

Kalman滤波技术在海表温度预测中的应用

吉进喜, 张立凤, 郭 渊

(解放军理工大学气象学院, 江苏 南京 211101)

摘要:以 EOF 分解方法为基础,把 AR 模型和 Kalman 滤波方法相结合,建立了海表温度的预报模型。首先对历史时间序列资料进行 EOF 分解,在此基础上,利用时间权重系数建立 AR(2)模型,并对此模型参数进行了改进,作为 Kalman 滤波的状态方程。然后用 Kalman 滤波方法对时间权重系数进行了滤波预测,并引入集合预报的思想对 SST 预测结果进行了重构,并与实况资料进行了相关性分析。以太平洋、印度洋、大西洋三大洋的热带海域为例进行了预测试验。试验结果表明,预测效果较好,相关系数平均达到了 98% 以上,而残差方差在 0.5 以内。

关键词: EOF 分解; AR 模型; Kalman 滤波; SST 预测

中图分类号: P457 **文献标识码:** A **文章编号:** 1003-0239(2010)03-0059-07

1 引言

SST (SST) 是重要的海洋水文要素之一, 它的变化对海洋和大气的物理状况和性质影响很大, 特别是对全球气候的变化和异常都有很大的影响。对 SST 的变化规律进行研究和预测是大气和海洋科学的一项重要内容, 在海洋开发和海洋经济上都有重要意义。Zhang Caiyun 等曾用 EOF 分解方法得到海表温度变化的 DV 模态、ENSO 模态、AC 模态等几个主要模态, 这些模态揭示了 SST 变化的时空特征^[1]。徐建军等通过对 50 年的印度洋、太平洋 SST 资料进行 EOF 分解, 也得出了近似的结论, 他认为 SST 的第一模态与年代际尺度变化相对应, 第二模态与 ENSO 事件相对应, 并模拟分析了其对中国降水的影响^[2]。董兆俊、滕军等在 SST 预测中引入混沌理论, 将相空间重构理论与模糊神经网络相结合, 建立了 SST 预测模型。实际的预测结果表明, 该模型预报精度较高, 但其预报区域较小, 对业务指导能力不是很强^[3]。Aiming Wu 采用多层神经网络方法建立了非线性的海表温度预报模型, 使得相关系数和均方

根误差较线性模型得到了极大的提高^[4]。侯瑞科曾采用平稳时间序列分析的方法, 建立了海表温度距平的回归预报方程^[5]。刘科峰、张韧等曾用经验正交分解和卡尔曼 (Kalman) 滤波相结合的方法, 对 T106 模式预报的位势高度误差进行了修正, 提高副热带高压的预报准确率^[6]。本文以 EOF 分解为基础, 采用 AR 模型和 Kalman 滤波相结合的方法建模分析海表温度的变化规律, 引入随机因子, 进而重构周平均 SST, 很好的预测了赤道太平洋、印度洋和大西洋三大洋的海表温度。

2 海表温度预测模型的建立

预测模型的建立是以 EOF 分解为基础, 以 Kalman 滤波为预报方法, 通过对海表温度 EOF 分解后的时间演变系数进行 Kalman 滤波预报, 得到预报的时间演变系数, 再与 EOF 分解得到的空间模态进行重构, 最后得到周平均 SST 场的预报结果, 具体流程如下:

2.1 周平均海表温度场的时空分解

采用经验正交分解方法 (EOF) 对多年周平均海

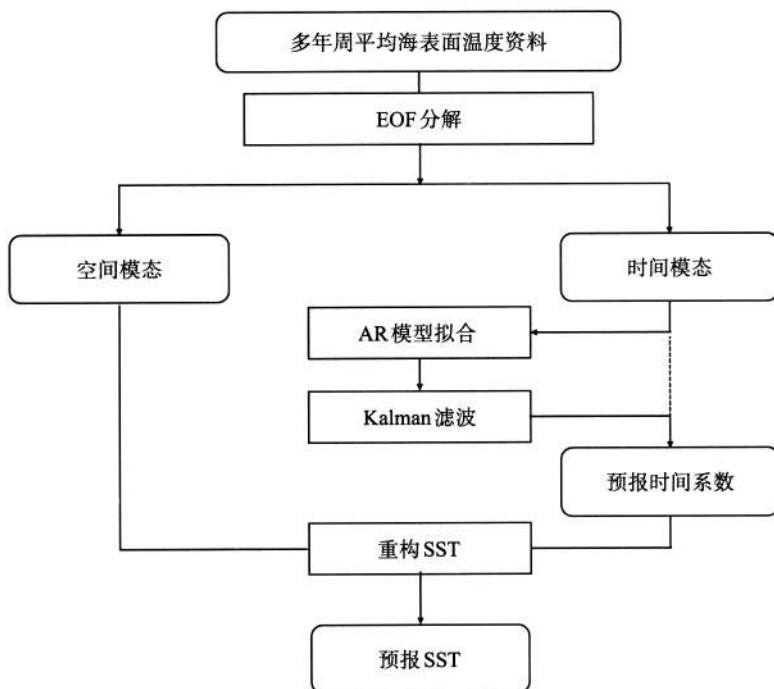


图1 预测模型建立的流程图

表面温度场进行时空分解。通过 EOF 分解, 可以把周平均 SST 距平场序列分解成各空间模态和时间模态乘积的累加, 公式^[7]如下:

$$X = \sum_{i=1}^N V_i T_i \quad (1)$$

其中, X 为周平均 SST 距平场, V_i 为 EOF 分解的第 i 个主要模态, T_i 为 EOF 分解的第 i 个主要模态的时间权重系数, N 为保留的模态数, 文中取 30。

物理量场经过 EOF 分解得到该物理量场的主要空间模态和时间模态。其中空间模态保留了物理量场空间分布的主要信息, 时间模态揭示了对应空间模态的时间演变信息, 对于空间模态由于其不随时间变化, 其不是预测关心的问题, 而时间系数的演变是预测要解决的主要问题。

2.2 时间权重系数的线性自回归模型-AR 模型

为了能够在预测物理量场时充分反映出空间分布信息, 保证预测结果不失真, 本文选择对时间权重系数进行滤波及预测, 保留空间模态不变。

对 SST 进行 EOF 分解得到的时间权重系数为一平稳时间序列 $\{T_k\}$, 对于平稳时间序列 $\{T_k\}$,

可建立线性自回归模型如下^[8]:

$$T_k = b_1 * T_{k-1} + \dots + b_p * T_{k-p} + \varepsilon_k \quad (2)$$

本文在取 2 阶自回归模型的基础上进行了改进。考虑到任意时刻的 SST 是由上一时刻的 SST 及在此时段内的变化决定的, 所以 (2) 式可改写为如下形式:

$$T_{k+1} = a * T_k + b * \text{delta} T_k + \varepsilon \quad (3)$$

$$\text{delta} T_k = T_k - T_{k-1} \quad (4)$$

其中, T 为 SST 距平 EOF 分解后的时间权重系数, 其只是时间的函数, T_k 为第 k 个时次的权重系数; $\text{delta} T_k$ 为 $k-1$ 至 k 时次时间权重系数的变化, 近似表示 k 至 $k+1$ 时段内的变化; 而 a 反映了前一个时次 SST 对后一时次的贡献, b 反映了第 k 个时次 SST 变化对后一时次的贡献, ε 近似反映了时间权重系数的两阶截断。

2.3 Kalman 滤波技术的应用

Kalman 滤波是采用最小均方误差准则的最佳滤波器, 它首先用于对随机过程的参量估计, 并很快地在各种最佳滤波和最佳控制问题中获得了广泛地应用。

Kalman滤波的出发点是建立被研究的系统的如下动态模型:

$$x_k = \Phi_{k,k-1}x_{k-1} + U_k, [k \geq 3] \quad (5)$$

$$y_k = H_k x_k + V_k \quad (6)$$

其中, $x_k = \begin{pmatrix} T_k \\ \text{delta}T_k \end{pmatrix}$ 是时间权重系数变化的状态矢量, Kalman滤波的目的在于对状态矢量进行最佳估计; $U_k = \begin{pmatrix} c+u_k \\ f \end{pmatrix}$ 为模式误差, 其中 u_k 为白噪声; $\Phi_{k,k-1} = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ 为系统转移矩阵; $y_k = \begin{pmatrix} T_k \\ \text{delta}T_k \end{pmatrix}$ 为测量矢量; H_k 为测量矩阵, 本文中取为单位阵; $V_k = \begin{pmatrix} v_k \\ 0 \end{pmatrix}$ 为测量噪音矢量, v_k 为白噪声。

