

ENSO 集合预报系统的检验评价

郑 飞^{1,2} 朱 江¹ 王 慧³

1 中国科学院大气物理研究所国际气候与环境科学中心, 北京 100029

2 中国科学院研究生院, 北京 100049

3 国家气象中心, 北京 100081

摘 要 讨论了一个热带太平洋海气耦合集合预报系统集合预报的检验问题。该集合预报系统模式为一个中等复杂程度的耦合模式, 其中大气部分为统计模式, 海洋部分为动力模式。初始扰动利用集合 Kalman 滤波同化得到, 模式误差扰动由一个一阶马尔可夫随机微分方程生成, 预报集合样本为 100 个。利用 1995~2005 年的观测资料进行了确定性预报检验, 包括相关系数和均方根误差。在概率预报检验方面, 包括 Talagrand 概率分布、离散度、Brier 评分 (BS)、命中率以及空报率的统计检验, 并且根据检验结果对预报系统进行了初步评价。确定性检验表明, 集合样本均值的预报水平在热带中太平洋区域要高于热带东太平洋和沿岸区域。同时概率预报检验结果表明, 集合预报系统有较高的概率预报技巧, 对确定性预报是一个完善和补充。

关键词 集合预报 检验 评价

文章编号 1006-9585 (2007) 05-0587-08 **中图分类号** P732 **文献标识码** A

The Verifications for ENSO Ensemble Prediction System

ZHENG Fei^{1,2}, ZHU Jiang¹, and WANG Hui³

1 *International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029*

2 *Graduate University of the Chinese Academy of Sciences, Beijing 100049*

3 *National Meteorological Center, Beijing 100081*

Abstract The verification of the ensemble forecast results of a tropical Pacific air-sea coupled ensemble prediction system is discussed. The model of the ensemble prediction system is an intermediate coupled model. The atmospheric component is a statistical model, and the ocean component is a dynamical model. An ensemble Kalman filter (EnKF) data assimilation system is implemented to provide the initial ensemble. A first-order linear Markov stochastic model is used to represent model errors. The ensemble size is 100. The deterministic verification, including correlation and root mean square error, and the probabilistic verification, including Talagrand probability distribution, spread, Brier score (BS), and hit rate and false alarm rate are applied to the ensemble prediction system. Then some evaluations are made according to the verification results. The verifications are covering the period of 1995—2005. The deterministic verifications show that the prediction skill of the ensemble mean over the central tropical Pacific is particularly higher than that of the ensemble mean in the eastern basin and coastal region. The probabilistic verifications show that the ensemble system has relative high probabilistic skill and it is a complement to the deterministic forecast.

Key words ensemble prediction, verification, evaluation

收稿日期 2006-07-18 收到, 2007-05-06 收到修定稿

资助项目 国家自然科学基金资助项目 40437017 和 40221503、中国科学院方向性项目 KZCX2-YW-202 和国家重点基础研究发展规划项目 2006CB403600

作者简介 郑飞, 男, 1979 年出生, 博士, 助理研究员, 主要从事资料同化与集合预报研究。E-mail: zhengfei@mail.iap.ac.cn

1 引言

由于观测的不准确（包括仪器误差，观测点在空间和时间上不够密集引起的插值误差）和资料分析、同化处理中引入的误差，我们所得到的作为数值模式初值的初始场总是含有不确定性。同时数值预报模式的方程组是非线性系统，对初始误差的敏感性，以及模式物理过程参数化所带来的随机性误差，都会使得预报误差随预报时间大幅度增长。“集合预报”正是基于这些不确定性的考虑，从一组相关不多的初值集合出发，同时考虑了数值模式中物理过程的不确定性等因素，进而得到一组预报值的方法。

20 世纪 90 年代初期，大气模式的集合预报方法开始在 NCEP（美国国家环境中心，National Centers for Environmental Prediction）和 ECMWF（欧洲中期天气预报中心，European Centre for Medium-Range Weather Forecasts）投入业务运行。90 年代末以来，集合预报在一些国家广泛应用于日常预报中，研究也更加深入，相应的检验方法也逐渐发展起来。对于 ENSO（El Niño—南方涛动，El Niño-Southern Oscillation）的跨季度气候预测模式，其本身也存在着很多不确定性因素，由于海洋介质的缓变性和记忆性，海洋模式预报的不确定性主要是由物理过程的不确定性导致的。目前关于 ENSO 的跨季度集合预测，开展的相对比较少，很多原理和方法都是借鉴大气中所用到的集合方法^[1~3]。本文作者结合一个太平洋区域 ICM（中等复杂程度海气耦合模式，Intermediate Coupled Model）^[4~6]，发展了一个 ENSO 的跨季度集合预测系统^[7]。该系统的初始集合场是通过 EnKF（集合卡曼滤波，Ensemble Kalman Filter）同化系统^[8,9]生成的，物理过程的不确定性采用一个一阶线性马尔可夫随机模式模拟得到，共 100 个集合预报成员。

