

张舰齐, 王丽琼, 左瑞亭, 等. 2017. 基于幅频分离的气候时间序列预测试验 [J]. 大气科学, 41 (3): 501–514. Zhang Jianqi, Wang Liqiong, Zuo Ruiting, et al. 2017. A predicting test on climatic time series based on amplitude-frequency separation [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 41 (3): 501–514, doi:10.3878/j.issn.1006-9895.1608.16135.

基于幅频分离的气候时间序列预测试验

张舰齐^{1,2} 王丽琼² 左瑞亭² 叶晶¹ 马秋丽¹ 叶成志³

¹ 中国人民解放军 95871 部队, 湖南衡阳 421000

² 中国人民解放军理工大学气象海洋学院, 南京 211101

³ 湖南省气象台, 长沙 410118

摘 要 从气候波动的瞬时频率与瞬时振幅出发, 结合最小二乘支持向量机技术, 提出了基于幅频分离技术的气候时间序列预测方法, 并对南京地区降水距平进行了 30 候的预测试验。结果表明, 幅频分离预测法能够对所有模式的振幅和高频模式的瞬时频率进行较好的预测, 而预测的瞬时频率累积误差会对模式分量的预测距平相关性产生敏感影响, 该新方法能够显著提高气候序列高频模式的预测效果。对于气候序列的低频模式分量, 集合经验模式分解的边界效应会对瞬时频率的求解产生较大误差, 使得序列边界区的幅角计算不准确, 导致对低频模式的最终预测效果不理想。对气候序列的高频分量采用幅频分离并进行最小二乘支持向量机预测, 而对其低频分量仅采用最小二乘支持向量机进行直接预测, 可同时提高高、低频分量的预测效果, 并最终提高整个气候序列的预测准确性。该分频预测方法可以使南京降水预测的 30 候距平相关保持在 0.4 以上。

关键词 幅频分离 高频分量 瞬时频率 最小二乘支持向量机 经验模式分解

文章编号 1006-9895(2017)03-0501-14

中图分类号 P467

文献标识码 A

doi:10.3878/j.issn.1006-9895.1608.16135

A Predicting Test on Climatic Time Series Based on Amplitude-Frequency Separation

ZHANG Jianqi^{1,2}, WANG Liqiong², ZUO Ruiting², YE Jing¹, MA Qiuli¹, and YE Chengzhi³

¹ No.95871 Unit of PLA, Hengyang, Hunan Province 421000

² Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101

³ Hunan Meteorological Observatory, Changsha 410118

Abstract From the perspective of the instantaneous frequency and amplitude of climatic wave series and by virtue of the technique of least square support vector machine (LS-SVM), a new prediction method of climatic series is proposed based on the separation of amplitude and frequency. A 30-pentad prediction test on Nanjing precipitation is conducted using this method. The results show that, the new prediction method based on the amplitude-frequency separation presents good prediction accuracies on both the amplitudes of all modes and the frequencies of higher frequency modes. The accumulated errors of predicted instantaneous frequencies have highly sensitive impacts on the anomaly correlations of modes. This method can distinctly improve the prediction of higher frequency modes. For the lower frequency modes, the boundary effect of ensemble empirical mode decomposition (EEMD) causes remarkable errors on the calculation of

收稿日期 2016-02-16; **网络预出版日期** 2016-08-16

作者简介 张舰齐, 男, 1990 年出生, 博士研究生, 讲师, 主要从事气候数值模拟与预测研究。E-mail: pla1990@yeah.net

通讯作者 左瑞亭, E-mail: ratinzuo@126.com

资助项目 国家自然科学基金资助项目 41475071, 国家青年科学基金资助项目 41305018, 财政部/科技部公益类行业专项 GYHY201306016

Funded by National Natural Science Foundation of China (Grant 41475071), National Youth Science Funds (Grant 41305018), Ministry of Finance / Ministry of Science and Technology of the Public Sector Special (Grant GYHY201306016)

instantaneous frequency, which subsequently leads to inaccurate argument and eventually results in unsatisfactory prediction on modes of lower frequencies. Thereby, implementing both amplitude-frequency separation and LS-SVM for the prediction of higher frequency modes of climatic series while merely using LS-SVM for the prediction of lower frequency modes can give perfect predictions on components of both higher and lower frequencies, and ultimately improve the prediction of the whole climatic series. The test implementing this frequency-based prediction method on prediction of precipitation in Nanjing shows that the anomaly correlation remains greater than 0.4 in its 30-pentad prediction.

Keywords Amplitude-frequency separation, High frequency, Instantaneous frequency, Least-square support vector machine, Ensemble empirical mode decomposition

1 引言

时间序列分析在统计学中占有十分重要的地位(吕金虎, 2002), 它是研究复杂系统特征规律的重要内容。特别是气候时间序列的分析研究, 对气候系统的复杂演变有着至关重要的指引作用, 而气候时间序列的分析也在气候预测领域中占有重要的地位(丑纪范, 2003)。但是, 气候时间序列是一个非线性非平稳过程(王革丽等, 2011), 而对非线性非平稳序列的预测较为困难, 一些尝试通过改善“遍历性”对时间序列进行预测的方法其效果也十分有限(王革丽等, 2004)。有研究指出, 气候时间序列的非平稳性是由于不断变化的外源强迫所造成的(Manuca and Savit, 1996; Wang et al, 2013), 把这些外源强迫考虑到预测模型中可以较好地提高预测效果(张彬等, 2012)。但实际上, 气候时间序列变化十分复杂, 导致其非平稳性的外部动力可能是由多种因素共同构成, 获取外部强迫信息并不容易。还有一些研究指出, 时间序列的非线性非平稳性的本质是波内频率调制现象(黄大吉等, 2003), 即振幅(能量)和频率在一个波内存在波动。对于气候时间序列的非线性非平稳的性质, 波内频率调制现象与不断变化的外源强迫在本质上是一致的, 从波内频率调制现象入手可能是提高预测准确性的有效方法。

为降低时间序列的非平稳性, 近些年发展起来的基于“气候层次”理论的分解隔离预测法(杨培才等, 2003; Wang and Yang, 2005; 郑祖光和刘莉红, 2010)有较多的应用, 所谓分解隔离预测法就是将多时间尺度的资料序列分解为若干具有特定时间尺度的分量, 对各个分量分别预测后再合成最终的预测结果。其中经验模态分解法(Empirical Mode Decomposition, 简称 EMD)的应用在一定程度上提高了非线性非平稳时间序列的分析预测效

果(Yang et al, 2010)。这种方法可将非平稳的时间序列迅速分解为有限个单分量模态(Intrinsic Mode Function, 简称 IMF), 从而降低了时间序列的非平稳性。然而, 一些研究证明, EMD 方法虽然能够降低时间序列的非平稳性, 但是对 EMD 分解得到的高频分量预测仍然是极其困难的(玄兆燕和杨公训, 2008), 高频分量仍然具有一定的非线性和非平稳特性。此外, 基于相空间重构和嵌入定理的提出也为非线性非平稳时间序列的预测提供了较好的借鉴(Takens, 1981), 诸如神经网络预测法(金龙等, 2000; 张军峰和胡寿松, 2007)、支持向量机(毛宇清等, 2007)和最小二乘支持向量机(朱佳等, 2010)也一定程度地提高了非线性非平稳时间序列的预测效果, 这些方法可以较好地运用到 EMD 分解后各个频段的分量预测中。

综合以上考虑, 本文拟先采用集合经验模态分解(EEMD)方法降低时间序列的非平稳性, 针对 EEMD 分解所得到高频分量预测的困难性, 从波内频率调制现象入手, 即从 EEMD 分量模态的瞬时频率与瞬时振幅出发进行预测, 尝试提高 EEMD 高频分量的可预测性, 最终提升气候序列的整体预测准确性。

2 方法和资料

2.1 方法

(1) 集合经验模态分解

EMD 是处理非平稳序列的有效手段(Huang et al., 1998), 由于其已广泛使用, 具体方法不再赘述。EEMD 是在建立在经验模态分解的基础上, 通过在原始序列中引入若干次白噪声, 利用白噪声具有频率均匀的统计特征, 使原始序列在不同尺度上具有一定连续性, 从而改变原始序列极值点的特性, 有效地克服了模态混淆现象。经过 EEMD 分解处理, 即可将一个非线性、非平稳序列快速分解为

相互独立的若干特征尺度模态 (IMF) 和一个剩余项 (Wu et al., 2011):

$$S(t) = \sum_{i=1}^n F_{\text{IMF}_i}(t) + R_n(t), \quad (1)$$

其中, $S(t)$ 为原始时间序列, $F_{\text{IMF}_i}(t)$ 为分解得到的 IMF 模态, $R_n(t)$ 为剩余项。

(2) Hilbert-Huang 变换求瞬时频率

当前对瞬时频率的计算已经有了相当多的方法 (Boashash, 1992; Boudraa et al., 2004), 本文采用了 Hilbert-Huang 变换法 (简称 HHT) 求解瞬时频率 (Huang, 1998), 主要是考虑到基于该变换的瞬时频率经典定义具有明确的物理意义, 且解析信号与原始信号的频谱完全相同, 可以利用瞬时频率与瞬时振幅重构原信号。

