

DERF2.0 对全球气温、降水和环流的 预测性能评估

（第 1 季度）

国家气候中心气候研究开放实验室

赵崇博 吴捷 任宏利 左金清

2014 年 12 月 20 日

摘要

本次报告依据 WMO 推荐中国气象局业务标准，分别使用确定性预报评估方法，如空间相似系数（ACC）、时间相关系数（TCC）、平均误差方差（MSSS）等定量评估以及 Ps 评分、Pg 检验等定性评估方法，和概率预报评估方法 ROC，对 DERF2.0 月尺度的气温(t2m)、降水和 500hPa 位势高度和高低层纬向风场的预报性能进行检验评估，按照业务处时间节点安排，本次报告主要提供第 1 季度评估结果，与上次报告相比，主要增加了基于 DERF2.0 模式信息的实时 MJO 监测预测产品评估检验，主要得到以下结论：

1) DERF2.0 第 1 季度多年的气候态在某些地区存在系统性的漂移，如在欧亚大陆中高纬度东部地区气温偏高，我国大部分地区气温偏低，热带西太平洋暖池和 SPCZ 地区降水偏强。

2) 对要素确定性预报技巧评估表现出很多通性，如对 500hPa 高度场预报技巧最高，纬向风场次之，2m 气温场的技巧略低于环流但高于降水；除气温外，模式对热带地区的预报性能好于热带外，特别对降水而言；非等级的定量评估（ACC、TCC、MSSS）等表明，模式的预报技巧对初值非常敏感，随着 lead time 的延长模式预报性能迅速下降，但等级评估（同号率、Ps 评分等）则对初值的敏感性较弱。由于预报员进行短期气候预测时，往往不能及时获取最新初值的预报结果，因此在参考模式较长 lead time 的定量预报结果时需多加注意。

3) 具体到东亚地区而言，模式对东亚地区的位势高度场预报优于其他北半球热带外地区，但对降水和温度的预报技巧要低于其他地区，同时预报性能的年际变化很大，对我国东部的降水和气温预报技巧较高，对我国西部预报技巧明显偏低。

4) 概率预报检验 ROC 的结果表明，模式对高/低异常值出现概率的预报技巧高于常值的预报技巧，对气温场（T2m）在热带外地区的预报技巧相对较高，对我国中东部地区降水异常的预报技巧也高于其他地区，另外模式在热带地区对 500hPa 高度场的概率预报技巧最高，对赤道中太平洋的纬向风场的预报技巧也高于其他地区。模式较高的概率预报技巧表明模式具有相对较多的样本能够准确判断出各个变量场的高/低值事件出现与否，可为模式预报结果的校正提供参考。

5) 对延伸期降水预报技巧检验表明，模式对赤道中东太平洋降水的预报技巧较高，可持续 7-8 候，而对热带外预报技巧较低，预报技巧随时间的延长很快减小，在第 3 候时，模式对我国长江以南地区的预报技巧相对较高，对我国西部的预报技巧很低。从概率预报上看，模式对 AN/BN 事件的预报技巧类似，均高于对 NN 事件的预报技巧。

本次报告除了对第 1 季度的各个要素月尺度预报结果进行检验外，也初步探讨了基于 DERF2.0 模式信息的实时 MJO、AO\监测预测产品评估检验，这部分工作还需要继续深入下去。

目录

1 概述.....	1
2 资料和方法.....	1
3 月尺度评估结果.....	1
3.1 2m 气温 (t2m)	1
3.1.1 气候态对比.....	1
3.1.2 空间距平相似系数 (ACC)	2
3.1.3 距平符号一致率 (Pc)	3
3.1.4 趋势异常综合检验 (Ps)	4
3.1.5 分级评分检验 (Pg)	5
3.1.6 时间距平相关系数 (TCC)	7
3.1.7 平均方差技巧 (MSSS)	8
3.1.8 气温 (t2m) 评估小结.....	9
3.2 降水 (prec)	9
3.2.1 气候态对比.....	9
3.2.2 空间距平相似系数 (ACC)	10
3.2.3 距平一致率 (Pc)	12
3.2.4 趋势异常综合检验 (Ps)	12
3.2.5 分级评分检验 (Pg)	13
3.2.6 时间距平相关系数 (TCC)	14
3.2.7 平均方差技巧 (MSSS)	16
3.2.8 降水 (prec) 评估小结.....	17
3.3 500hPa 位势高度 (HGT 500)	17
3.3.1 气候态对比.....	17
3.3.2 时间距平相关系数 (TCC)	18
3.3.3 平均方差技巧 (MSSS)	20
3.3.4 500hPa 位势高度评估小结.....	21
3.4 200hPa 和 700hPa 纬向风场.....	21
3.4.1 气候态对比.....	21
3.4.2 时间距平相关系数 (TCC)	22
3.4.3 均方技巧评分 (MSSS)	24
3.4.4 纬向风场评估小结.....	25
4 概率预报评估 (ROC)	25
4.1 2m 气温 (t2m)	26
4.2 降水 (prec)	28
4.3 500hPa 位势高度.....	30

4.4 纬向风场.....	32
4.5 ROC 评估小结.....	36
5 对延伸期降水预报的检验评估.....	36
5.1 确定性预报检验.....	36
5.1.1 时间距平相关系数 (TCC)	36
5.1.2 均方技巧评分 (MSSS)	37
5.2 概率预报检验 (ROC/RD)	38
5.3 延伸期降水预报小结.....	41
6 基于 DERF2.0 实时预报产品.....	42
6.1 MJO 监测预测实时产品.....	42
6.2 AO 预测实时产品和实时检验.....	44
7 全文小结.....	45
主要参考文献.....	46

DERF2.0 对全球气温、降水和环流的预测性能评估

（第 1 季度）

1 概述

根据科技业务处工作安排的开放室业务维持项目，开放室参照 WMO 等国际推荐标准，对现有 DERF2.0（1991-2010 共 20 年）历史回报资料，从月尺度和月内尺度两个方面，对气温、降水和环流等要素和重要的季节内过程进行检验评估。本次报告提交第 1 季度（1、2、3 月）的评估结果，与第 4 季度报告相比，增加了基于 DERF2.0 模式信息的实时 MJO 监测预测产品评估检验，以供参考。