集合预报与传统的单一的确定性预报不同，它是不确定性预报，从一组数据中提取出有用信息。如何检验和评价集合预报系统至今还在研究探讨当中，本文借鉴气象上常用的一些集合预报检验方法应用于本文的集合预报的检验^[10]。

2 集合预报系统和检验资料介绍

本文的 ENSO 集合预报系统是在 Zhang 等^[4,5]发展的一个太平洋区域 ICM 基础上，采用 EnKF 同化方法^[8,9]为集合预报系统提供初始的集合样本场。试验比较表明该同化方法能够为集合预报系统提供比较精确的并与模式动力相协调的集合初值场，同时初始集合样本的标准差空间分布特征与观测误差的空间分布比较一致，说明初始样本的范围基本上可以将观测及其误差范围包含在内。对于本文的大样本而言，每个初始样本都可以代表一种观测出现的可能性，亦即每个样本的初始场都有同样的可能性代表海气的实际包含误差在内的真实状态。针对模式物理过程的不确定性，该集合预报系统采用一个一阶线性马尔可夫随机模式来模拟模式模拟过程中所存在的不确定性。本文所要检验的集合预报结果是从 1995~2004 年的 10 年每个月开始预报 12 个月，每个月开始的预报均有 100 个集合样本，这样总共就有 12 000 个预报结果需要检验。关于该集合预报系统的详细介绍，可参看文献 [7]。本文只对系统输出的 SSTA（海表温度距平，Sea Surface Temperature Anomaly）结果进行检验，因此所用到的检验观测资料是由 Smith 等^[11]重建的月平均 SST（海表温度，Sea Surface Temperature）资料得到 1995~2005 年 11 年的 SSTA 资料。另外文中与模式控制试验的结果进行了对比，控制试验的初始化方法具体见文献 [4, 5]。

3 确定性检验

一般情况下，由于计算平均的过程中能把不可能预报的随机信息过滤掉，集合平均预报通常比单个确定性预报，甚至比用更高分辨率模式所产生的单个预报准确^[12,13]，集合平均是集合预报最初级的应用。本文针对集合平均预报作确定性检验：相关系数和均方根误差。

3.1 相关系数

相关系数（Correlation）是用来衡量两个变量变化之间的关系的一个参数^[14]。可以用来检验两个变量的序列的变化是否一致，或者是两个场

的型态是否一致。

$$C = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (1)$$

图 1 给出的实线是预报样本均值与观测的 Niño3.4 指数的相关系数随预报时间的变化。可以看出预报样本均值在 Niño 3.4 区域与观测的变化吻合得比较好, 预报 12 个月相关系数仍在 0.7 左右, 并且每个月都要好于持续性预报。图 2 是预报样本均值与观测的相关系数的空间分布随预报时间的变化。可以看出, 两者在热带中太平洋的变化吻合得很好, 场型也是比较一致的。而热带东太平洋地区要稍差一些。

3.2 均方根误差

均方根误差 (Root Mean Square Error, 简称 RMSE) 是用来衡量两个变量之间平均差异的一个参数, 表示的是两个变量的平均偏离程度^[14]。

$$E_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}. \quad (2)$$

图 3 给出的实线是预报样本均值与观测的 Niño 3.4 指数的均方根误差随预报时间的变化。可以看出样本均值在 Niño 3.4 区域与观测的平均差异随预报时间在增大, 预报 12 个月均方根误差保持在 0.8 °C 左右, 并且每个月都要好于持续性预报。图 4 是预报样本均值与观测的均方根误差的空间分布。可以看出, 两者在中太平洋的平均差异要比东太平洋地区低很多。而东太平洋地区

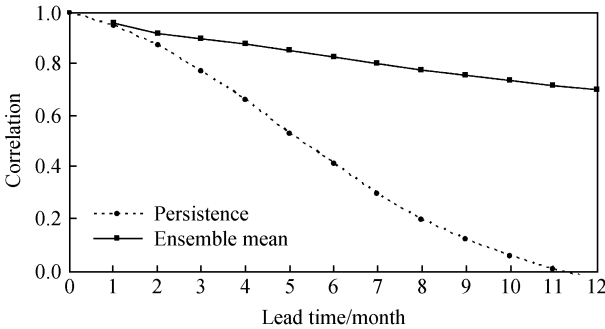


图 1 预报样本均值与观测的 Niño 3.4 指数相关系数随预报时间的变化和持续性预报
Fig. 1 Anomaly correlations of the Niño 3.4 SST anomalies for the model ensemble mean hindcast and for persistence, are shown as functions of lead time