对第 j 个 IMF 模态 $F_{\text{IMF}_j}(t)$ 进行 Hilbert 变换, 得到其正交序列 $H_j(t)$:

$$H_j(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{F_{\text{IMF}_j}(t')}{t-t'} dt', \quad (2)$$

其中, P 是柯西主值。进一步由原模态和其正交模态构造解析信号序列 $Z_j(t)$:

$$Z_j(t) = F_{\text{IMF}_j}(t) + iH_j(t) = a_j(t)e^{i\theta_j(t)}, \quad (3)$$

其中, $a_j(t)$ 和 $\theta_j(t)$ 分别为解析信号的振幅和幅角 (或位相)。则瞬时频率 $\omega_j(t)$ 可定义为

$$\omega_j(t) = \frac{d\theta_j}{dt}. \quad (4)$$

由此经过 HHT 后, 可以把资料序列 $S(t)$ 表述为各 IMF 分量 $F_{\text{IMF}_j}(t)$ 的瞬时频率 $\omega_j(t)$ 和瞬时振幅 $a_j(t)$ 的形式 (这里未含剩余项):

$$S(t) = \sum_{j=1}^n F_{\text{IMF}_j}(t) = \sum_{j=1}^n a_j(t)e^{i\int \omega_j(t) dt}. \quad (5)$$

(3) 最小二乘支持向量机

支持向量机 (Support Vector Machine, 简称 SVM) 是最近几年兴起的机器学习方法, 对比传统的神经网络学习法, 支持向量机解决了神经网络中无法避免的局部极小值问题。Suykens and Vandewalle (1999) 提出的最小二乘支持向量机 (Least Square Support Vector Machine, 简称 LS-SVM) 是支持向量机的一种改进, 其将传统的不等式约束改为等式约束, 将误差平方和损失函数作为训练集的经验损失, 把二次规划问题转化为求解线性方程组, 提高了求解问题的速度和收敛精度。最小二乘支持向量机预测原理可简述如下:

对于资料序列向量 $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$, 采用非线性变换函数 $\mathbf{z} = \boldsymbol{\varphi}(\mathbf{x})$, 得到 m 维特征空间向量 \mathbf{z} , 高维特征空间中的计算会涉及到两向量的内积:

$$\langle \mathbf{z}_i \cdot \mathbf{z}_j \rangle = \sum_{k=1}^m \boldsymbol{\varphi}_k(x_i) \cdot \boldsymbol{\varphi}_k(x_j) = \boldsymbol{\varphi}(x_i)^\top \cdot \boldsymbol{\varphi}(x_j) = \langle \boldsymbol{\varphi}(x_i) \cdot \boldsymbol{\varphi}(x_j) \rangle. \quad (6)$$

由于特征空间维数较高, 上式计算中支持向量机往往用核函数 $K(x_i, x_j)$ 代替内积运算 $\langle \boldsymbol{\varphi}(x_i) \cdot \boldsymbol{\varphi}(x_j) \rangle$, 而核函数一般满足 Mercer 条件 (Cristianini and Shawe-Taylor, 2000) 即可, 在 SVM 中核函数有多种定义, 本文取径向基函数 RBF, 即高斯径向基核 (表示两点的距离):

$$K(x_i, x_j) = \exp \left[-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right], \quad (7)$$

其中, σ 为核参数, 调节核函数的平滑程度。这样对于非线性输入空间, 构造了高维特征空间 $\{\boldsymbol{\varphi}(\mathbf{x}), \mathbf{y}\}$, 其中, $\mathbf{x} \in R^1$, $\boldsymbol{\varphi}(\mathbf{x}) \in R^m$, $\mathbf{y} \in R^1$, R 为实数集, 其回归决策函数为

$$f(\mathbf{x}) = \langle \boldsymbol{\omega} \cdot \boldsymbol{\varphi}(\mathbf{x}) \rangle + b, \quad (8)$$

其中, $\boldsymbol{\omega}$ 和 b 为回归决策的参数。引入平方误差损失函数 e_i^2 代替标准向量机中的损失函数, 用等式约束:

$$\langle \boldsymbol{\omega} \cdot \boldsymbol{\varphi}(x_i) \rangle + b + e_i = y_i, \quad (9)$$

代替标准向量机的误差不等式约束, 则最优化目标函数转化为

$$\text{Min } F(\boldsymbol{\omega}, e) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2, \quad (10)$$

其中, γ 为正则化参数, 控制函数逼近误差的大小, 代表一定复杂性。

将约束条件:

$$\langle \boldsymbol{\omega} \cdot \boldsymbol{\varphi}(x_i) \rangle + b + e_i = y_i, \quad (11)$$

引入拉格朗日乘子 λ , 则目标函数的拉格朗日函数可写为

$$J = L(\boldsymbol{\omega}, e, b) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \lambda_i \left(\langle \boldsymbol{\omega} \cdot \boldsymbol{\varphi}(x_i) \rangle + b + e_i - y_i \right). \quad (12)$$

由泛函极小值条件可得:

$$\begin{pmatrix} b \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{E}^\top \\ \mathbf{E} & K(x, y) + \frac{1}{\gamma} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mathbf{Y} \end{pmatrix}, \quad (13)$$

其中, $\lambda = (\lambda_1, \lambda_2, \lambda_3 \cdots \lambda_n)^T$, $Y = (y_1, y_2, y_3 \cdots y_n)^T$, $E = (1, 1, 1 \cdots 1)^T$ 为 $n \times 1$ 维列向量。由此求解 λ 和 b , 可得非线性模型的决策函数:

$$f(x) = \sum_{i=1}^n \lambda_i K(x_i, x) + b, \quad (14)$$

至此可构建数据序列的预测模型。

(4) 边界效应的处理

EEMD 分解和 Hilbert 变换存在着边界效应, 边界效应的存在使得 EEMD 分解所得到的各项分量的瞬时频率与瞬时振幅在边界处存在较大的误差。为克服边界效应, 当前已经有了若干种解决边界效应的方法(胡爱军等, 2008; 沈路等, 2009), 例如极值延拓法、镜像延拓法等, 但是由于分量序列在端点以外没有任何序列, 任何延拓都是人为的, 人为性过大也就使得在端点处的瞬时频率与瞬时振幅求解与实际相差甚远。因此, 本文尽可能充分利用原始资料的演变规律, 使用最小二乘支持向量机对原始资料序列延拓两个极大值和极小值, 以减弱边界效应。

2.2 资料

本研究所使用的资料来自国家气象局的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”, 主要用到南京站 2008~2013 年共计 6 年的逐日降水观测资料。本文取降水观测资料的候距平值进行研究, 求解距平时用到的气候均值为 2013 年前 30 年历史平均。用以本文预测建模的历史资料取为 2009~2012 年, 预测时段为 2013 年的前 30 候。实测资料的 IMF 分量取自 2008~2013 年共计 6 年的候降水距平序列分解结果, 取其 2009~2012 年共计 288 候进行比对分析。

3 预测方法分析

3.1 基本预测方法设计

对于南京地区降水的预测问题, 本文采用两种方法进行预测对比, 即基于 EEMD、HHT 幅频分离以及 LS-SVM 的预测方法(后简称“幅频分离预测法”)和仅基于 EEMD 而不进行 HHT 幅频分离的直接 LS-SVM 的预测法(后简称“直接预测法”)。具体而言, 幅频分离预测法首先利用 LS-SVM 对资料序列进行延拓, 对延拓后的序列进行 EEMD 分解, 对得到的 IMF 模态分量进行 HHT 变换求解瞬时频率与瞬时振幅, 然后对瞬时频率和瞬时振幅进行 LS-SVM 预测, 最后将预测的瞬时频率和振幅合

成得到各模态分量的预测值, 再由各模态分量预测值进一步合成得到原资料序列的预测结果。在 EEMD 分解中扰动白噪声与原始信号的信噪比为 0.3, 集合样本数取 1000。直接预测法与之类似, 只是对前者分解得到的 IMF 模态分量不进行 HHT 幅频分离求解, 而直接对其各模态分量进行 LS-SVM 预测及后续合成。