2 资料和方法

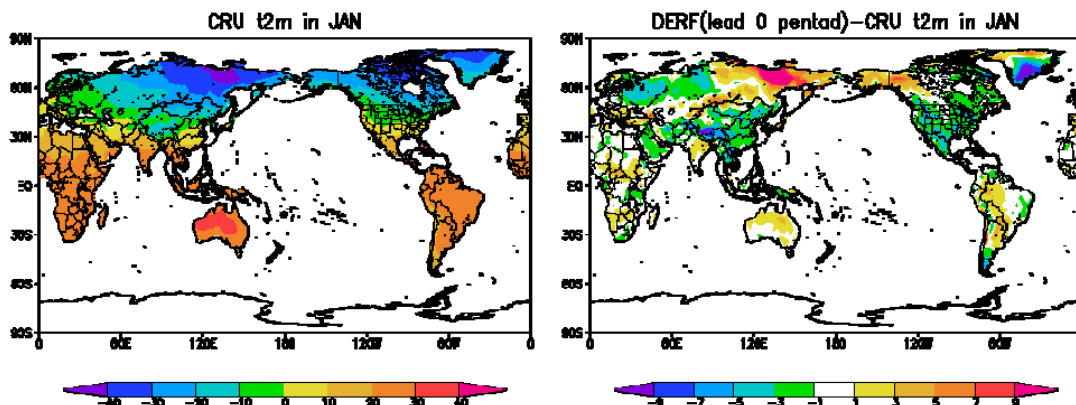
对 DERF2.0 回报资料的处理方式和评估区域与第 4 季度报告中所述一致，检验要素包括降水 (prec)、2m 气温 (t2m)，500hPa 位势高度场 (z)，200hPa 和 700hPa 纬向风场 (uwnd)。检验方法包括确定性预报检验，如气候态对比、空间距平相似系数 (ACC)、距平符号一致率 (Pc)、趋势异常综合检验 (Ps 评分)、分级评分检验 (Pg)、时间距平相关系数 (TCC) 和均方根技巧 (MSSS) 等，也包括概率预报检验，如 ROC 和 RD 等。

3 月尺度评估结果

3.1 2m 气温 (t2m)

3.1.1 气候态对比

首先将模式的气候态与观测进行对比，从而对模式模拟性能进行简单评估。经过对比，模式分别超前 0 候、1 候、2 候和 3 候的气候态差别不大，这里仅给出超前 0 候时模式与观测气候态的偏差，也能最好的反映模式的性能。如图 1 所示，DERF2.0 的 t2m 第 1 季度气候态在欧亚大陆西部高纬度地区以及青藏高原均整体偏低 1℃ 以上，对欧亚大陆中东部高纬地区模拟的温度偏高，我国大部温度偏低。



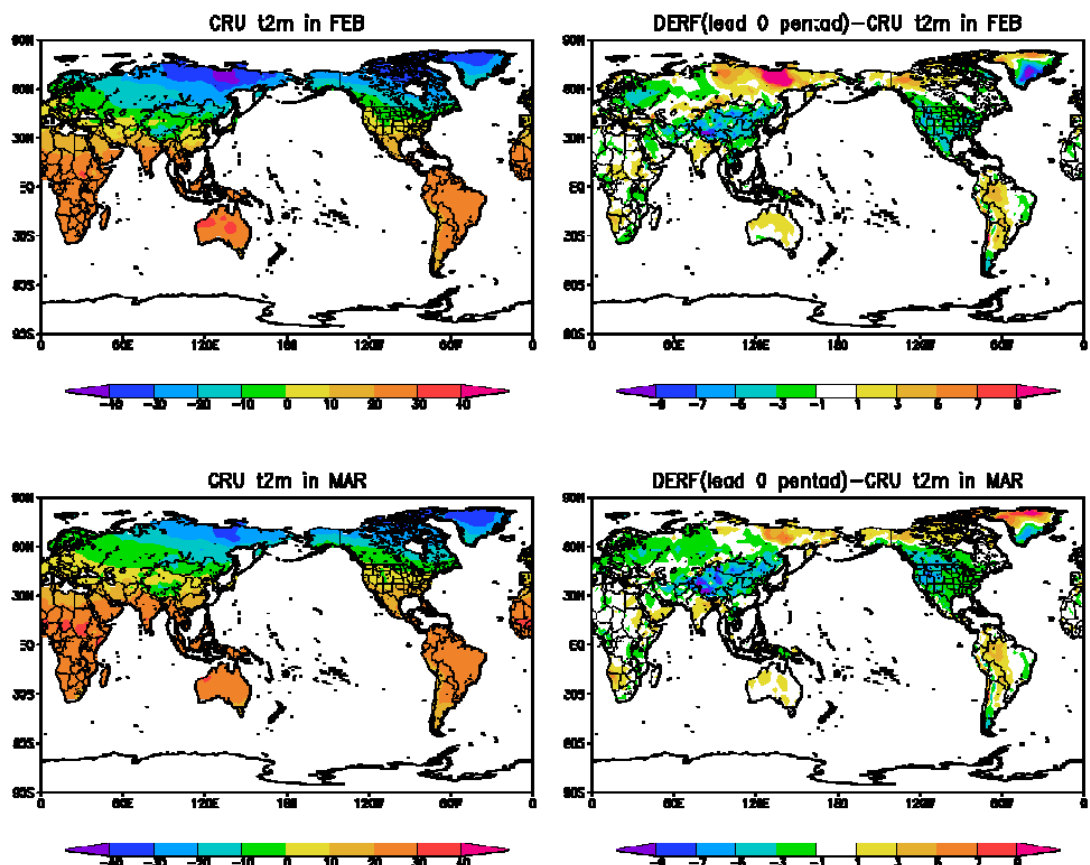
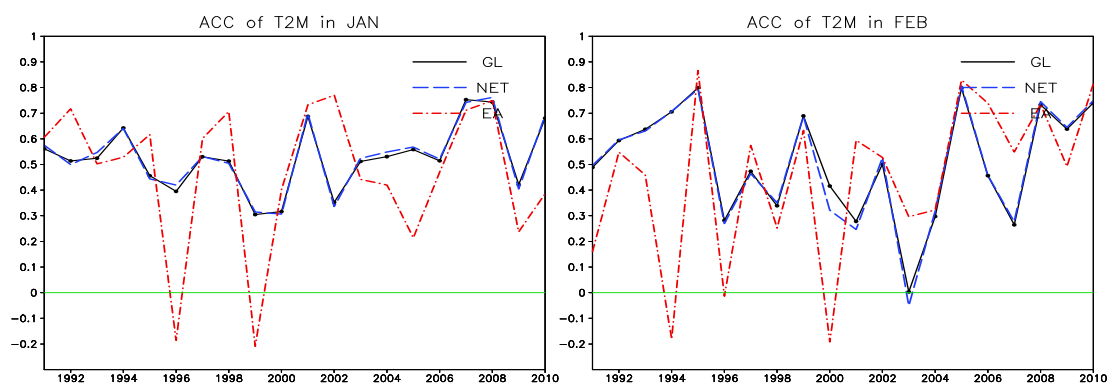


