 (11): 890-895. Chou Jifan, Xu Ming. 2001. Advancement and prospect of short-term numerical climate prediction [J]. Chinese Science Bulletin, 46 (18): 1497-1502, doi: 10.3321/j.issn:0023-074X.2001.11.002.
- Cortes C, Vapnik V. 1995. Support-vector networks [J]. Machine Learning, 20 (3): 273-297, doi:10.1007/BF00994018.
- 管鹏. 2007. 非平稳时间序列建模与预测 [D]. 兰州大学硕士学位论文, 5-11. Guan Peng. 2007. Non-stationary time series modeling and forecasting [D]. M. S. thesis (in Chinese), Lanzhou University, 5-11.
- 韩建平, 钱炯, 董小军. 2010. 采用镜像延拓和 RBF 神经网络处理 EMD 中端点效应 [J]. 振动、测试与诊断, 30 (4): 414-417. Han Jianping, Qian Jiong, Dong Xiaojun. 2010. Suppression of end-effect in empirical mode decomposition by mirror extension and radial basis function neural network prediction [J]. Journal of Vibration, Measurement and Diagnosis (in Chinese), 30 (4): 414-417, doi: 10.3969/j.issn.1004-6801.2010.04.015.
- 胡广书. 2003. 数字信号处理——理论、算法与实现 (2 版) [M]. 北京: 清华大学出版社, 649pp. Hu Guangshu. 2003. Digital Signal Processing-Theory, Algorithm and Implementation (2nd ed.) (in Chinese) [M]. Beijing: Tsinghua University Press, 649pp.
- Huang N E, Shen Z, Long S R, et al. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society A, 454 (1971): 903-995, doi: 10.1098/rspa.1998.0193.
- Huang N E, Chern C C, Huang K, et al. 2001. A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999 [J]. Bulletin of the Seismological Society of America, 91 (5): 1310-1338, doi: 10.1785/0120000735.
- 霍玉洪. 2009. 基于非线性时间序列分析的气象分析 [D]. 安徽大学硕士学位论文, 10-11. Huo Yuhong. 2009. The meteorological analysis on the basis of nonlinear time series analysis [D]. M. S. thesis (in Chinese), Anhui University, 10-11, doi: 10.7666/d.203668.
- 纪玲玲, 林振山, 王昌雨, 等. 2009. 最小二乘回归支持向量机对非线性时间序列预测的试验分析 [J]. 解放军理工大学学报 (自然科学版), 10 (1): 92-97. Ji Lingling, Lin Zhenshan, Wang Changyu, et al. 2009. Experiments on nonlinear time series prediction with least square support vector regression machine [J]. Journal of PLA University of Science and Technology (Natural Science Edition), 10 (1): 92-97, doi:10.7666/j.issn.1009-3443.20090117.
- 李丽娟. 2008. 最小二乘支持向量机建模及预测控制算法研究 [D]. 浙江大学博士学位论文, 14-21. Li Lijuan. 2008. The study of modeling algorithm based on LS-SVM and predictive control algorithm [D]. Ph. D. dissertation (in Chinese), Zhejiang University, 14-21.
- 廖润贤. 1999. 论大气模式的设计 [J]. 气象学报, 57 (5): 513-524. Liao Dongxian. 1999. On the design of the atmospheric model [J]. Acta Meteorologica Sinica (in Chinese), 57 (5): 513-524, doi:10.11676/qxxb1999.049.
- Linderherd A. 2005. Compression by image empirical mode decomposition [C]// IEEE International Conference on Image Processing 2005. Genova: IEEE, doi: 10.1109/ICIP.2005.1529810.
- 刘科峰, 张初, 洪梅, 等. 2009. 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, 20 (3): 354-359. Liu Kefeng, Zhang Ren, Hong Mei, et al. 2009. Subtropical high forecast model of least square support vector machine [J]. Journal of Applied Meteorological Science (in Chinese), 20 (3): 354-359, doi: 10.11898/1001-7313.20090312.