3.2 资料延拓对边界效应及 EEMD 模态分解的影响分析

EEMD 分解的三次样条插值函数法会在边界处产生边界效应(黄大吉等, 2003), 为解决该边界效应问题, 本文的两种预测方法都在分解前先对 2009~2012 年共计 288 候的资料进行 20 候的 LS-SVM 延拓。为进行对比, 同步将未用 LS-SVM 延拓资料序列直接进行 EEMD 分解, 分解中采用常用的极值延拓法进行简单的端点处理。为将两者与实际值对比, 本文还同时取 2008~2013 年共计 6 年的降水候距平进行 EEMD 分解, 取其 2009~2012 年共 288 候分解结果作为实测 IMF 分量。图 1 分别给出了上述采用和未采用 LS-SVM 延拓进行 EEMD 分解的累积绝对误差对比情况。显然, 经过 LS-SVM 延拓后的 EEMD 分解其各模态误差均显著小于未延拓情况, 除高频模态外, 累积绝对误差均能减小 50% 左右。此外, 随着分解模态层级的加深(模态序号增加), 边界误差逐渐向资料内部传播, 而经 LS-SVM 的延拓处理还可延缓误差向内部传播的速度。图 2 给出了两种情况与实际值的均方根误差和相关性对比, 显然, 延拓处理后各模态均方根误差显著减低, 且其相关性都有提高, 尤其是第 4、5 模态的提升效果最为明显。

为进一步分析各模态分量在资料序列中的相对重要性, 图 3 给出了经过 LS-SVM 延拓序列分解的各模态的方差贡献。各模态的方差贡献随序号递减, 其中高频分量所占方差贡献较大, IMF1 的方差贡献最大高达约 57%, 前 3 项的累积方差贡献约为 85%, 决定了整个资料序列的主要变化特征, 这也表明后续高频分量的预测将对整体序列预测的至关重要性。

3.3 瞬时频率的累积误差对模态分量预测的相关性影响分析

依据 HHT 原理, 任一特征模态分量可表示为

$$F_{IMj}(t) = a_j(t) e^{i \int \omega_j(t) dt} = a_j(t) \cos \left[\int \omega_j(t) dt \right] =$$

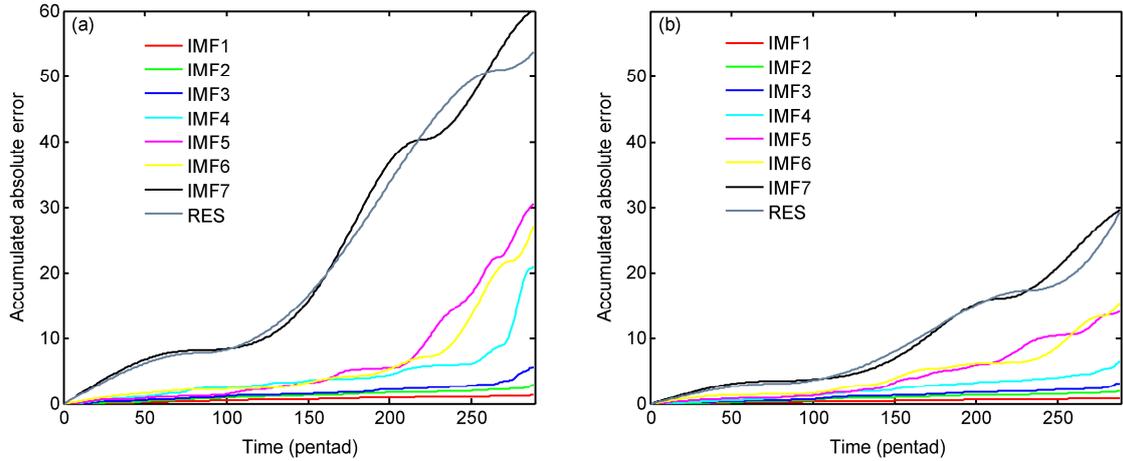


图1 原始降水距平(a)未经LS-SVM延拓和(b)经过LS-SVM延拓两种情况下进行集合经验模态分解(EEMD)得到的各IMF(Intrinsic Mode Function)模态分量及剩余项(RES)的累积绝对误差

Fig. 1 Accumulated absolute errors of IMF (Intrinsic Mode Function) modes and residual term (RES) of precipitation anomaly decomposed by EEMD (Ensemble Empirical Mode Decomposition): Precipitation anomaly series (a) without LS-SVM (Least Square Support Vector Machine) extension; (b) with LS-SVM extension

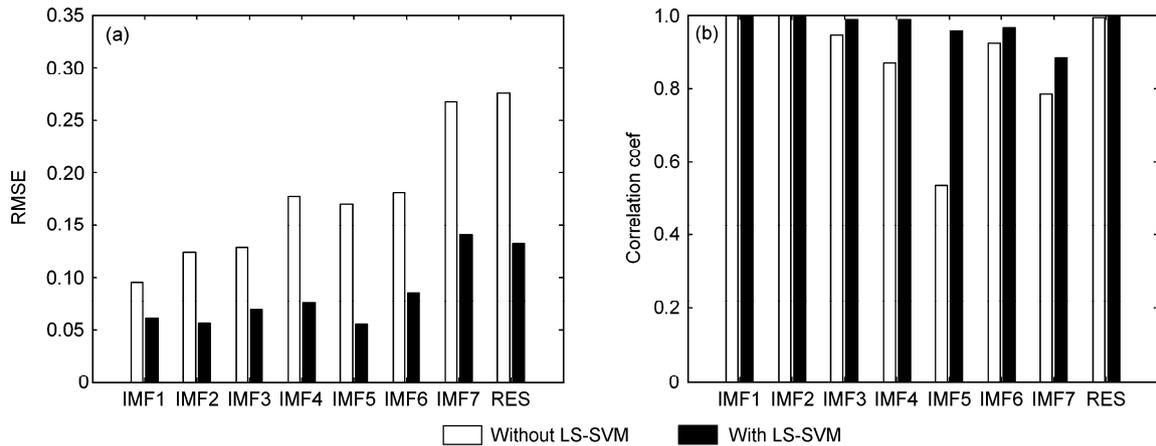


图2 未经LS-SVM延拓和经过LS-SVM延拓两种情况下IMF分量与真实IMF分量的(a)均方根误差和(b)相关性

Fig. 2 IMF of precipitation anomaly decomposed by EEMD with LS-SVM and without LS-SVM: (a) Root-mean-square error (RMSE) of IMF with real IMF; (b) correlation between the IMF and real IMF

$$a_j(t)\cos[\theta_j(t)], \quad (15)$$

如暂不考虑振幅预测影响, 设该模态的真值为 $F_{IMj}^*(t)$, 即

$$F_{IMj}^*(t) = a_j(t)\cos[\theta_j(t)]. \quad (16)$$

则其预测值为

$$F_{IMj}(t) = a_j(t)\cos[\theta_j(t) + \Delta\theta_j(t)], \quad (17)$$

其中, $\Delta\theta_j(t)$ 为该模态在 t 时刻的累积幅角误差。依据幅角计算原理, 每一步的幅角计算误差都将会被带到下一步的幅角计算中, 因此累积幅角误差 $\Delta\theta_j(t)$ 应为

$$\Delta\theta_j(t_k) = \sum_{i=1}^{i=k} \Delta\theta_j(t_i). \quad (18)$$

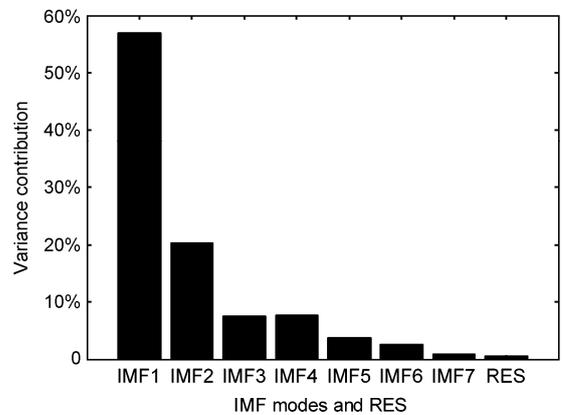


图3 原始降水距平序列经过LS-SVM延拓后进行EEMD分解所得IMF分量和剩余项RES的方差贡献

Fig. 3 Variance contributions of IMF modes and RES of precipitation anomaly decomposed by EEMD with LS-SVM extension

预测值 $F_{IMj}(t)$ 与真值 $F_{IMj}^*(t)$ 的相关分布如图 4 所示, 若需相关性不低于 0.6, 则幅角预测的累积误差应控制在 $(2k\pi - 0.9, 2k\pi + 0.9)$, 其中 k 为整数, 对应的瞬时频率误差控制范围应为 $[k - 0.9/(2\pi), k + 0.9/(2\pi)]$ 。取 $k = 0$ 时为第一阶段误差范围。由于幅角的累积误差是积分关系, 其并不是固定不变的值, 或者是具有某种特定规律, 但是瞬时频率或幅角的预测累积误差可对相关性有一定的指示作用, 累积幅角误差控制的越小则相关性越好。可见, 如能将瞬时频率累积误差控制在特定的范围区间, 则可保证较高的相关性。