图1 全球多年(20a)平均的陆地1月、2月、3月的2m温度(t2m),左列为CRU观测,右侧为DERF(超前0候预报)与观测的差值图

3.1.2 空间距平相似系数(ACC)

空间距平相关系数(ACC)是短期气候预测中最常用的检验方法之一,反映模式的总体预测性能。分别计算不同 lead time 的空间相似系数,这里给出全球(GL)、北半球热带外(NET)和东亚(EA)区域 ACC 的年际变化(图2)和多年平均值(图3)。并对每个月的结果进行多年平均,讨论模式对 t2m 距平预报空间模态的把握能力和对初值的敏感性。如图2所示,与第4季度相比,第1季度的 ACC 评分有所提高,但东亚区域平均明显低于全球和北半球热带外平均值。同时,ACC 对模式的初值比较敏感,在超前2候预报时,各个月的 ACC 评分均有明显下降。



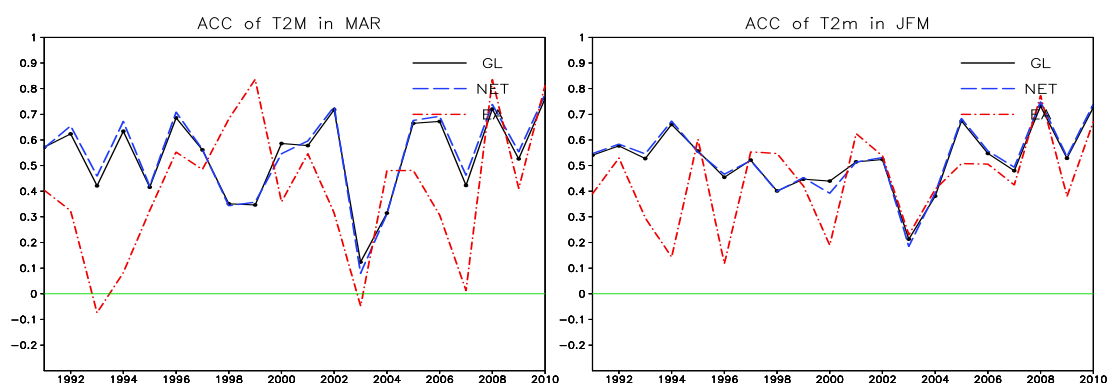


图 2 DERF2.0 提前 0 候起报的不同地区平均的（全球（GL）、北半球热带外（NET）、东亚（EA））陆地 1 月、2 月、3 月的 2m 温度（t2m）与观测的空间距平相似系数（ACC）逐年演变

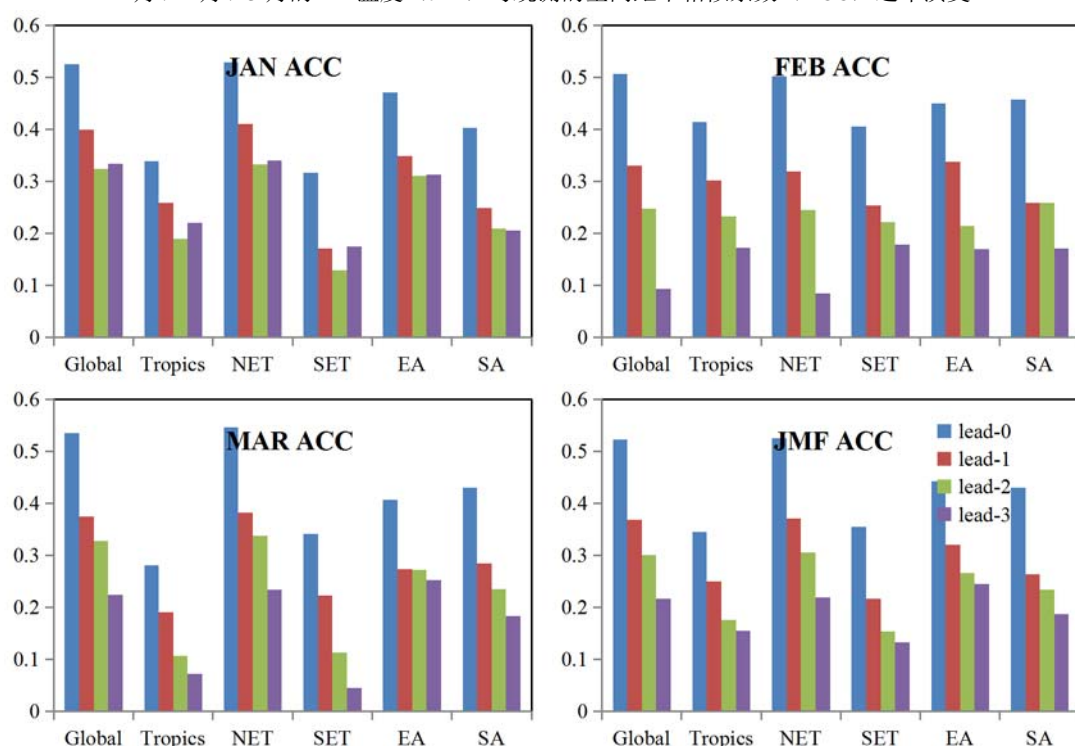


图 3 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 2m 温度（t2m）的空间距平相似系数在不同区域的多年平均，横坐标代表多个区域（见表 3）

3.1.3 距平符号一致率（Pc）

距平符号一致率 Pc 反映的是一定范围内预测值与实况值距平符号一致的站（格）点数占总站（格）点数的比例。只有当同号率大于 50%，异常的主要趋势被反映出来时，再考察强度预测才有意义，因此 ACC 需要与其配合使用。由图 4 可以看出，不同地区的同号率差异没有 ACC 那么明显，约在 55%~70%之间，北半球热带外在超前 0 候预报时则能够达到 70%，因此 DERF2.0 的预测总体上能够反映出 t2m 异常的主要趋势，与第 4 季度相比略有提升。同时，同号率对初值并不如 ACC 敏感，也就是较长 lead time 的预报仍能反映出异常的主要趋势，特别是在热带地区，这一定对气候预测中的定性预报十分重要。