Kalman滤波就是对由(1~2)式所描述的系统, 求状态矢量 x_k 的最佳估计值 \hat{x}_k , 使估计误差方差最小, 这里使用的是最小均方误差准则。

已知在 $k-1$ 时次的最佳估计 \hat{x}_{k-1} 及相应的方差矩阵

$$P_{k-1} = E(e_{k-1}e_{k-1}^T) = E[(x_{k-1} - \hat{x}_{k-1})(x_{k-1} - \hat{x}_{k-1})^T] \quad (7)$$

根据 $k-1$ 时次的 \hat{x}_{k-1} , 由(5)式可得 k 时次的预测值 $\hat{x}_{k,k-1}$:

$$\hat{x}_{k,k-1} = \Phi_{k,k-1}\hat{x}_{k-1} + U_k \quad (8)$$

相应的预测方差矩阵为:

$$P_{k,k-1} = \Phi_{k,k-1}P_{k-1}\Phi_{k,k-1}^T + \sigma_u^2 \quad (9)$$

根据 $\hat{x}_{k,k-1}$ 及(2)式可得到 k 时次的预测值:

$$y_k = H_k \hat{x}_{k,k-1} \quad (10)$$

根据实际得到的 k 时次测量值的预测值:

$$\hat{x}_k = \hat{x}_{k,k-1} + K(y_k - H_k \Phi_{k,k-1}\hat{x}_{k-1}) \quad (11)$$

其中, K 为增益矩阵。

可以证明, 对于最佳处理有

$$K_k = P_{k,k-1}H_k^T(H_kP_{k,k-1}H_k^T + \sigma_v^2)^{-1} \quad (12)$$

$$P_k = (I - K_kH_k)P_{k,k-1} \quad (13)$$

总的说来, 对于(1~2)式所表示的动态系统, (7~12)式就是它的Kalman滤波方程。

将AR(2)模型所建立的(15)式作为Kalman滤

波过程中的状态方程(1), 分两步实施对时间权重系数的滤波和预测:

第一步: 滤波过程。在对历史时间权重系数进行滤波的过程中, 为了更准确的预测, 本文采用自适应过程, 即用EOF分解得到的时间权重系数 T_k 代替(10)式中的预测值 y_k , 选用临近的52周, 即约一个周期的历史资料进行滤波。按此方法调整增益矩阵, 能有效提高历史拟合率, 为预测打下良好的基础。

第二步: 预测过程。由(7~12)式可对未来时间权重系数进行预测, 由(10)式可得到未来任意时次的最优估计 \hat{x}_k 。

2.4 SST场重构

根据kalman滤波所得到的时间权重系数的最佳估计, 由(13)式可以得到预测物理量距平场的最佳估计, 此时, \hat{X} 为预测得到的海表温度距平场, \hat{T}_i 为第 i 个模态的时间权重系数的最佳估计。

并进一步得到预测SST场:

$$SST_{re} = SST_{mean} + \hat{X} \quad (14)$$

SST_{re} 表示重构海表温度场, SST_{mean} 表示平均海表温度场。

2.5 集合预报思想的引入

考虑到SST演变的随机性质, 在Kalman滤波技术预测SST时, 即在每一步预测时, 都引入了近似表达SST变率的随机扰动, 不确定随机量的引入, 更加形象的描述了海温异常, 但同时必然带来预报结果的不确定性。为了消除预测效果的不确定性, 本文借鉴集合预报的思想^[9], 通过生成不同的随机量, 利用海表温度预测模型预测SST, 最后将各个预测SST等权平均, 得到SST的最佳预测场。经这样做可以滤掉预报中不确定随机扰动对结果的影响, 从而提高预报质量。试验证明, 集合思想的引进是可行的。