3.4 资料延拓对边界区瞬时频率的计算影响分析

幅频分离预测法涉及到对瞬时频率的 LS-SVM 预测, 但在边界处, 瞬时频率具有一定的误差, 这可由 EEMD 分解的边界效应造成, 另一方面也可由 HHT 变换过程中的“频谱泄露”造成, 这是有限长资料序列的傅里叶变换引起的。对样本资料直接进行幅频分离预测势必引起明显的频率预测误差而导致预测结果不可靠。因此, 本文在 EEMD 分解前, 提前对资料样本进行 LS-SVM 延拓, 相当于将原资料序列的边界变为延拓后序列的“内区”, 理论上可以有效削弱瞬时频率的计算误差。如图 5 给出了对资料序列延拓与未延拓两种情况下边界区各模式的瞬时频率计算结果, 瞬时频率的计算在边界区域具有一定的误差, 且随着 IMF 的频率降低而逐渐增大, 序列延拓可以一定程度削弱边界区的频率计算误差, 尤其是对高频分量效果明显, 前 4 个较高频分量的误差总体得到了较好控制, 但低频分量的效果不甚明显。

4 预测效果分析

4.1 特征模式分量的预测分析

采用上述幅频分离预测法和直接预测法对各模式进行预测分析。为节省篇幅, 本文仅重点分析前 5 个分量的预测结果。如图 6 到图 10 所示, 幅频分离预测法中瞬时频率的累积预测误差对模式预测的相关性具有很敏感的影响, 如图 6a 和 c 所示, 当累积频率误差逐渐超出第一误差范围时, 相关性也在逐渐下降(起始阶段 1~5 候), 而累积频率误差重新回到第一误差范围内时(5~10 候), 相关性则变好。后续时段同样具有相类似的特征。采用幅频分离法预测得到的 IMF 分量的相关性平均都可达到 0.6 以上, 而采用直接预测法得到的 IMF

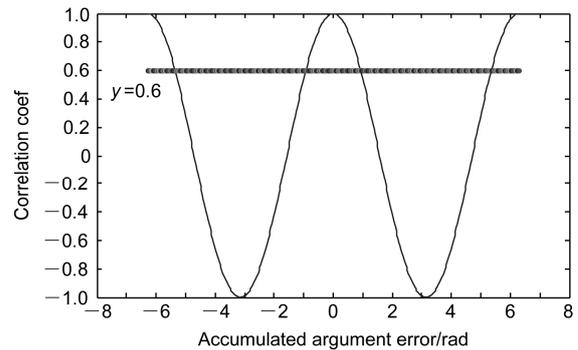


图 4 模式分量预测结果的相关性与幅角累积误差的关系

Fig. 4 The relationship between the accumulated argument error and correlations based on the prediction of IMF mode

分量的相关性则平均不超过 0.2。幅频分离预测的优势同样对振幅的预测也十分显著, 如图 6f, 平均而言, 直接预测法的振幅预测误差较大。上述预测特征在 IMF2 到 IMF4 的预测中同样体现较好, 如图 7 到图 9 所示, IMF2 的 ACC (Anomaly Correlation Coefficient) 在预测的 30 候内平均在 0.5 以上, IMF3 分量的 ACC 在 30 候以内平均在 0.4 以上, 相关性检验均超过了 95% 的信度检验。而直接预测法所得到的 ACC 评估则较不理想。

以上分析表明, 幅频分离预测法对高频分量预测具有较好的优越性, 使用最小二乘支持向量机对瞬时频率作出预测, 其累积误差与模式的相关性存在敏感影响关系, 某时刻瞬时频率预测的误差并不影响模式整体预测的相关性, 但是其累积误差却可以对整体预测相关性造成较大的影响, 如果能够较好的控制瞬时频率的累积误差在特定范围内变化, 则可以较好地提高预测相关性。如果说瞬时频率的变化从某一方面反映了外源强迫的变化, 则对瞬时频率的预测就是在推断外源强迫的变化, 这种规律可能是线性的也可能是非线性的。从另一方面考虑, 时间序列的非线性本质是波内频率调制现象, 对频率与振幅的不断变化作出较好的推断就可以改进时间序列预测的准确性。

不可否认的是, 幅频分离预测法虽然对高频分量具有一定的优势, 但是模式分量预测的 ACC 随着其模式的序号增加而呈现下降趋势, 如图 9 对 IMF4 分量的预测已经与直接预测法的结果相差无几。而 IMF5 分量的预测结果显示, 幅频分离预测法的结果使得位相相反, 虽然瞬时频率的预测难度不大, 并且累积误差控制在第一误差范围, 而图 5

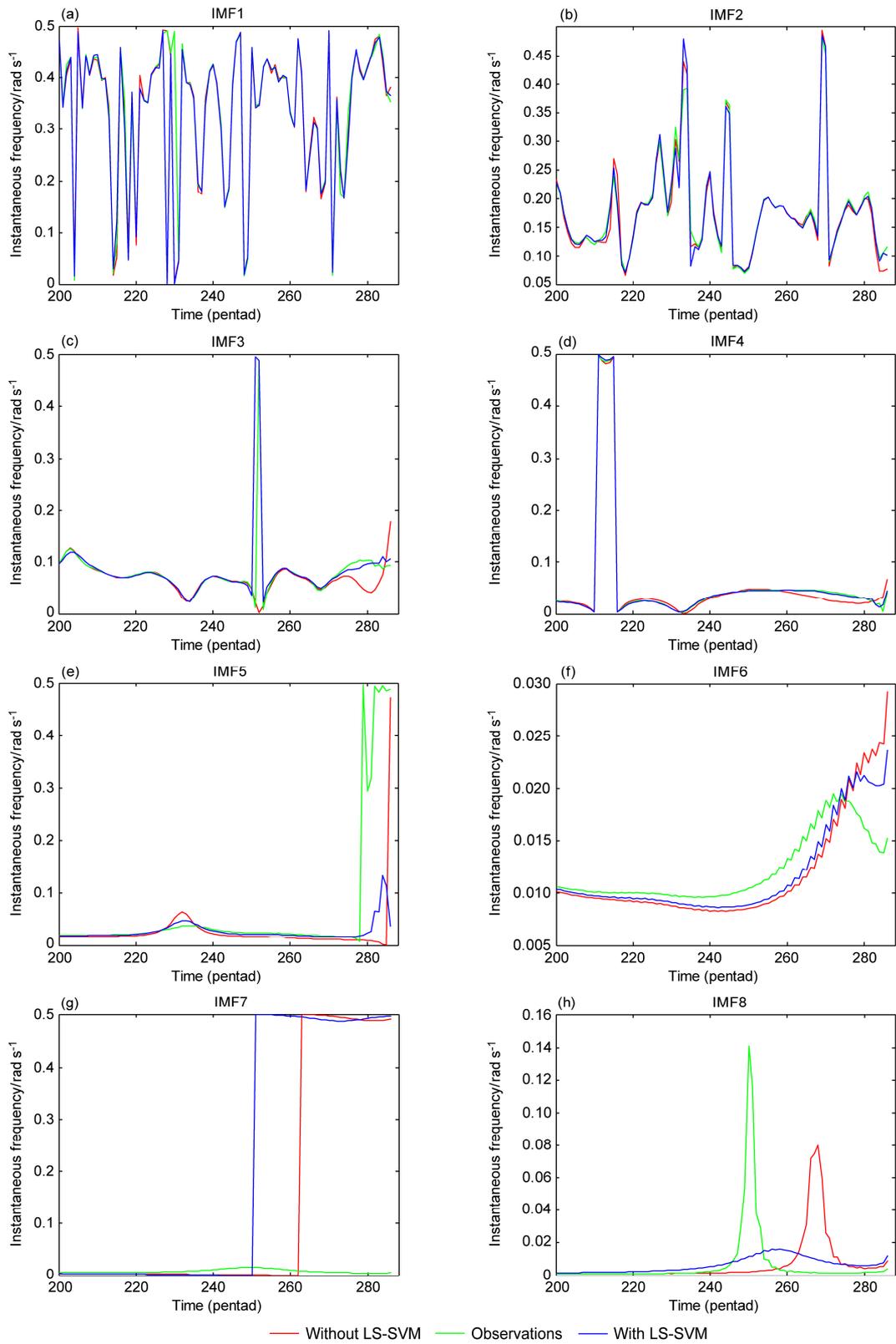


图5 (a-g) 各 IMF 模态和 (h) RES 延拓后与未延拓下的边界区瞬时频率

Fig. 5 Instantaneous frequencies in the boundary region with and without extension corresponding to (a-g) IMF1-IMF7 and (h) RES