特别是靠近南美沿岸的差异是最明显的。

4 概率预报检验

集合预报系统也是一种概率预报系统。概率

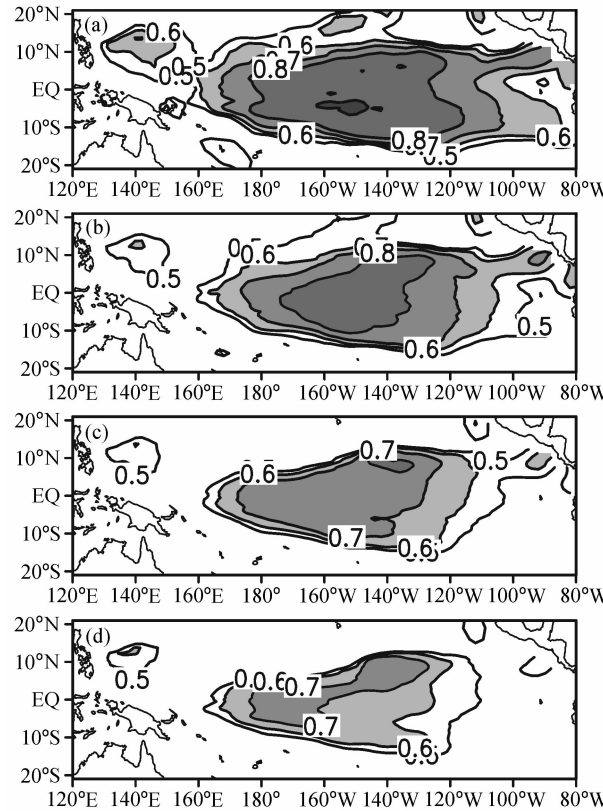


图 2 预报样本均值与观测的相关系数的水平分布: (a)、(b)、(c) 和 (d) 分别为预报第 3、6、9 和 12 个月, 等值线间隔 0.1
Fig. 2 Horizontal distributions of anomaly correlations between observed and predicted ensemble mean SST anomalies at 3 (a), 6 (b), 9 (c), and 12 (d) months of lead time, respectively. The contour interval is 0.1

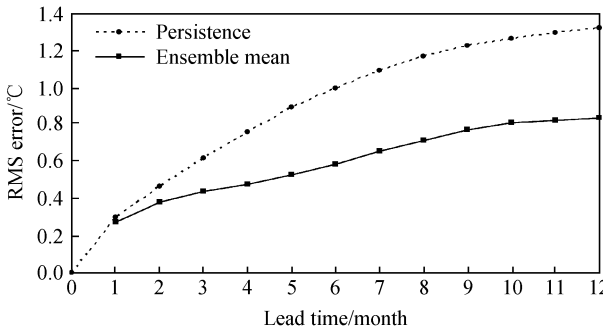


图 3 同图 1, 但为均方根误差
Fig. 3 Same as in Fig. 1, but for RMSE

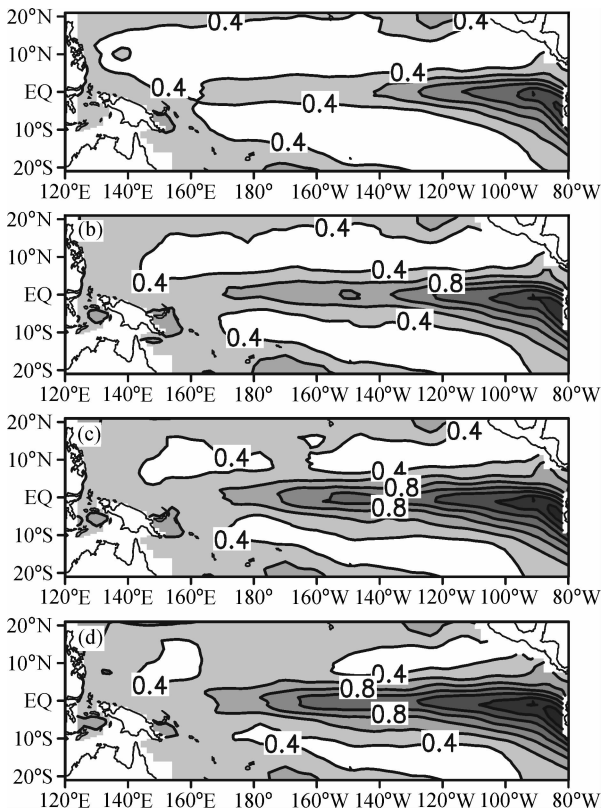


图4 同图2, 但为均方根误差。等值线间隔 $0.2\text{ }^{\circ}\text{C}$, 阴影区域为均方根误差大于 $0.4\text{ }^{\circ}\text{C}$ 的区域

Fig. 4 Same as in Fig. 2, but for RMSE. The contour interval is $0.2\text{ }^{\circ}\text{C}$, and the shaded area is above $0.4\text{ }^{\circ}\text{C}$