3 海表温度预测模型在海表温度预测中的应用

为了验证海表温度预测模型的可行性, 本文选取三个个例设计试验, 并从不同侧面对试验结果进行了深入分析。

3.1 资料说明

研究采用资料为美国海气总署发布的1981年12月~2007年7月共1344个周的 $1^\circ \times 1^\circ$ 周平均海表温度,下文中称为实况资料,其陆面温度为海表面温度资料插值所得。全球谱模式T106L19气候态海表温度资料,下文中称为气候态SST。

3.2 试验方案设计

为了验证建立的SST预报模型的可行性及有效性,本文利用1344周周平均SST资料,对2007年4月四个时次的SST进行预测,选取了太平洋

($120^\circ\text{E} \sim 80^\circ\text{W}$, $15^\circ\text{S} \sim 15^\circ\text{N}$),大西洋($45^\circ \sim 15^\circ\text{W}$, $30^\circ\text{S} \sim 30^\circ\text{N}$),印度洋($40^\circ \sim 110^\circ\text{E}$, $30^\circ\text{S} \sim 10^\circ\text{N}$)三个区域作为3个试验个例,具体见表1。

3.3 试验及结果分析

3.3.1 周平均海温EOF分解的时间权重系数分析

图2给出了三个大洋海表面温度EOF分解的第一模态对应的时间权重系数,从图上可以清楚看出,SST的时间权重系数呈现出有规律的振荡,约有50周的周期,这正是地球公转周期,而其不同的振幅反映了不同时间模态在SST场中的不同比重,揭示了不同年份独特的SST场结构。

表1 试验方案设计

试验个例	试验一	试验二	试验三
试验区域	赤道太平洋 ($120^\circ\text{E} \sim 80^\circ\text{W}$, $15^\circ\text{S} \sim 15^\circ\text{N}$)	赤道大西洋 ($45^\circ \sim 15^\circ\text{W}$, $30^\circ\text{S} \sim 30^\circ\text{N}$)	赤道印度洋 ($40^\circ\text{E} \sim 110^\circ\text{E}$, $30^\circ\text{S} \sim 10^\circ\text{N}$)