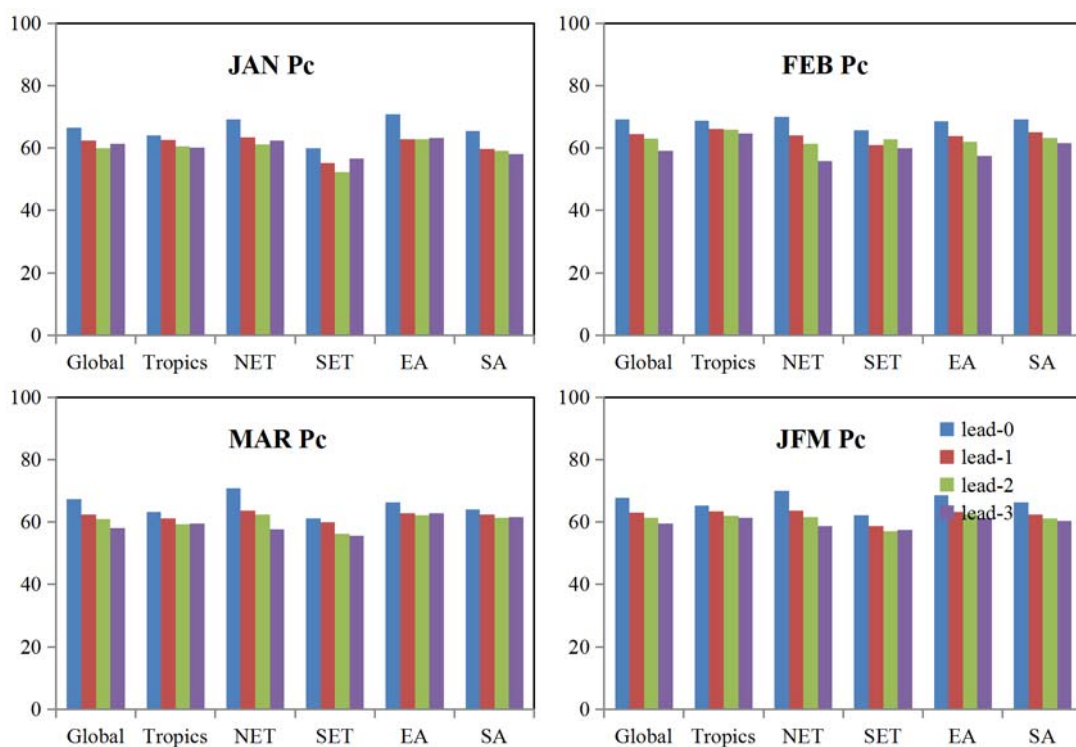
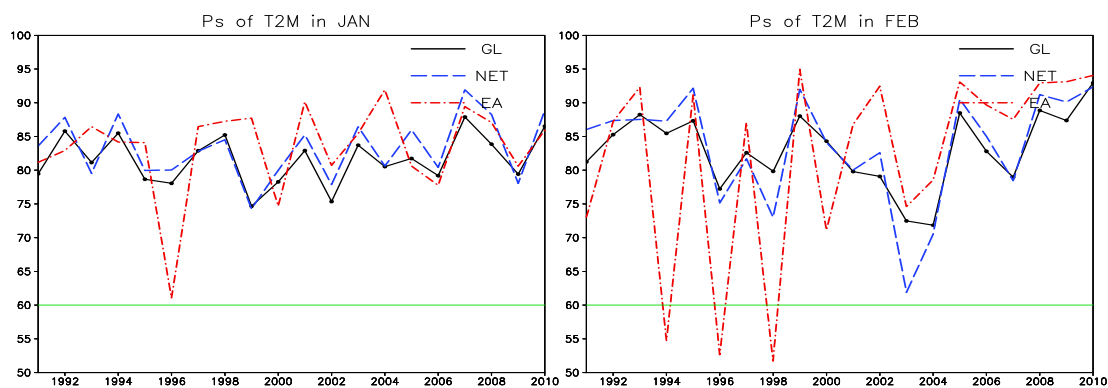


图 4 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 2m 温度 (t2m) 的距平符号一致率 (Pc) 的在不同地的多年平均

3.1.4 趋势异常综合检验 (Ps)

Ps 评分也是对分类确定性预报的检验方法, 综合考虑了预报趋势项、异常项和漏报项的影响, 这里同样给出不同区域平均的 Ps 评分的年际变化和多年平均。总体而言, Ps 评分与 ACC 类似, 年际变化较大, 全球总体评分在 75-90 分之间, 在东亚地区年际变化非常明显, 但也基本都能高于 65 分, Ps 评分对初值的敏感程度比 ACC 要小很多。