预报系统并不预报某种条件下的大气或海洋状态, 而是预报大气或海洋状态(事件)发生的概率分布。目前的天气预报中早已存在了这种事件发生的概率预报, 而且具有发展的趋势。这里按世界气象组织在1989年规定的方法对集合预报系统中的概率预报产品进行检验^[15]。

4.1 Talagrand 分布

Talagrand 认为一个“好”的集合预报系统的标准应是每个预报成员似乎以同样的概率发生; 换言之, 观测实况也应以相同的概率落在它们附近^[16,17]。从平均统计意义上看, 集合预报中的每个成员的准确率应大致相同, 即某个或某些预报成员不应该总是比其他成员准确。否则, 集合预报方法就失去意义了。这可称为“成员等同性”。

图5是针对所有模式格点的分析得到的在不同的预报月的 Talagrand 分布, 可以从图上看出发出预报样本的概率分布比较平直, 虽然落在两头的几率仍然是最高的(说明观测仍然有落在样本之外的情况出现, 即样本范围相对而言仍然有些偏小), 但是与理想状况的平均概率相比差距不大, 说明样本有很好的同一性, 没有特别明显的区分。

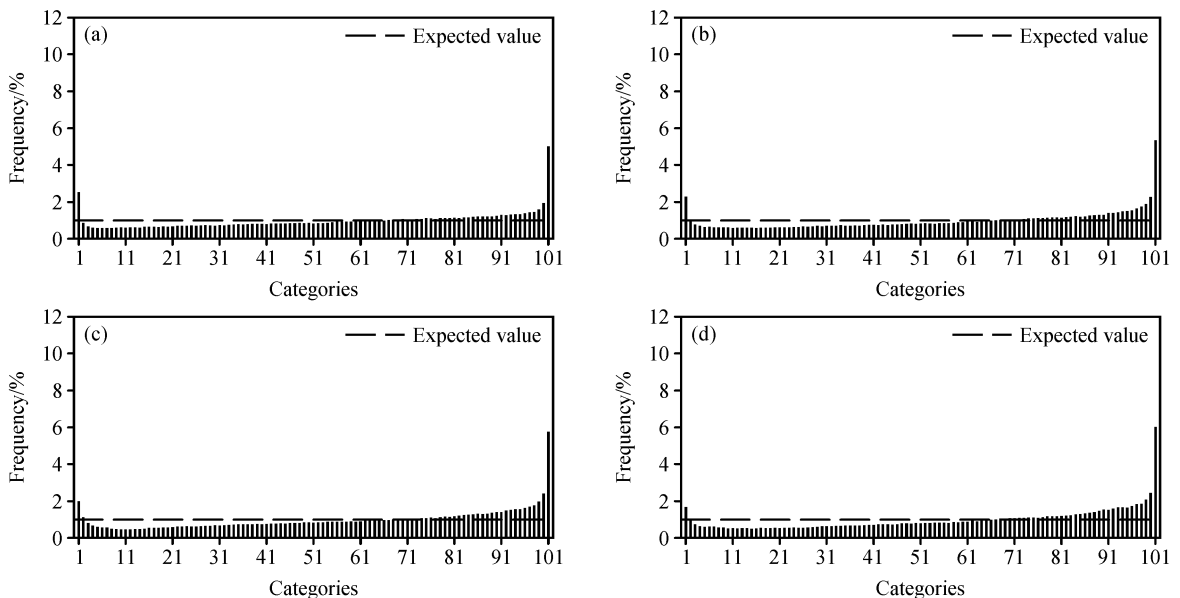


图5 Talagrand 分布: (a)、(b)、(c) 和 (d) 分别为预报第 3、6、9 和 12 个月。虚线表示平均概率, 横坐标为样本序号, 按照从小到大的顺序排列

Fig. 5 Talagrand diagram for the full ensemble SST anomaly hindcast: (a) 3-month lead time, (b) 6-month lead time, (c) 9-month lead time, and (d) 12-month lead time hindcasts. The dashed line marks the theoretical frequency for a perfectly reliable ensemble prediction system, the abscissa is the ensemble category, and defined by the ordered ensemble members from smallest to largest