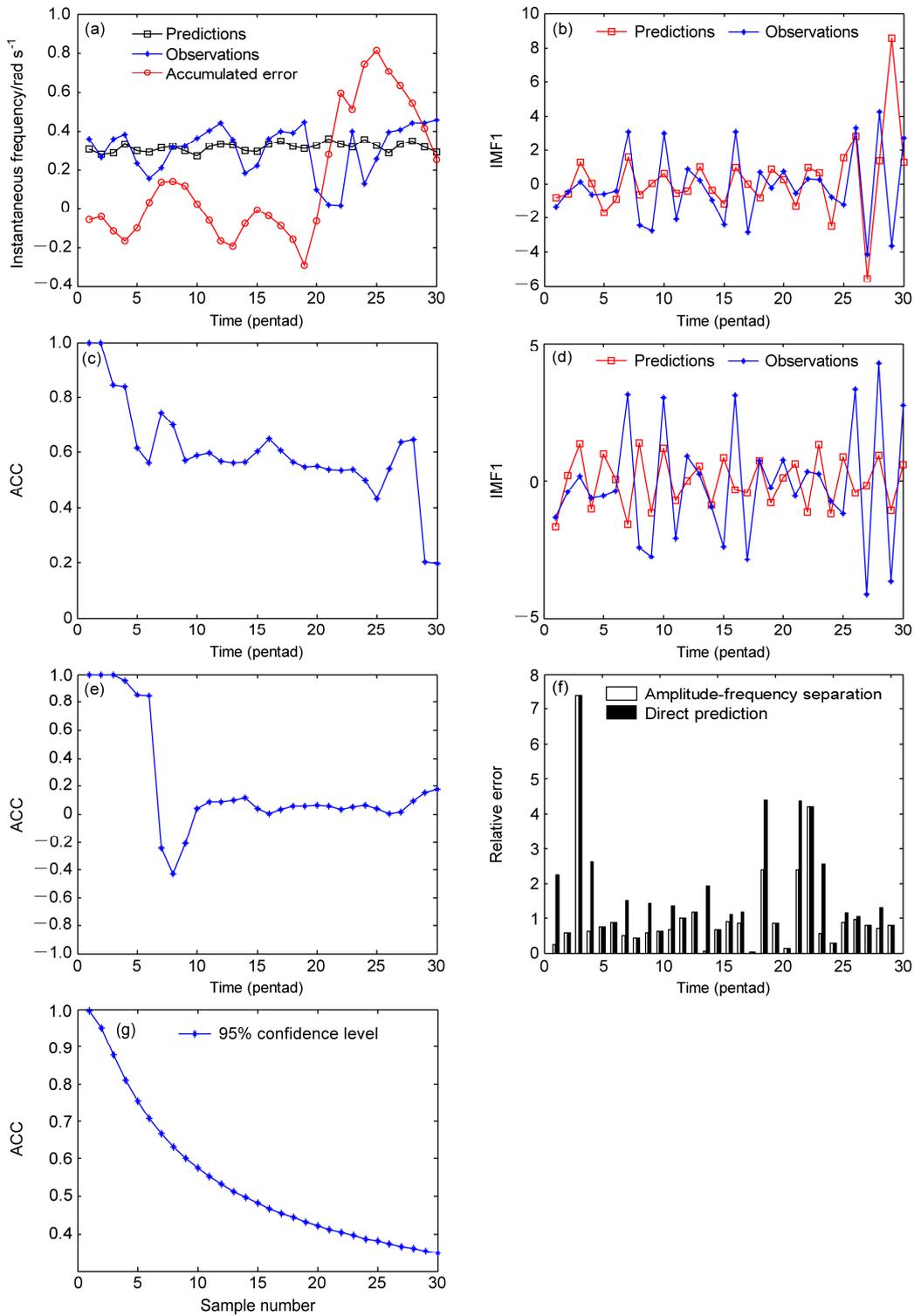


图 6 IMF1 模态分量的预测分析: (a) 瞬时频率预测累积误差; (b) 幅频分离预测法预测结果与实际对比; (c) 幅频分离预测法预测的距平相关系数; (d) 直接预测法预测结果与实际对比; (e) 直接预测法预测的距平相关系数; (f) 幅频分离与直接预测振幅的相对误差; (g) ACC (Anomaly Correlation Coefficient) 的相关性检验

Fig. 6 The predictions of IMF1: (a) Accumulated errors of predictions of instantaneous frequencies; (b) comparison of predictions based on amplitude-frequency separation method and observations; (c) anomaly correlation coefficients of predictions based on amplitude-frequency separation method; (d) comparison of prediction based on the direct prediction method and observations; (e) anomaly correlation coefficients of prediction based on the direct prediction method; (f) relative errors of predictions by amplitude-frequency separation method and by direct method; (g) significance test of ACC at the 95% confidence level

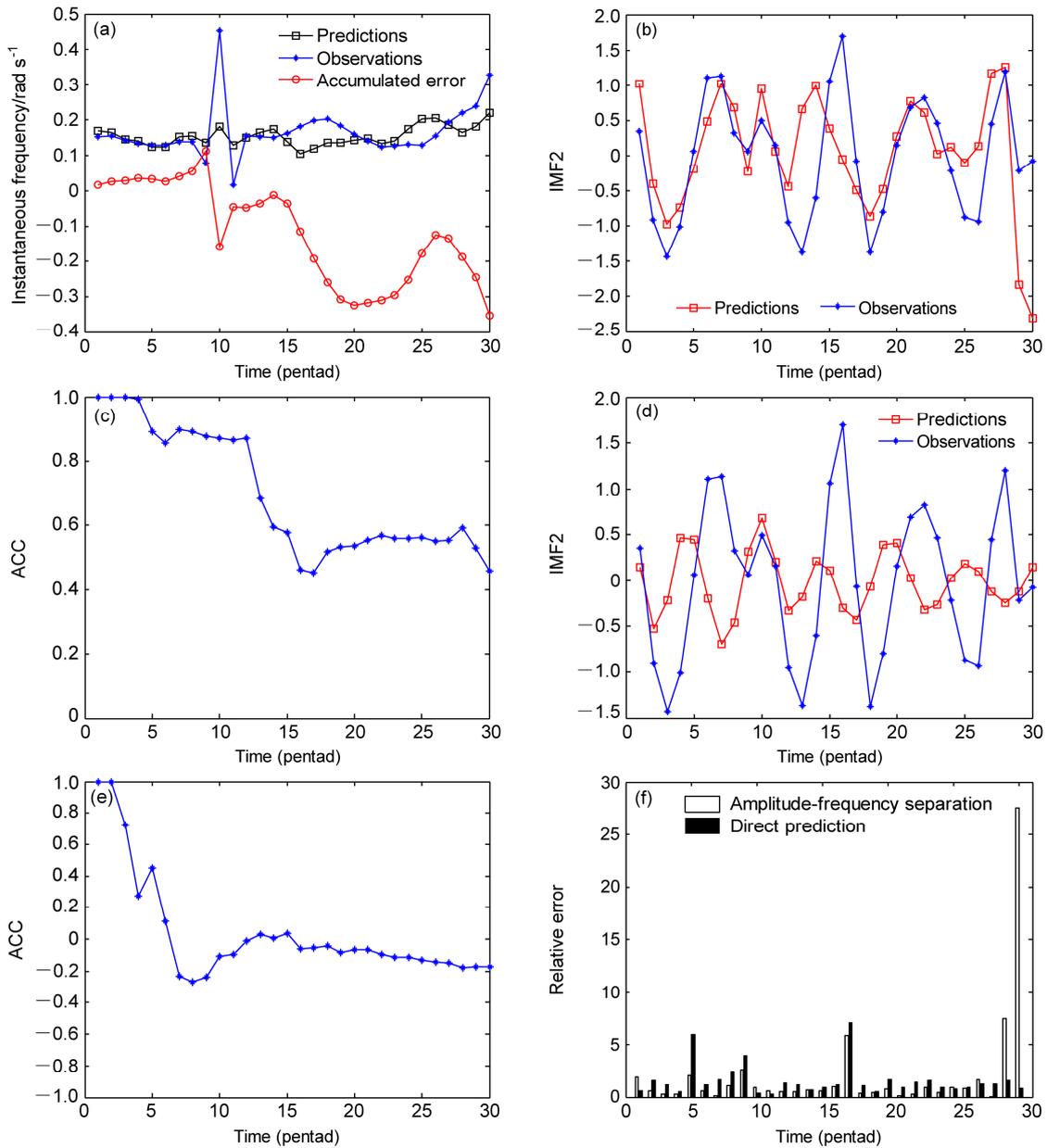


图7 同图 6a-f, 但为 IMF2
Fig. 7 Same as Fig. 6a-f, but for IMF2

显示, IMF5 边界处的瞬时频率具有较大的误差, 因而边界处的幅角计算值与真实值相比也具有较大的误差, 这种误差带到了后续预测中, 从而使得后续预测产生较大的误差, 甚至使得位相相反。相比而言, 直接预测法对低频分量的预测效果较好, 预测的 ACC 值相对较高。

4.2 合成预测结果分析

各模态的预测分析表明, 幅频分离预测法对

高频分量的预测效果较好, 而对变化相对较为简单的低频分量预测效果不理想, 而直接预测法的效果正好相反。因此, 本文尝试一种合成预测, 即将两者结合, 对高频分量 (IMF1 到 IMF4) 采用幅频分离预测法, 而对低频分量 (其余模态和剩余项) 采用直接预测法, 然后将二者结果进行合成重构, 暂称之为合成预测。图 11 给出了合成预测、幅频分离预测和直接预测三种方法对降水