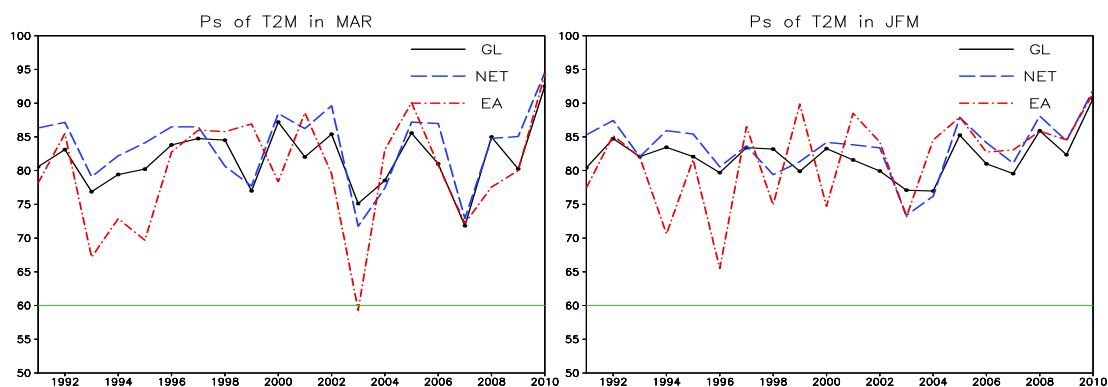


图 5 同图 2，但为趋势异常综合检验（Ps 评分）逐年演变

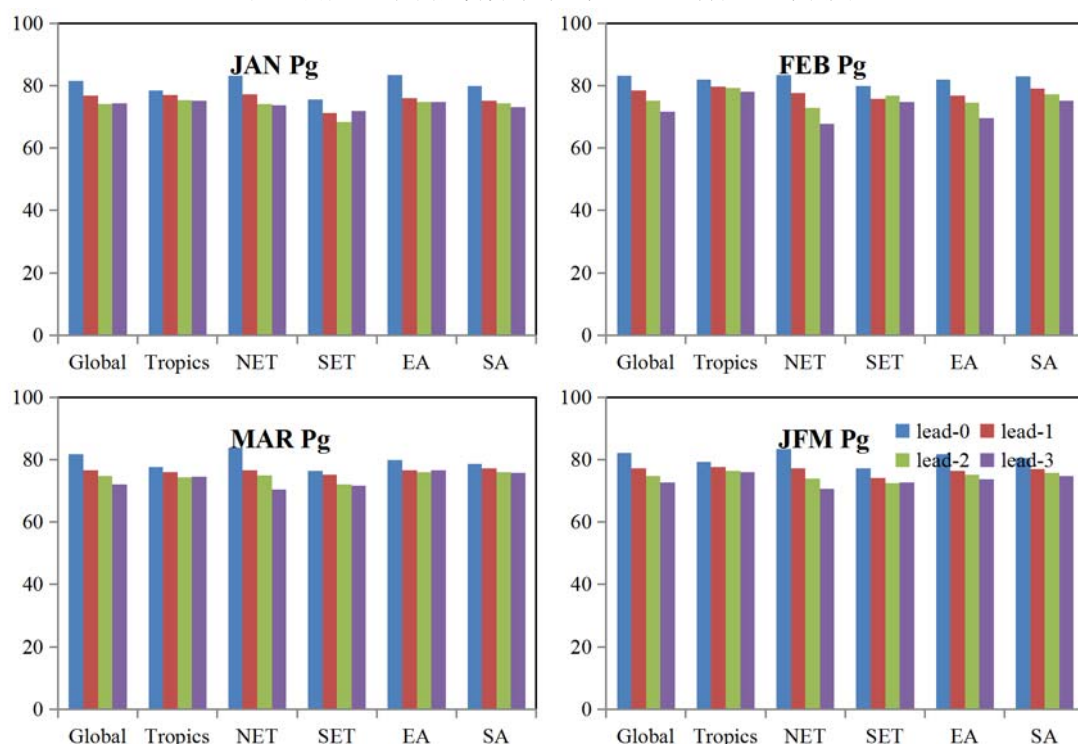


图 6 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 2m 温度 (t2m) 的趋势异常综合检验 (Ps 评分) 的在不同地的多年平均

3.1.5 分级评分检验 (Pg)

Pg 评分主要反映的是模式对预测要素在量级上的把握程度。综合图 7 图 8 而言，第 1 季度各月气温预测的 Pg 评分在我国大部分地区均在 70-75 分。就全球而言，热带地区评分相对较高，均超过 70 分。同时，与同号率类似，Pg 评分对初值不是十分敏感，特别是热带地区。

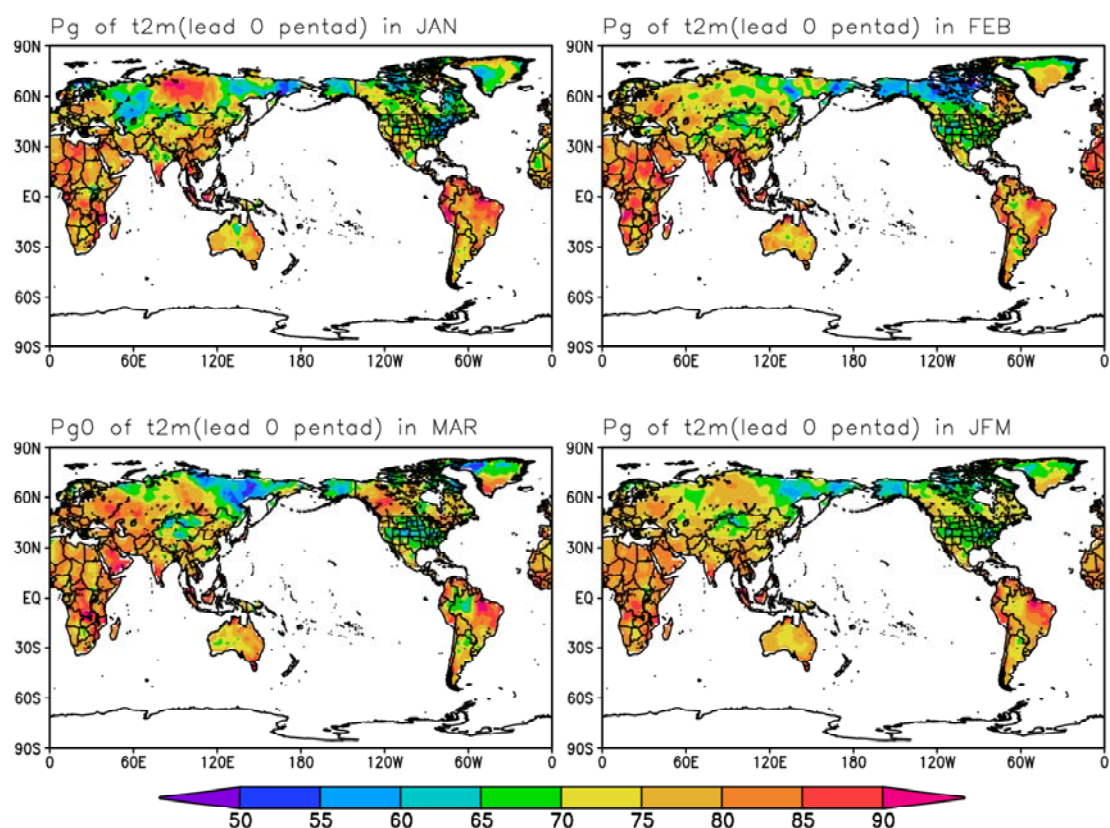


图 7 DERF2.0 超前 0 候起报的的全球 1 月、2 月、3 月和第 1 季度的 2m 温度 (t2m) 的 Pg 评分多年平均分布

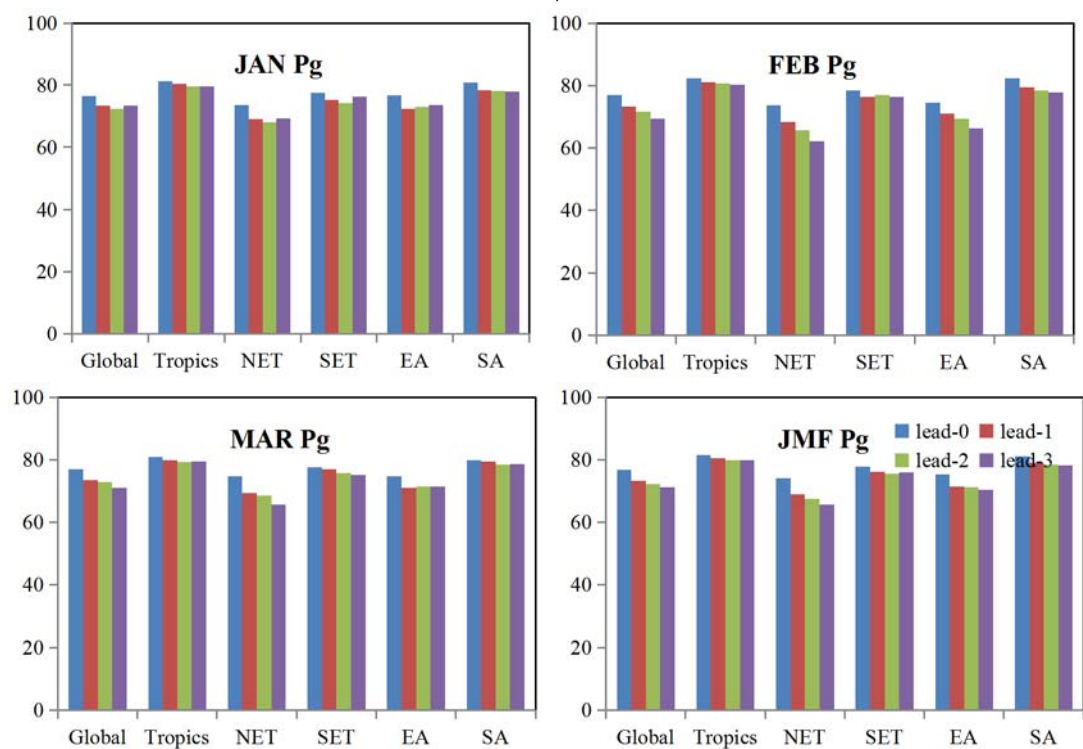


图 8 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 2m 温度 (t2m) 的 Pg 评分在不同地区的多年平均