相比不同月预报的 Talagrand 分布, 预报时间越短, Talagrand 分布越理想。

4.2 离散度

集合预报各成员的结果是离散的, 说明了未来模式大气或海洋的预报不确定性。按照 Buizza 定义的离散度 (Spread)^[18], 可以将其看作为各扰动预报与控制 (未扰动) 预报之间的平均距离。设有 N 个成员的集合预报为 $f_i (i=1, 2, 3, \cdots, N)$, 则集合预报的离散度 S 可用下式计算:

$$S(t)=\sqrt{\frac{1}{N}\sum_{i=1}^N[f_i(t)-f_0(t)]^2}, \quad (3)$$

其中, t 表示预报时效, 这里 f_0 是集合预报的平均值。这样 S 即为样本的标准差。

从图 6 可以看出, 在不同的预报月, Niño 3.4 指数的样本均值的均方根误差要大于平均的样本标准差。这说明在该区域或者说该指数的样本离散度仍然偏小, 样本的范围不能完全涵盖所有的观测, 亦即观测仍然会落在样本之外, 从前面的 Talagrand 图上也可以得到一致的结论。另外, Niño 3.4 指数样本均值的均方根误差要小于控制预报的均方根误差, 说明集合平均预报要比控制预报的可信度高。

4.3 事件分类

在进行以下两种检验之前, 需要先对所统计的事件进行分类 (Tericle category)。针对需要比较分析的预报结果而言, 需要定义模式每个网

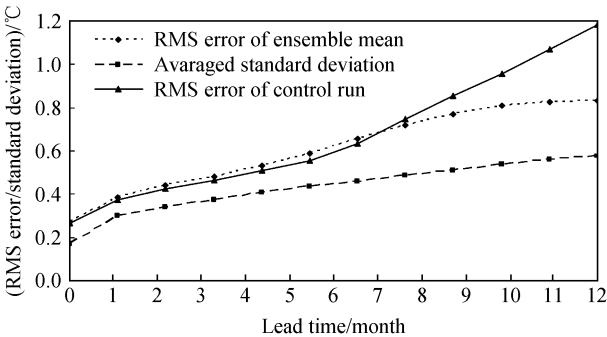


图 6 样本均值与观测的 Niño 3.4 指数的均方根误差随预报时间的变化、平均的样本 Niño 3.4 指数的标准差随预报时间的变化和控制试验结果与观测的 Niño 3.4 指数均方根误差随预报时间的变化

Fig. 6 Anomaly RMS error of the Niño 3.4 SST anomalies for the model ensemble mean hindcast and for control run, and averaged standard deviations of the model ensembles are shown as functions of lead time

格点上的 SST 异常值出现的暖事件 (Warm events, 事件出现在高百分位 66.6%~100.0%的范围)、冷事件 (Cold events, 事件出现在中百分位 33.3%~66.6%的范围) 以及正常事件 (Normal conditions, 事件出现在低百分位 0~33.3%的范围)。我们采用 1971~2000 年 (气候态) 的 ERSST (扩展重构的海表温度, Extended Reconstructed Sea Surface Temperature)^[11] 距平资料在每个网格点进行分析, 这里存在一个假定即每个格点的 SSTa 的 30 年时间分布都是符合正态分布的。具体的分类方法见图 7。

4.4 Brier 评分

Brier^[19] 定义了一种均方概率误差, 称之为 Brier 评分 (Brier Score, 简称 BS):

$$B_s=\frac{1}{N}\sum_{i=1}^N(f_i-O_i)^2, \quad (4)$$

其中, N 为二态分类事件的预报数; f_i 为事件发生的预报概率; 如果事件发生 $O_i=1$, 事件不发生 $O_i=0$ 。在这一形式中, 评分的取值范围是 0~1, 且越小越好, 即 $B_s=0$ 表示概率预报最佳, 预报

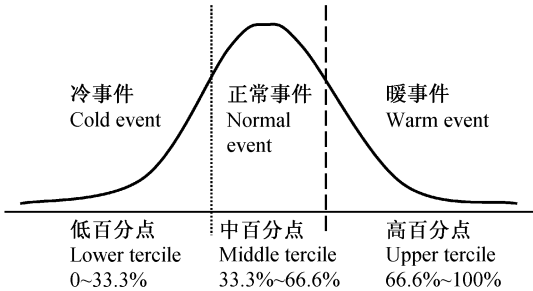


图 7 冷事件、正常事件和暖事件 3 种事件分类示意图
Fig. 7 A schematic illustration of tericle category for cold event, normal condition and warm event

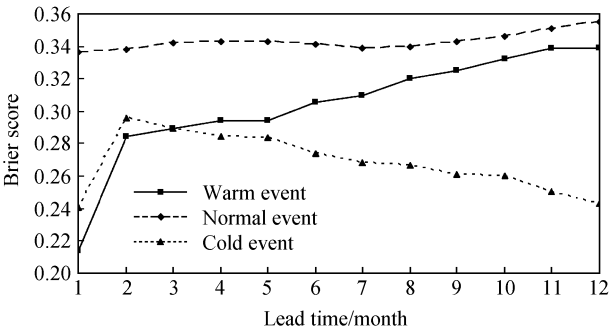


图 8 Niño 3.4 区 3 种事件的 Brier 评分
Fig. 8 Brier scores for three events in the Niño 3.4 region