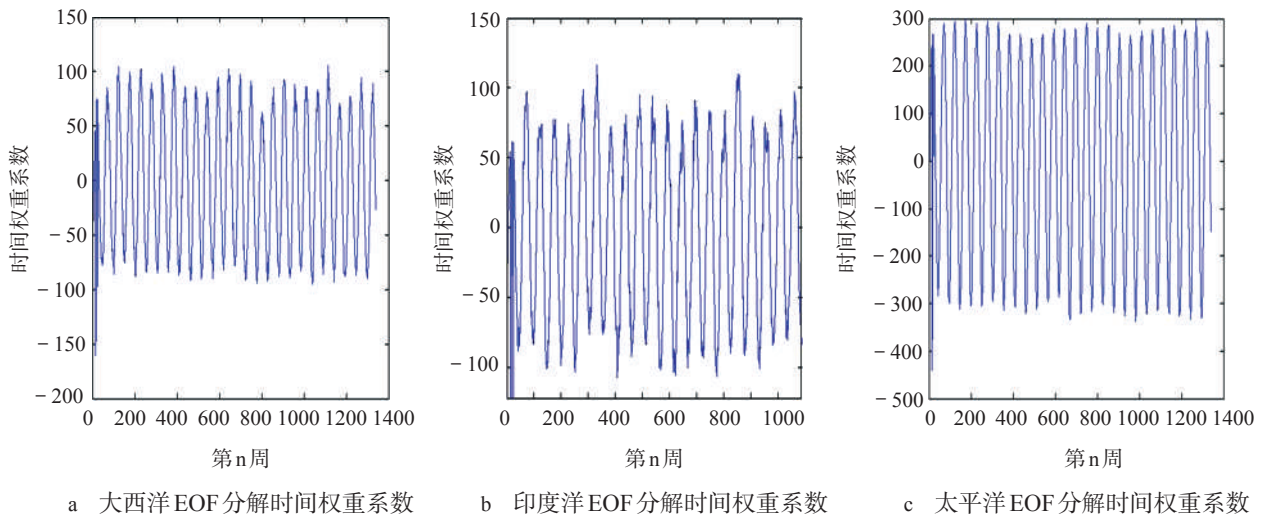


图2 三个试验中EOF分解第一主模态对应的时间系数

3.3.2 AR模型的参数拟合

利用EOF分解得到的第一模态的时间权重系数,用(15~16)式进行拟合,可得到各组试验的拟合参数见表2。

表2 时间权重系数的拟合参数

大西洋	1.7237	0.9832	-0.0001
印度洋	1.5932	0.9820	-0.0069
太平洋	1.9348	0.9854	-0.0088

从表2的结果可以看出, 的值很小,表明AR

的2阶模型已能够描述该时间序列的演变。

3.3.3 集合思想的应用

分析(1~2)式可以发现, 为模式误差, 为测量随机量,两者皆为白噪声。其取值范围对于模型的预报结果是有影响的,取值太小不能准确描述其异常变化,取值太大就会破坏了其规律变化。本文通过大量试验,发现随机量范围越小,预报效果越理想,但对于海表温度异常的描述能力越差,随机过程被“隐藏”于其线性变化过程中,所以随机量范围不能太小。为了确保海表

温度预测模型的稳定性,随机量取值范围不能太大。本文通过设计随机量取值范围从1~30不等的对比试验,发现选取范围为 $|U_k|, |V_k| < 5$ 时,既能描述海表温度变化的异常特性,又能保持预测模型的稳定性。

但同时也发现,随机量选取范围小时,预测结果很不稳定,预测SST与实况SST的相关系数起伏较大,不能很好的实现SST的预

测。本文取30个集成员,也即在Kalman滤波技术中生成30组不同的随机量,利用海表温度预测模型得到30个不同的预测SST,最后将这30组预报SST等权平均,得到SST的最佳预测场。

集合思想的引入是否能达到上述的效果?本文设计试验对比分析了集合前后第一周预测SST与实况SST相关系数及均方根误差(见图3)。从图中可以看出,集合后相关系数的提高不是很明