3.1.6 时间距平相关系数 (TCC)

与 ACC 不同，时间相关系数 (TCC) 能够给出每个格点的预报技巧，这里分别给出逐月和第 1 季度 TCC 的空间分布 (图 9) 和区域平均 (图 10)。由图可见，对全球而言，与第 4 季度相比第 1 季度的预报技巧有显著提升，全球平均 TCC 在 0.4 左右；区域平均来看，北半球热带外和东亚地区的预报技巧较高，如欧亚中高纬度，我国西南和东北地区等。此外，TCC 对初值的依赖性也非常明显，随着 lead time 的变长技巧显著下降，特别是从 lead 0 候到 lead 1 候下降非常明显。

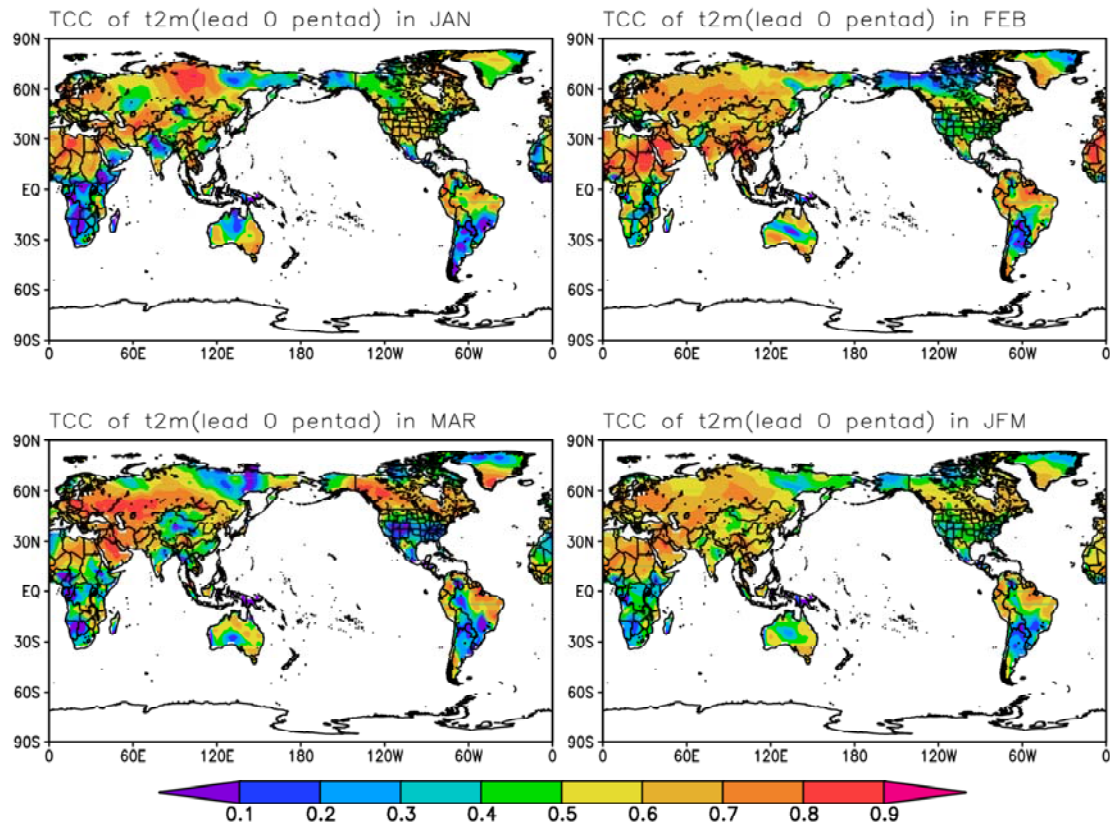
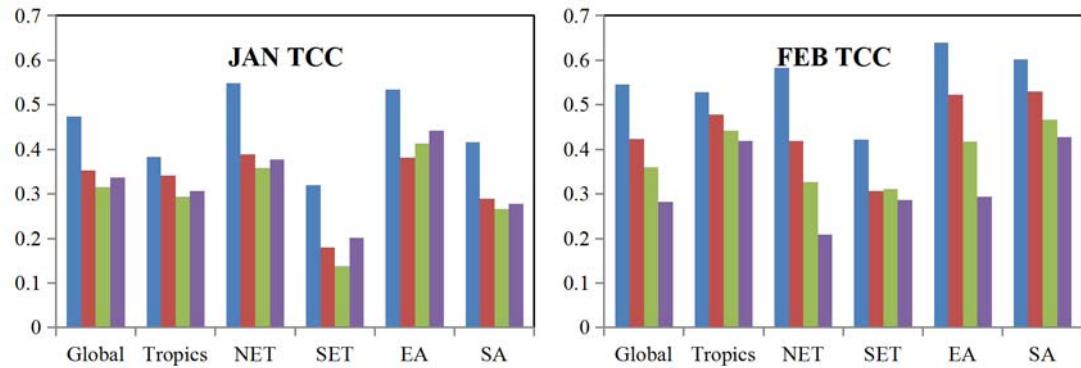


图 9 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的 2m 温度(t2m)的时间距平相关系数(TCC)分布