正确; $B_s=1$, 表示评分最差, 预报失效^[20,21]。

图 8 给出的是 Niño 3.4 区所有格点关于 3 种事件平均的 Brier 评分。可以看出对于正常事件的 Brier 评分最低, 且随预报时间的变化不大; 对于暖事件的 Brier 评分要高于正常事件, 其随预报时间的增长而降低; 而对于冷事件的 Brier 评分是最高的, 其在预报 2 个月以后随预报时间的增长而增高, 这个现象主要是因为模式预报有偏冷的趋势, 亦即模式趋向于冷事件, 当冷事件发生时 B_s 就自然会低, 而当冷事件没有发生时 (发生暖事件或者正常事件), B_s 就相对要高一些。模式的预报过冷趋势也是需要在接下来的工作中进一步探讨的问题。

4.5 相对作用特征

相对作用特征 (Relative Operating Characteristic, 简称 ROC) 是信号探测理论在数值天气预报中的一种应用, 对二分类要素序列进行检验。在每个格点上, 考虑一个事件 (如暖事件) 发生或不发生两种状态。用实况去检验预报, 其结果必是下列情况之一: 预报准确, 正确否定, 漏报和空报^[22]。因此可以构成如表 1 的双态分类联列。

定义命中率 (H) 和假警报率 (F):

$$H = \frac{a}{a+c}, \quad (5)$$

$$F = \frac{b}{b+d}. \quad (6)$$

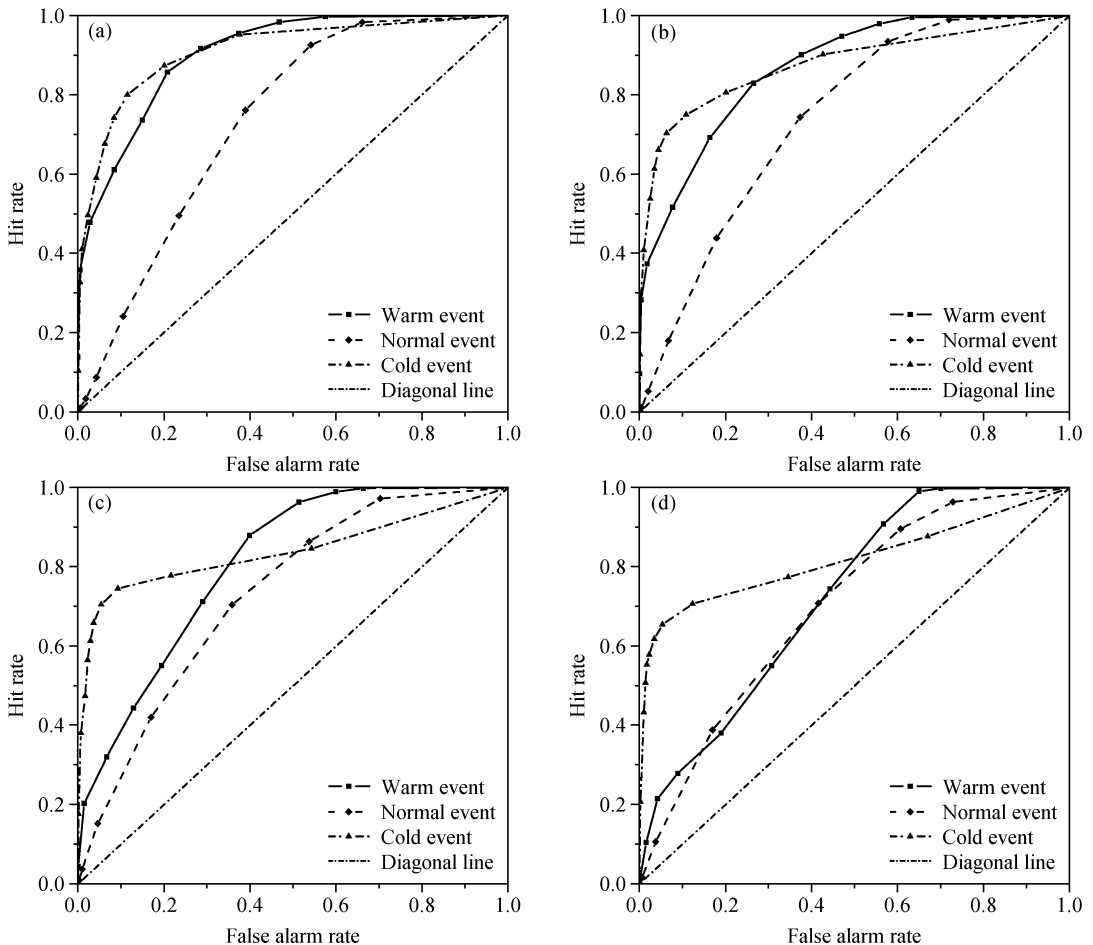


图 9 Niño 3.4 区 ROC 曲线: (a)、(b)、(c) 和 (d) 分别为预报第 3、6、9 和 12 个月。对角虚线为基准线, ROC 曲线落在对角线左上方认为是有预报技巧