- 刘莉红, 郑祖光, 琚建华. 2008. 基于 EMD 方法的我国年气温和东部年降水量序列的振荡模态分析 [J]. 高原气象, 27 (5): 1060-1065. Liu Lihong, Zheng Zuguang, Ju Jianhua. 2008. Vacillations of the time series of annual temperature and precipitation in China based on EMD method [J]. Plateau Meteorology (in Chinese), 27 (5): 1060-1065.
- 穆穆, 李建平, 丑纪范, 等. 2002. 气候系统可预报性理论研究 [J]. 气候与环境研究, 7 (2): 227-235. Mu Mu, Li Jianping, Chou Jifan, et al. 2002. Theoretical research on the predictability of climate system [J]. Climatic and Environmental Research (in Chinese), 7 (2): 227-235, doi: 10.3878/j.issn.1006-9585.2002.02.10.
- 孙继昌, 刘金平, 梁家志. 2004. 1998 年洪水调查及评价 [J]. 水文, 24 (5): 14-19. Sun Jichang, Liu Jinping, Liang Jiazhi. 2004. Investigation and evaluation of the floods in 1998 [J]. Hydrology (in Chinese), 24 (5): 14-19, doi: 10.3969/j.issn.1000-0852.2004.05.004.
- Suykens J A K, Vandewalle J. 1999. Least squares support vector machine classifiers [J]. Neural Processing Letters, 9 (3): 293-300, doi: 10.1023/A:1018628609742.
- 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 [J]. 气象学报, 63 (4): 516-525. Wan Shiquan, Feng Guolin, Zhou Guohua, et al. 2005. Extracting useful information from the

- observations for the prediction based on EMD method [J]. *Acta Meteorologica Sinica* (in Chinese), 63 (4): 516–525, doi: 10.11676/qxsb2005.051.
- Wu Z H, Huang N E. 2009. Ensemble empirical mode decomposition: A noise-assisted data analysis method [J]. *Advances in Adaptive Data Analysis*, 1 (1): 1–41, doi: 10.1142/s1793536909000047.
- 袁重光, 李旭, 曾庆存. 1996. 跨季度气候距平数值预测研究小结 [J]. *气候与环境研究*, 1 (2): 150–159. Yuan Chongguang, Li Xu, Zeng Qingcun. 1996. Summary of numerical prediction research on extraseasonal climate anomalies [J]. *Climatic and Environmental Research* (in Chinese), 1 (2): 150–159, doi: 10.3878/j.issn.1006-9585.1996.02.06.
- 曾庆存, 张学洪, 郭裕福, 等. 1999. 气候系统模式、气候数值模拟及气候预测理论研究 [J]. *中国科学院院刊*, 14 (1): 51–53. Zeng Qingcun, Zhang Xuehong, Guo Yufu, et al. 1999. Research on climate system model, climate numerical simulation and climate prediction theory [J]. *Proceedings of the Chinese Academy of Sciences* (in Chinese), 14 (1): 51–53.
- 张海勇, 马孝江, 盖强. 2002. 一种新的时变参数 AR 模型分析方法 [J]. *大连理工大学学报*, 42 (2): 238–241. Zhang Haiyong, Ma Xiaojiang, Gai Qiang. 2002. A new method for time-varying AR model [J]. *Journal of Dalian University of Technology* (in Chinese), 42 (2): 238–241, doi: 10.3321/j.issn:1000-8608.2002.02.023.
- 张军. 2008. 灰色预测模型的改进及其应用 [D]. 西安理工大学硕士学位论文, 5–6. Zhang Jun. 2008. Improvement of grey forecasting model and its application [D]. M. S. thesis (in Chinese), Xi'an University of Science and Technology, 5–6.
- 张明阳, 王克林, 刘会玉, 等. 2005. 基于 EMD 的洪涝灾害成灾面积波动的多时间尺度分析 [J]. *中国农业气象*, 26 (4): 220–223, 238. Zhang Mingyang, Wang Kelin, Liu Huiyu, et al. 2005. Analysis of fluctuation of inundated area of flood disaster at multi-time scale based on empirical mode decomposition (EMD) method in China [J]. *Chinese Journal of Agrometeorology* (in Chinese), 26 (4): 220–223, doi: 10.3969/j.issn.1000-6362.2005.04.004.