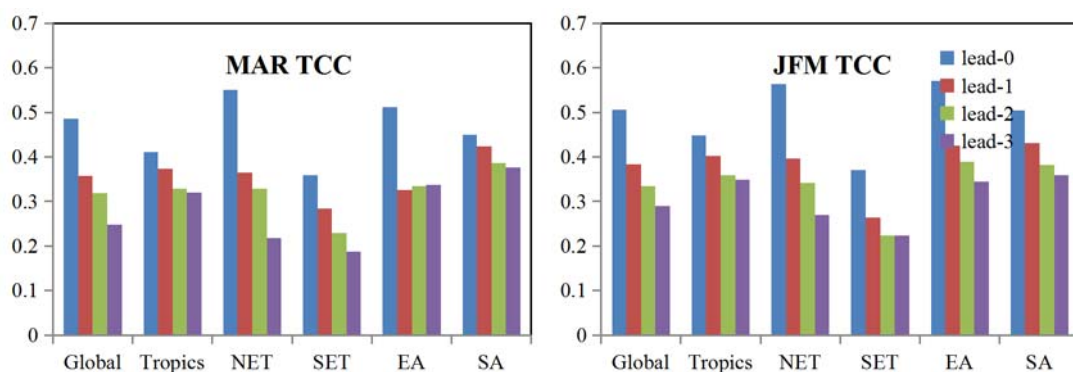


图 10 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 2m 温度 (t2m) 的时间距相关系数 (TCC) 在不同地区的平均

3.1.7 平均方差技巧 (MSSS)

MSSS 为 WMO 推荐的确定性预报的主要评估方法, 主要考虑预报误差和“气候”预报误差的比值, MSSS 值越接近 1, 表明模式预测越准确, 大于 0 表明有预报意义。此处需要注意, 由于预报中最关注对异常的预报而不是气候态的偏差, 因此计算预报误差时未考虑气候态的漂移, 仅考虑对距平预报的偏差。

全球平均而言, MSSS 总体得分在超前 3 候预报时开始降低至 0 以下, 在空间上看, MSSS 与 TCC 的评估结果较为类似, 欧亚中高纬度, 我国东北地区评分大于 0, 但在热带、东亚和南亚的评分很低。

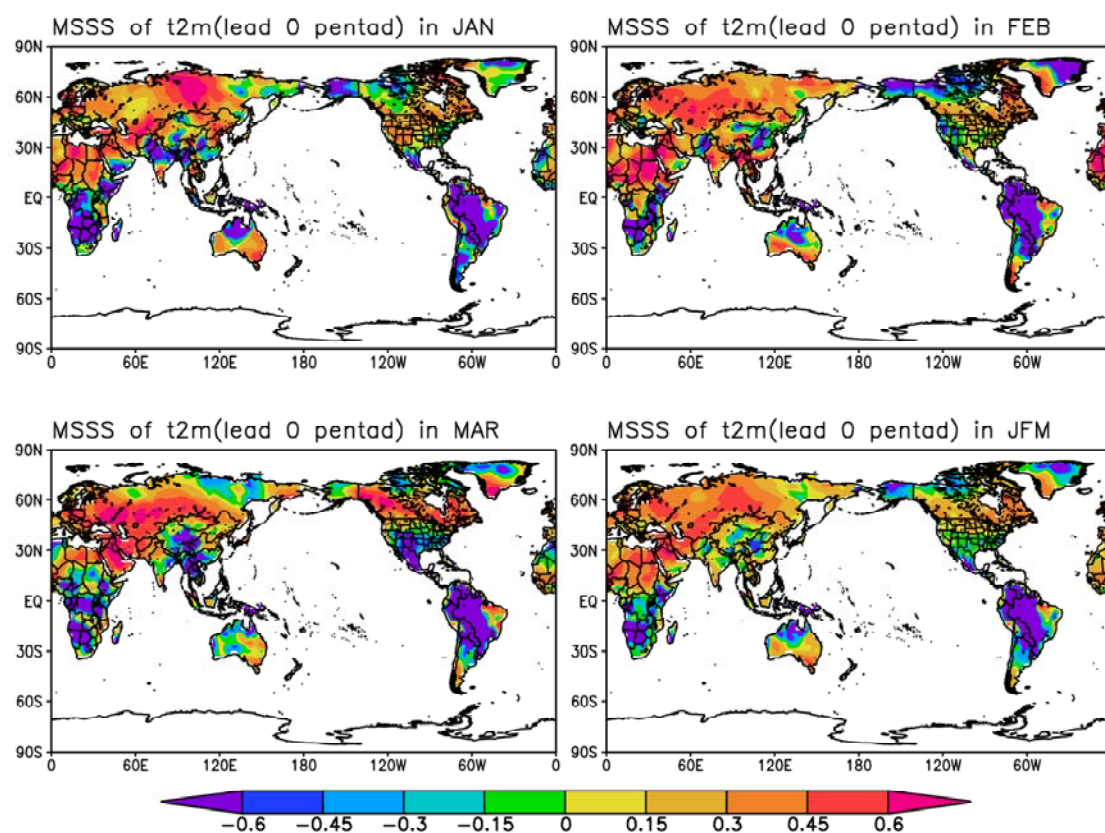


图 11 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的 2m 温度 (t2m) 的平均方差技巧 (MSSS) 分布

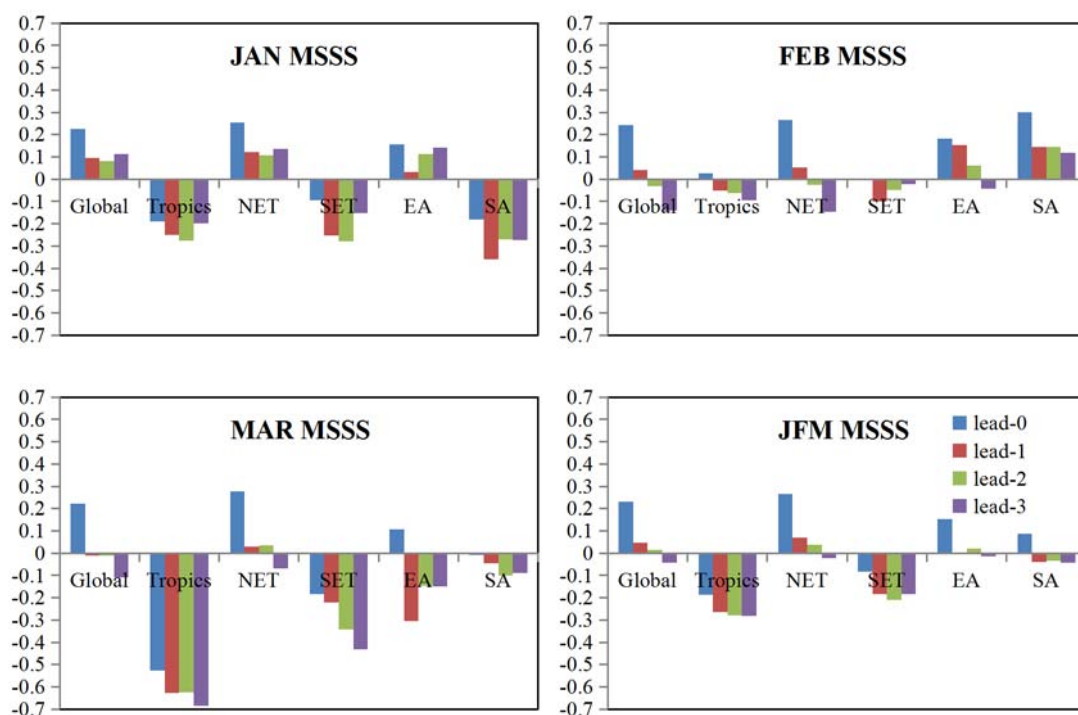


图 12 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 2m 温度(t2m)的平均方差技巧(MSSS)在不同地区的平均