Fig. 9 Niño 3.4 ROC curves for a lead time of (a) 3 months, (b) 6 months, (c) 9 months, and (d) 12 months. The dashed diagonal line is a benchmark line, and the system has prediction skill when the ROC curve lies top left of the diagram

表 1 双态分类联列
Table 1 Contingency

预报	观测	
	出现	不出现
出现	<i>a</i>	<i>b</i>
不出现	<i>c</i>	<i>d</i>

注: *a* 表示预报准确, *c* 表示漏报, *b* 表示空报, *d* 为正确否定

根据设置不同预报事件发生的概率临界值 (0~100%), 得到不同的命中率和假警报率可在笛卡尔坐标上绘成一曲线, 这曲线称之为 ROC 曲线。曲线落在对角线左侧意味着有预报技巧, 落在对角线上及右侧即无预报技巧。

图 9 是不同预报月的 3 种事件的 ROC 曲线。对于不同预报月, 集合预报系统在 Niño 3.4 区对于 3 种事件均是有概率预报技巧的, ROC 曲线均落在了对角线的左侧。相比而言, 在不同预报月, 对于冷事件的预报效果均是最好的, 暖事件次之, 而正常事件最差, 这和 Brier 评分是相同的结论。进一步, 可以得到 ROC 曲线所包含的面积, 即 ROC Area (简称 ROCA):

$$A_{\text{ROC}} = \int_0^1 H(x) dx = \sum_{i=1}^M (H_{i+1} + H_i) (F_{i+1} - F_i).$$

(7)

式中, M 为概率临界值的分类数, $A_{\text{ROC}} > 0.5$ 说明有预报技巧。图 10 为 3 种不同事件的 A_{ROC} 随预报时间的变化, 结论与图 9 比较类似, 3 种事件的 A_{ROC} 均在 0.5 以上, 同时冷事件的 A_{ROC} 值最大, 正常事件的 A_{ROC} 值最小。从另一个侧面描述了该集合预报系统的概率预报技巧。

5 结论

简介了本文作者等发展的 ENSO 集合预报系统简况和集合预报产品及其检验方法。并且对 1995~2005 年的集合预报结果进行了统计检验, 初步得到如下结果:

(1) 确定性检验表明, 作为集合预报的初始应用, 集合预报的样本均值在 Niño 3.4 区预报技巧较高, 与观测的变化吻合得比较好, 预报 12 个月平均的偏离程度为 0.8 °C 左右。热带中太平洋区域的相关系数和均方根误差的空间分布也是如此。而在热带东太平洋以及南美沿岸等

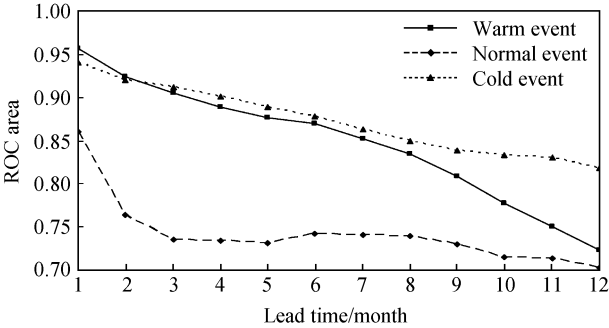


图 10 Niño 3.4 区 3 种事件的 ROC 面积
Fig. 10 ROC area for three events in the Niño 3.4 region

区域, 集合预报的样本均值与观测的相关系数要小于热带中太平洋, 均方根误差也在该区域出现最大值。

(2) 模式整个区域的第 3、6、9 和 12 个月集合预报的 Talagrand 分布直方图总体来说除两端之外, 分布还是比较均匀的。预报时间越短, Talagrand 分布越理想。

(3) Niño 3.4 指数的离散度检验表明, 集合平均预报要比控制预报的可信度高, 同时样本均值的均方根误差要大于平均的样本标准差。这说明了在 Niño 3.4 区域样本的范围不能完全涵盖所有的观测, 亦即观测仍然会落在样本之外。表明该预报系统的离散度仍然偏小。

(4) 通过对 Niño 3.4 区 3 种事件概率预报的 Brier 评分、命中率和空报率的统计检验以及 ROC 面积检验, 表明该集合预报系统对 3 种不同事件的概率预报都存在一定的概率预报技巧, 其中对冷事件的概率预报评分相对最高, 对正常事件的概率预报评分最低。