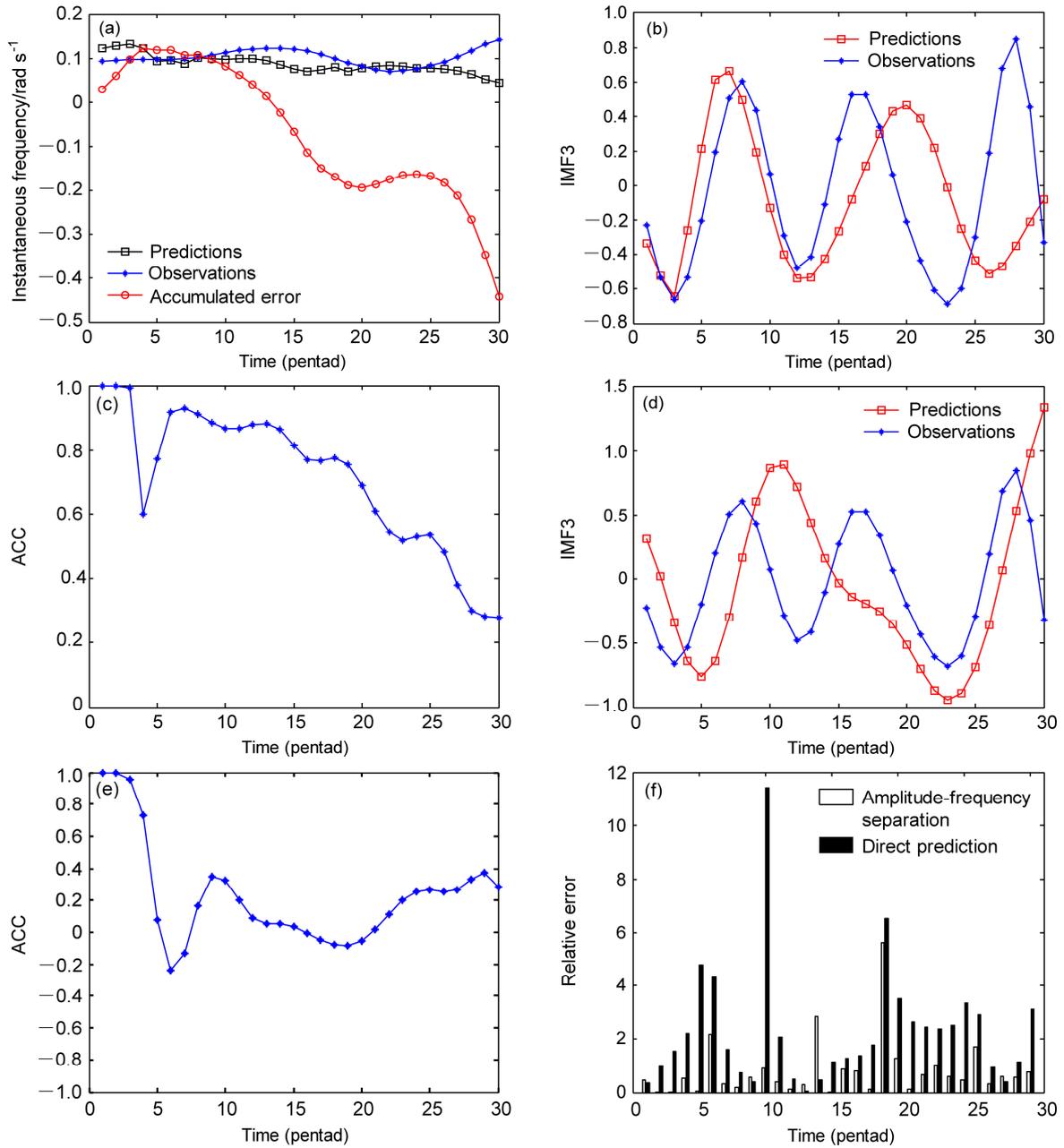


图8 同图 6a-f, 但为 IMF3
Fig. 8 Same as Fig. 6a-f, but for IMF3

距平的预测结果对比, 从图 11a 的降水距平预测图清晰可见, 合成预测方法的距平预测结果在趋势和偏差幅度上明显要优于另两种方法, 而幅频分离预测方案的结果又稍优于直接预测法。图 11b 给出了三者的不同预测时效的距平相关, 随时效延长, 预测降水距平相关逐渐递减, 但合成预测的距平相关始终保持最高, 幅频分离法其次, 直

接预测法的 ACC 预测效果不理想。图 11c 的均方根误差也呈现出与距平相关一致的结果, 合成预测的均方根误差最小, 其次是幅频分离法, 直接预测法有相对较高的均方根误差。

5 结论

本文设计了基于幅频分离预测法对南京地区

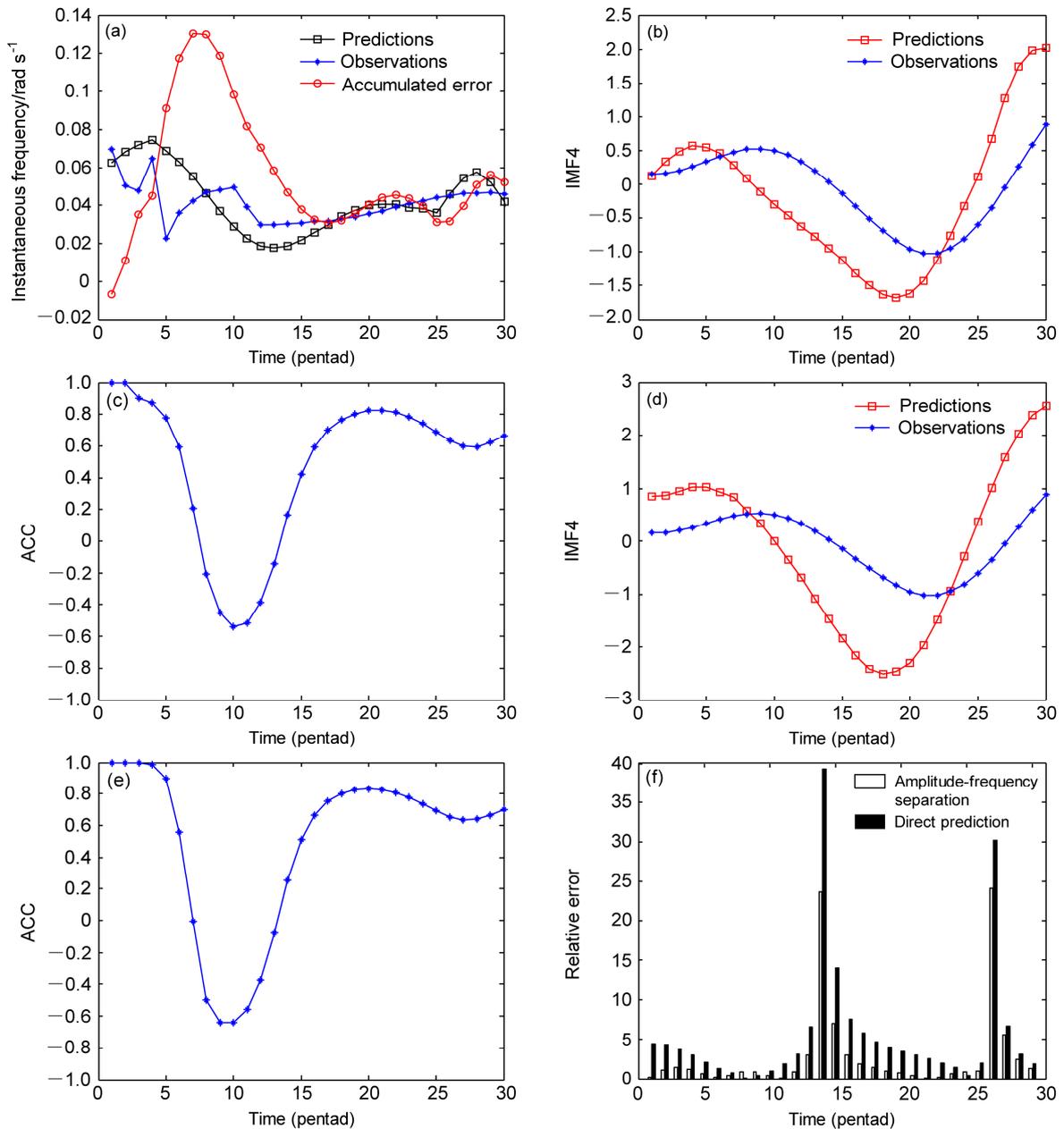


图9 同图 6a-f, 但为 IMF4

Fig. 9 Same as Fig. 6a-f, but for IMF4

降水距平进行 30 候的预测，并将幅频分离法、直接预测法和合成预测法进行了分析比较，得到了各个方法的优势及缺陷，主要结论如下：

(1) EEMD 分解在边界处会产生误差，并且误差会逐渐向内部扩散，这种误差尺度对中低频 IMF 分量具有较大的影响。对原始序列进行延拓后可在一定程度上削弱误差，延拓后分解得到的 IMF 模态