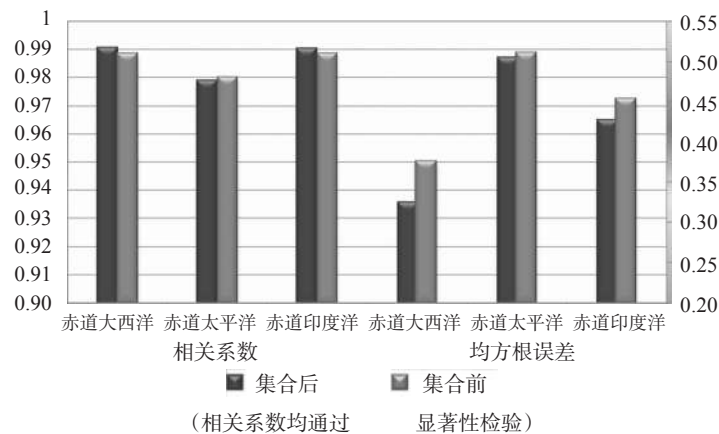


图3 集合前后第一周预测SST与实况SST相关系数及均方根误差

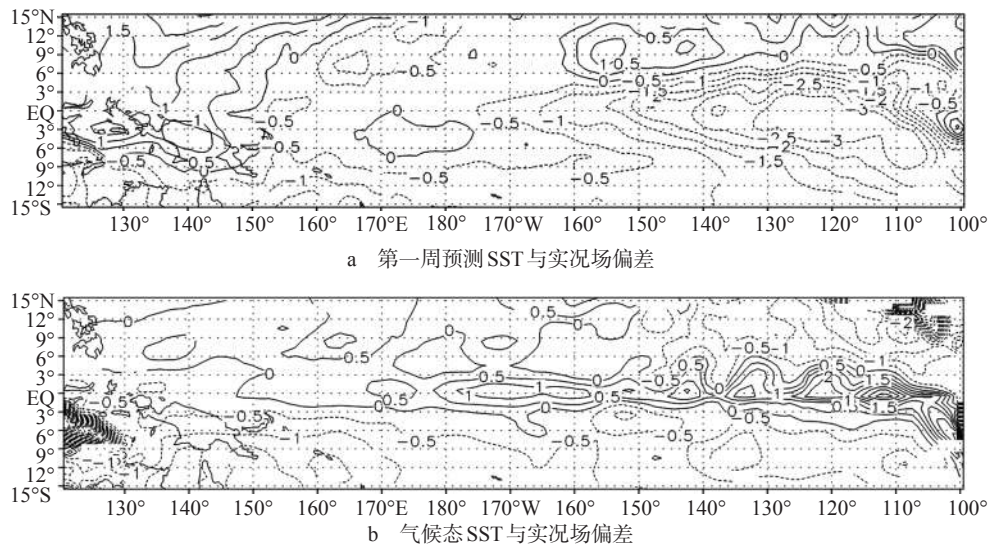


图4 赤道太平洋海表温度偏差场

显,除赤道太平洋集合前后预测效果相当外,其他两个海域均提高了0.2个百分点;从均方根误差更是可以看出,集合后误差明显降低,赤道大西洋误差减小了0.05℃。集合思想的引入确实提高预测效果。

3.3.4 预测SST与实况SST的相关分析

为了验证模型的有效性及其可行性,对预测的4个时次的SST场与实况资料进行相关分析。首先分析第一周预测SST与第一周实况SST的偏差场,以赤道太平洋为例(见图3),并与气候态SST

与实况 SST 的偏差场作对比, 定性揭示模型的预测效果。

由第一周预测 SST 与实况 SST 偏差场(见图 3a)可以看出, 赤道中太平洋及西太平洋南半球海域误差较小, 都在 1° 以内, 预测误差主要出现在赤道东太平洋, 最大偏差能达到 3° 以上, 菲律宾及以东也有大值中心, 误差达到 1.5° 。而从气候态 SST 与实况 SST 的偏差场(见图 3b)可看出, 赤道东太平洋偏差达到 5° 以上, 而菲律宾以东洋面偏差亦达到 2° 以上。模型对赤道东太平洋及西太平洋北赤道地区预测效果较差, 其原因除个例和资料对预测结果可能产生的影响外, 赤道太平洋地区复杂的海温状况也会使得预测效果较差, 但相对于气候态, 还是有相当大的改进。

用相关系数定量描述试验区域内预测海温与实况海温线性相关的密切程度, 用均方根误差反映预测 SST 与实况 SST 之间的整体差值大小^[7]。图 5 给出了四个时次 3 个海域预测 SST 与实况 SST 的相关系数及均方根误差, 从图中可以看出: 预测 SST 与实况 SST 的相关系数较高, 随着预报时间的

推移, 预测 SST 与实况 SST 相关系数是逐渐下降的, 除赤道太平洋第四周降至 0.963 外, 其他两个海域第四周相关系数均在 0.975 以上。而从均方根误差的分析来看, 三个试验预测 SST 与实况 SST 的均方根误差均控制在 1.0 以内, 误差相对较小, 与相关系数分析类似, 均方根误差是随预报时间增大的, 赤道大西洋误差增长较缓, 赤道印度洋误差增长最快, 这与不同海域海表温度变化的复杂程度是相关的。

通过以上分析可以发现, 对于海表温度状况较为复杂的区域, 预测效果仍较差, 模型的预测效果与选取区域、个例及资料是有关的。另外, 文中对海陆分界的处理不是很理想, 没有考虑海陆分布对海表温度的影响, 而是将规则区域内的海表温度作为研究的对象, 这样处理有利于资料的处理, 但同时也将造成地形对海表温度影响加大, 这也是亟需进一步解决的问题。但整体来看, 海表温度预测模型较为成功的实现了区域海表温度的预测。