以上简单分析说明, 本文作者等发展的 ENSO 集合预报系统, 比原有的确定性预报模式模拟结果有一定的改善, 预报水平优于一些更加复杂的耦合环流模式^[23], 同时概率预报也是对确定性预报的一个完善和补充。但是通过分析也发现, 该集合预报系统对热带东太平洋区域预报技巧相对稍差一些, 这是模式本身物理过程所造成的, 希望能够在以后的工作中通过在随机过程中考虑预报偏差等方法来提高该地区的预报能力, 同时能够将数值天气集合预报中的一些方法^[24]应用到 ENSO 集合预报中做进一步的探讨。

参考文献 (References)

- [1] Kirtman P B. The COLA anomaly coupled model: Ensemble ENSO prediction. *Mon. Wea. Rev.*, 2003, **131**: 2324~2341
- [2] Kleeman R, Moore A M. A new method for determining the reliability of dynamical ENSO predictions. *Mon. Wea. Rev.*, 1999, **127**: 694~705
- [3] Moore A M, Kleeman R. Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, 1998, **124**: 557~584
- [4] Zhang R H, Zebiak S E, Kleeman R, et al. A new intermediate coupled model for El Niño simulation and prediction. *Geophys. Res. Lett.*, 2003, **30** (19): 2012, doi: 10.1029/2003GL018010
- [5] Zhang R H, Zebiak S E, Kleeman R, et al. Retrospective El Niño forecast using an improved intermediate coupled model. *Mon. Wea. Rev.*, 2005, **133**: 2777~2802
- [6] Zheng Fei, Zhu Jiang, Zhang R H, et al. Improved ENSO forecasts by assimilating sea surface temperature observations into an intermediate coupled model. *Adv. Atmos. Sci.*, 2006, **23** (4): 615~624
- [7] Zheng Fei, Zhu Jiang, Zhang Rong Hua, et al. Ensemble hindcasts of SST anomalies in the tropical Pacific using an intermediate coupled model. *Geophys. Res. Lett.*, 2006, **33**: L19604, doi: 10.1029/2006GL026994
- [8] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, 2003, **53**: 343~367
- [9] Evensen G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.*, 2004, **54**: 539~560
- [10] 皇甫雪官. 国家气象中心集合数值预报检验评价. 应用气象学报, 2002, **13**: 29~36
Huangfu Xueguan. The verification for ensemble prediction system of national meteorological center. *Quarterly Journal of Applied Meteorology* (in Chinese), 2002, **13**: 29~36
- [11] Smith T M, Reynolds R W. Improved extended reconstruction of SST (1854—1997). *J. Climate*, 2004, **17**: 2466~2477
- [12] Du J, Mullen S L, Sanders F. Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, 1997, **125**: 2427~2459
- [13] Tracton M S, Du J. Short-range ensemble forecasting (SREF) at the National Centers for Environment Prediction. Preprints of 12th Conference on Numerical Weather Prediction, Phoenix, Arizona, Amer. Meteor. Soc., 1998. 269~272
- [14] 黄嘉佑. 气象统计分析与预报方法. 北京: 气象出版社, 1990
Huang Jiayou. *Application of Statistic to Weather Forecast* (in Chinese). Beijing: China Meteorological Press, 1990
- [15] Stanski H R, Wilson L J, Burrows W R. Survey of common verification methods in meteorology. World Weather Watch Tech. Rept. No. 8, WMO/TD No. 358, WMO, Geneva, 1989. 114 pp
- [16] Talagrand O, Vautard R, Strauss B. Evaluation of probabilistic prediction systems. Proceedings, *ECMWF Workshop on Predictability*. 1997
- [17] Hamill T M. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, 2001, **129**: 550~560
- [18] Buizza R, Palmer T N. Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, 1998, **126**: 2503~2518
- [19] Brier G W. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, 1950, **78**: 1~3
- [20] Murphy A H. A new vector partition of the probability score. *J. Appl. Meteor.*, 1973, **12**: 595~600
- [21] Mason S J. On using “climatology” as a reference strategy in the Brier and ranked probability skill scores. *Mon. Wea. Rev.*, 2004, **132**: 1891~1895
- [22] Mason S J, Graham N E. Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, 1999, **14**: 713~725
- [23] 周广庆, 李旭, 曾庆存. 一个可供 ENSO 预测的海气耦合环流模式及 1997/1998 ENSO 的预测. 气候与环境研究, 1998, **3**: 349~357
Zhou Guangqing, Li Xu, Zeng Qingcun. A coupled ocean-atmosphere general circulation model for ENSO prediction and 1997/1998 ENSO forecast. *Climatic and Environmental Research* (in Chinese), 1998, **3**: 349~357
- [24] 蔡其发, 张立凤, 张铭. 中期数值天气预报的集合预报试验. 气候与环境研究, 1999, **4**: 365~374
Cai Qifa, Zhang Lifeng, Zhang Ming. The experiment of ensemble forecasting about numerical medium-range forecasting. *Climatic and Environmental Research* (in Chinese), 1999, **4**: 365~374