分量在均方根误差有较大的降低，相关性有较大幅度的提高。

(2) 基于幅频分离的预测法对高频 IMF 分量预测有明显优势。IMF 分量的高频部分占有较大的方差贡献，对高频分量的预测较为关键，基于幅频分离的预测方法从非线性时间序列的本质出发，通过把握时间序列的瞬时频率与瞬时振幅的变化规律

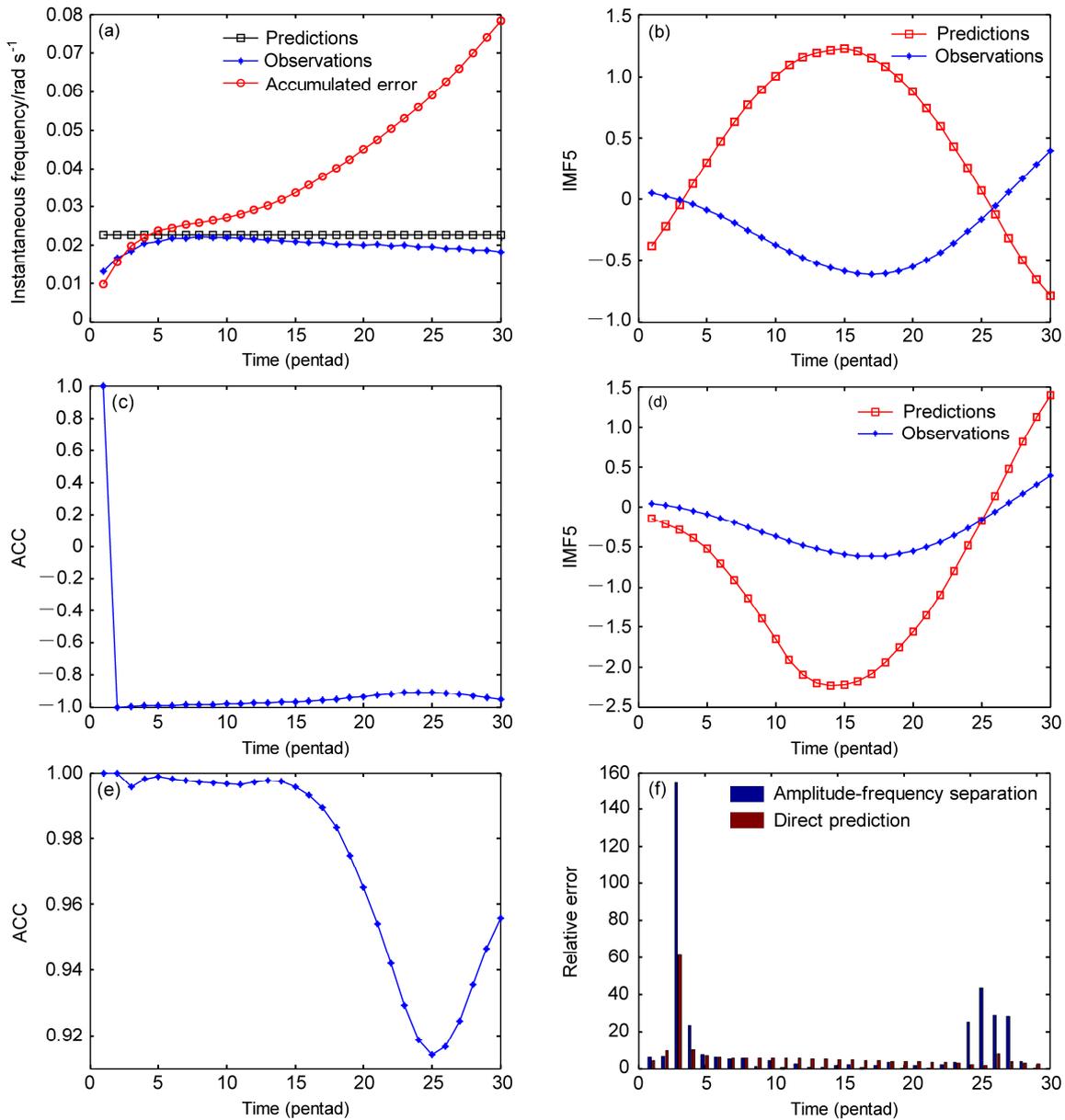


图 10 同图 6a-f, 但为 IMF5
Fig. 10 Same as Fig. 6a-f, but for IMF5

从而提高预测准确性。

(3) 基于幅频分离预测法对相对较为简单的低频 IMF 分量预测效果差。由于任意时刻的幅角值依赖于 IMF 分量在边界处的幅角值, 而 EEMD 分解以及求解瞬时频率的边界误差使边界处的幅角计算不够准确, 而经过原始资料的延拓后可削弱瞬时频率在边界处的误差, 特别是高频分量 (IMF1 到 IMF4) 效果较为明显。起始时刻的幅角准确性使高频分量预测较准确, 而低频分量在边界处的幅角误

差, 使得预测结果与真值的位相相反, 预测效果较不理想。

(4) 累积幅角预测误差对 IMF 分量的预测相关性具有指示作用, 而瞬时频率的预测准确性决定了累积幅角误差。瞬时频率预测的累积误差与 IMF 分量的相关性有着较好的对应关系, 在高频分量 (IMF1 到 IMF4) 的预测中体现更为明显。

(5) 基于幅频分离预测法对所有 IMF 分量的瞬时振幅预测具有较好的效果, 这种较好的效果不依

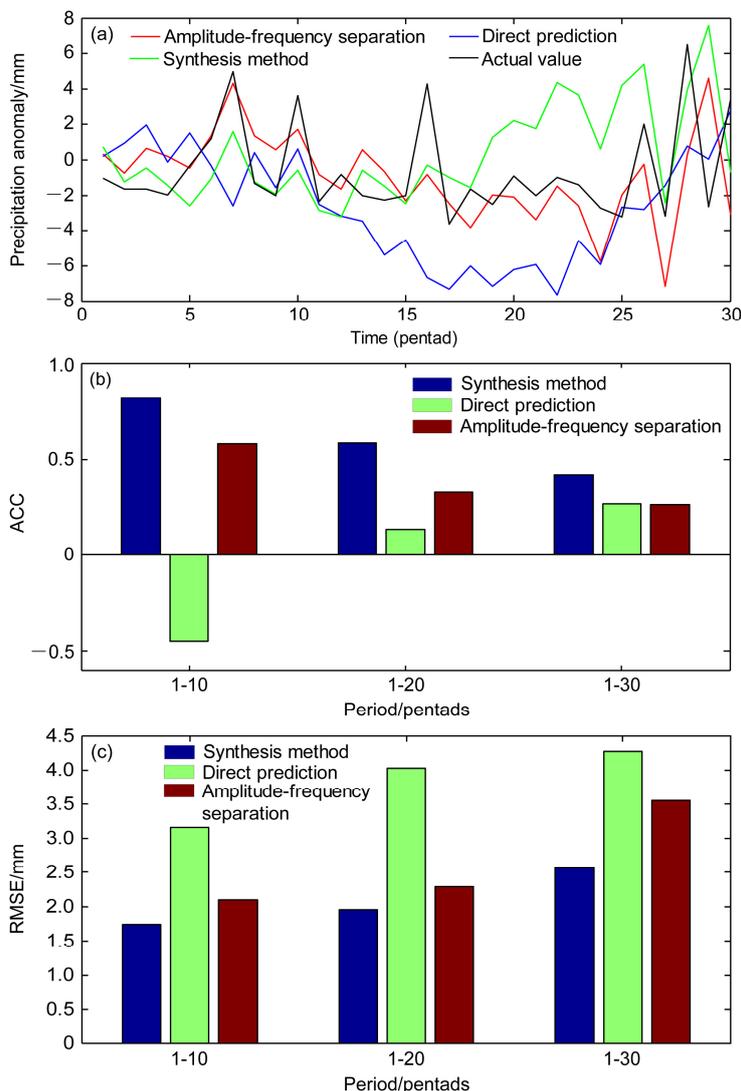


图 11 同预测方法下降水距平预测结果的 (a) 对比、(b) 相关系数以及 (c) 均方根误差

Fig. 11 Predictions of precipitation anomalies based on different prediction methods: (a) Comparison of predictions of precipitation anomalies; (b) correlation coefficients of precipitation anomalies; (c) the RMSE of precipitation anomalies

依赖于任何初始变量值。

针对降水序列中不同频谱分量分类预测，对复杂多变的高频分量采用幅频分离预测，而对变化规律相对简单的低频分量和剩余项采用直接预测法，这种合成预测法的预测效果相比单纯的幅频分离和直接预测方法都更具优势。本文预测试验虽是针对降水，对其它气象要素一样具有普适性，但是该方法的有效预测时效，以及与其他预测方法的进一步比较尚待以后深入研究。

参考文献 (References)

Boashash B. 1992. Estimating and interpreting the instantaneous frequency

of a signal. I. Fundamentals [J]. Proc. IEEE, 80 (4): 520–538, doi:10.1109/5.135376.