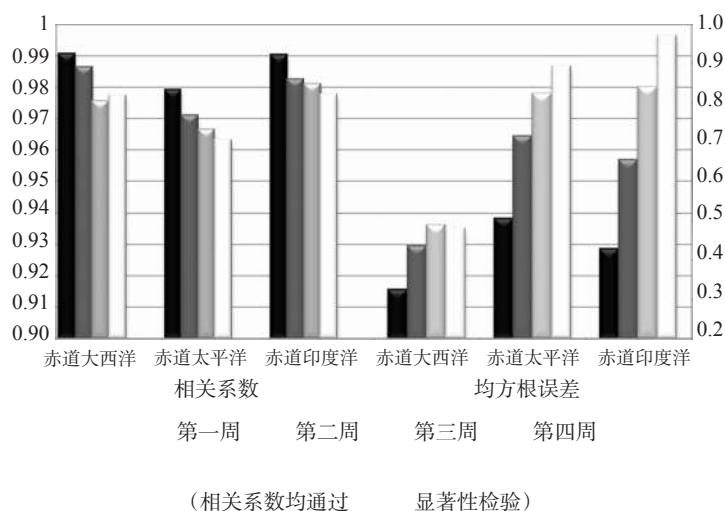


图 5 预测 SST 与实况 SST 相关系数及均方根误差

4 讨论

把 EOF 分解方法和 Kalman 滤波方法相结合, 建立了 SST 预测模型。以太平洋、印度洋、大西洋的热带海域为个例进行了预测试验。试验结果

表明: 把 EOF 分解和 Kalman 滤波技术相结合, 建立了 SST 预测模型, 计算量小, 较之海气耦合模式更经济, 更便捷。采用 EOF 分解后, 对时间系数进行预测, 可以保留 SST 变化的主要模态信息, 不会导致信息失真。采用自回归 AR 模型, 求得回归系数, 作为 Kalman 滤波的状态方程, 使得

状态方程更为明确。引入集合预报的思想对 SST 预测结果进行了重构,并与实况资料进行了相关性分析,预测效果较好,赤道地区相关系数达到了 97%以上,残差方差在 0.5 以内。但试验参数对区域的依赖性较高,为了使方法具有更广泛的普适性,一些关键技术还有待进一步研究。

参考文献:

- [1] Zhang Caiyun, Chen Ge, A global analysis of multi-mode sea surface temperature pattern[J], Acta Oceanologica Sinica, 2007, 26(1):12-22.
- [2] 徐建军, 王东晓. 印度洋-太平洋 SST 的年际、年代际异常及其对亚洲季风的影响[J], 海洋学报, 2000,20(3):34-43.
- [3] 董兆俊, 滕军等. 基于相空间重构与模糊神经网络耦合的 SST 预测模型[J], 热带海洋学报, 2008,27(4):73-76.
- [4] Aiming Wu, William W. Hsieh, Benyang Tang. Neural network forecasts of the tropical Pacific sea surface temperatures[J]. Neural Networks, 2006,19:145-154.
- [5] 侯瑞科. 平稳时间序列分析在海温预报中的应用[J]. 海洋预报, 1996, 13(1):41-45.
- [6] 刘科峰, 张韧等. EOF 分解与 Kalman 滤波相结合的副高位势场数值预报优化[J]. 解放军理工大学学报(自然科学版), 2006, 7(3):291-296.
- [7] 黄嘉佑. 气象统计与预报方法[M]. 北京, 气象出版社, 2004: 121-141.
- [8] 范剑青, 姚琦伟著. 非线性时间序列-建模、预报及应用[M]. 高等教育出版社, 2005:8-19.
- [9] Eugenia Kalnay. 大气模式、资料同化和可预报性[M]. 北京, 气象出版社, 2005:190-222.

Prediction of sea surface temperature by Kalman filtering

Ji Jin-xi, ZHANG Li-feng

(PLA University of Science and Technology, Nanjing Jiangsu 211101)

Abstract: Based on the EOF decomposition, combining the AR model and the Kalman filtering method, a sea surface temperature prediction model is established. First, the time-series of historical sst is decomposed by EOF, on the basis of which, AR(2) model is fitted from the time weighting Coefficient as the state equation in the Kalman filtering. Then the filtering and prediction of the time Coefficient are done by Kalman filtering method, in the process of which, the idea of ensemble prediction is introduced. In the end, the future sst is reconstructed. On the same time, correlation and root-mean-square error is analyzed. Three tests are done in the areas: tropical Pacific, tropical Atlantic, tropical Indian Ocean. The results show a good prediction effect, and the mean correlation coefficient is more than 98%.

Key Words: EOF decomposition; AR model; Kalman filter; SST prediction