Boudraa A O, Cexus J C, Salzenstei F, et al. 2004. IF estimation using empirical mode decomposition and nonlinear Teager energy operator [C]// IEEE First International Symposium on Control, Communications and Signal Processing. Hammamet, Tunisia: IEEE, 45–48, doi:10.1109/ISCCSP.2004.1296215.

Cristianini N, Shawe-Taylor J. 2000. An introduction to support vector machines and other kernel-based learning methods [M]. Cambridge university press.

丑纪范. 2003. 短期气候预测的现状 问题与出路(一) [J]. 新疆气象, 26 (1): 1–4. Chou Jifan. 2003. Short term climatic forecast: Present condition, problems and way out [J]. Xinjiang Meteorology (in Chinese), 26 (1): 1–4, doi:10.3969/j.issn.1002-0799.2003.01.001.

- 胡爱军, 安连锁, 唐贵基. 2008. HILBERT-HUANG 变换端点效应处理新方法 [J]. 机械工程学报, 44 (4): 154–158. Hu Aijun, An Liansuo, Tang Guiji. 2008. New process method for end effects of HILBERT-HUANG transform [J]. Chinese Journal of Mechanical Engineering (in Chinese), 44 (4): 154–158, doi:10.3321/j.issn:0577-6686.2008.04.028.
- 黄大吉, 赵进平, 苏纪兰. 2003. 希尔伯特-黄变换的端点延拓 [J]. 海洋学报, 25 (1): 1–11. Huang Daji, Zhao Jinping, Su Jilan. 2003. Practical implementation of the Hilbert-Huang transform algorithm [J]. Acta Oceanologica Sinica (in Chinese), 25 (1): 1–11, doi:10.3321/j.issn:0253-4193.2003.01.001.
- 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]. Proc. Roy. Soc. A Math. Phys. Eng. Sci., 454 (1971): 903–995, doi:10.1098/rspa.1998.0193.
- 金龙, 秦伟良, 姚华栋. 2000. 多步预测的小波神经网络预报模型 [J]. 大气科学, 24 (1): 79–86. Jin Long, Qin Weiliang, Yao Huadong. 2000. A multi-step prediction model of wavelet neural network [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 24 (1): 79–86, doi:10.3878/j.issn.1006-9895.2000.01.08.
- 吕金虎. 2002. 混沌时间序列分析及其应用 [M]. 武汉: 武汉大学出版社, 1–10. Lv Jinhui. 2002. Chaotic Time Series Analysis and Its Application (in Chinese) [M]. Wuhan: Wuhan University Press, 1–10.
- Manuca R, Savit R. 1996. Stationarity and nonstationarity in time series analysis [J]. Physica D: Nonlinear Phenomena, 99 (2–3): 134–161, doi:10.1016/S0167-2789(96)00139-X.
- 毛宇清, 王咏青, 王革丽. 2007. 支持向量机方法应用于理想时间序列的预测研究 [J]. 气候与环境研究, 12 (5): 676–682. Mao Yuqing, Wang Yongqing, Wang Geli. 2007. An application study on prediction and analysis for ideal time series based on the SVM method [J]. Climatic and Environmental Research (in Chinese), 12 (5): 676–682, doi:10.3878/j.issn.1006-9585.2007.05.10.
- 沈路, 周晓军, 张志刚, 等. 2009. Hilbert-Huang 变换中的一种端点延拓方法 [J]. 振动与冲击, 28 (8): 168–171. Shen Lu, Zhou Xiaojun, Zhang Zhigang, et al. 2009. A boundary-extension method in Hilbert-Huang transform [J]. Journal of Vibration and Shock (in Chinese), 28 (8): 168–171, doi:10.3969/j.issn.1000-3835.2009.08.038.
- 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.
- Takens F. 1981. Detecting strange attractors in turbulence [M]//Rand D, Young L S. Dynamical Systems and Turbulence, Warwick 1980. Berlin, Heidelberg: Springer-Verlag, 366–381, doi:10.1007/BFb0091924.
- 王革丽, 杨培才, 吕达仁. 2004. 场时间序列预测方法及其预测能力的试验分析 [J]. 大气科学, 28 (4): 536–544. Wang Geli, Yang Peicai, Lü Daren. 2004. Method of spatio-temporal series and tests analysis on its predictable skill [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 28 (4): 536–544, doi:10.3878/j.issn.1006-9895.2004.04.06.
- 王革丽, 杨培才, 毛宇清. 2008. 基于支持向量机方法对非平稳时间序列的预测 [J]. 物理学报, 57 (2): 714–719. Wang Geli, Yang Peicai, Mao Yuqing. 2008. On the application of non-stationary time series prediction based on the SVM method [J]. Acta Phys. Sinica (in Chinese), 57 (2): 714–719, doi:10.3321/j.issn:1000-3290.2008.02.013.
- 王革丽, 杨培才, 卞建春, 等. 2011. 一个包含外强迫因子的非平稳时间序列的预测方法 [J]. 科学通报, 56 (25): 2107–2110. Wang Geli, Yang Peicai, Bian Jianchun, et al. 2011. A novel approach in predicting non-stationary time series by combining external forces [J]. Chinese Science Bulletin, 56 (28–29): 3053–3056, doi:10.1007/s11434-011-4638-1.
- Wang G L, Yang P C. 2005. A compound reconstructed prediction model for nonstationary climate processes [J]. Int. J. Climatol., 25 (9): 1265–1277, doi:10.1002/joc.1158.
- Wang G L, Yang P C, Zhou X J. 2013. Nonstationary time series prediction by incorporating external forces [J]. Adv. Atmos. Sci., 30 (6): 1601–1607, doi:10.1007/s00376-013-2134-z.
- Wu Z H, Huang N E, Wallace J M, et al. 2011. On the time-varying trend in global-mean surface temperature [J]. Climate Dyn., 37 (3–4): 759–773, doi:10.1007/s00382-011-1128-8.
- 玄兆燕, 杨公训. 2008. 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, 34 (1): 97–101. Xuan Zhaoyan, Yang Gongxun. 2008. Application of EMD in the atmosphere time series prediction [J]. Acta Automatica Sinica (in Chinese), 34 (1): 97–101, doi:10.3724/SPJ.1004.2008.00097.
- 杨培才, 卞建春, 王革丽, 等. 2003. 气候系统的层次结构和非平稳行为: 复杂系统预测问题探讨 [J]. 科学通报, 48 (13): 1470–1476. Yang Peicai, Bian Jianchun, Wang Geli, et al. 2003. Hierarchy and nonstationarity in climate systems: Exploring the prediction of complex systems [J]. Chinese Science Bulletin, 48 (19): 2148–2154, doi:10.1360/03wd0175.
- Yang P C, Wang G L, Bian J C, et al. 2010. The prediction of non-stationary climate series based on empirical mode decomposition [J]. Adv. Atmos. Sci., 27 (4): 845–854, doi:10.1007/s00376-009-9128-x.
- 张彬, 金莲姬, 王革丽. 2014. 非平稳时间序列的区域预测研究 [J]. 气候与环境研究, 19(1): 89–96. Zhang Bin, Jin Lianji, Wang Geli. 2014. Regional prediction of non-stationary time series [J]. Climatic Environ. Res. (in Chinese), 19 (1): 89–96, doi:10.3878/j.issn.1006-9585.2012.12157.
- 张军峰, 胡寿松. 2007. 基于一种新型聚类算法的 RBF 神经网络混沌时间序列预测 [J]. 物理学报, 56 (2): 713–719. Zhang Junfeng, Hu Shousong. 2007. Chaotic time series prediction based on RBF neural networks with a new clustering algorithm [J]. Acta Phys. Sinica (in Chinese), 56 (2): 713–719, doi:10.3321/j.issn:1000-3290.2007.02.019.
- 郑祖光, 刘莉红. 2010. 经验模态分析与小波分析及其应用 [M]. 北京: 气象出版社, 42–43. Zheng Zuguang, Liu Lihong. 2010. Empirical Mode Analysis and Wavelet Analysis and Its Application (in Chinese) [M]. Beijing: China Meteorological Press, 42–43.
- 朱佳, 王振会, 金天力, 等. 2010. 基于小波分解和最小二乘支持向量机的大气臭氧含量时间序列预测 [J]. 气候与环境研究, 15 (3): 295–302. Zhu Jia, Wang Zhenhui, Jin Tianli, et al. 2010. Combination of wavelet decomposition and least square support vector machine to forecast atmospheric ozone content time series [J]. Climatic and Environmental Research (in Chinese), 15 (3): 295–302, doi:10.3878/j.issn.1006-9585.2010.03.09.