3.1.8 气温 (t2m) 评估小结

综合以上几种方法的评估，可对 DERF2.0 对陆地 t2m 的预报性能检验结果进行初步小结。首先，模式的平均态存在一定偏差，表现为欧亚大陆西部高纬度地区以及青藏高原地区偏冷，欧亚大陆中高纬东北部偏暖；第二，DERF2.0 对气温具有一定的预报技巧，在北半球热带外预报技巧相对较高，但对东亚地区的预报技巧地区低于北半球热带外平均，对我国而言，模式对我国西南和东北地区预报技巧相对较高；第三，第 1 季度 OND 三个月的技巧与第 4 季度相比技巧略有提升，但在非等级预报评分中 (ACC、TCC、MSSS)，模式对初值均非常敏感，随着 lead time 的延长，预报技巧显著下降，但在等级预报 (同号率、Pg 评分) 中，对初值并不敏感，Ps 评分接近 80 分，这提示预报员，在参考 lead time 较长的预报时，预报的量值可能会与观测差距较大，但模式的定性预报仍有较高的参考价值。

3.2 降水 (prec)

3.2.1 气候态对比

与对气温 (t2m) 的分析类似，这里同样首先进行 DERF2.0 全球降水预测结果与 GPCP 观测的气候态对比分析。如图 13 所示，DERF2.0 预测的降水气候态的模式与观测相比，雨带的分布基本一致，特别是在热带反映出了热带辐合带 (ITCZ) 和南太平洋辐合带 (SPCZ) 的主要特征，但热带西太平洋暖池和 SPCZ 地区降水明显偏强，MC 降水偏弱。

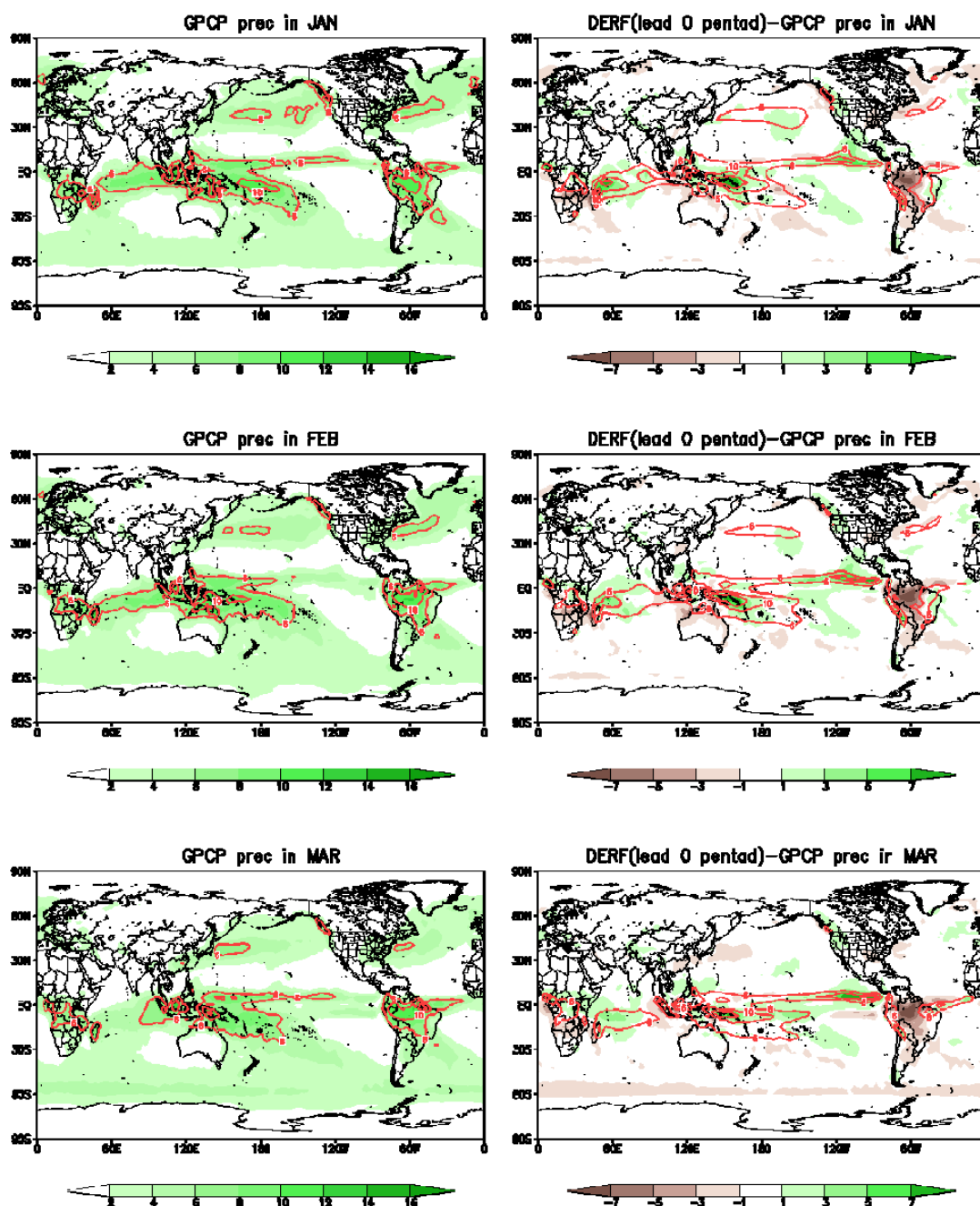


图 13 全球多年 (20a) 平均的 1 月、2 月、3 月降水量 (单位: mm/day), 左列为 GPCP 观测, 右侧阴影为 DERF (超前 0 候) 预报的气候态与 GPCP 观测之差, 红色等值线标出降水量大于 6mm/day 的区域

3.2.2 空间距平相似系数 (ACC)

图 14 和图 15 给出降水距平空间相似系数的年际变化和不同 lead time 起报的多年平均, 与 2m 气温相比, 降水的 ACC 评分明显偏低, 全球平均而言在 0.3 左右, 与第 4 季度相当。同时各个区域评分差距比较明显, 热带地区相似系数较高, 且在起报时间超前较长时仍下降不明显, 反之在东亚地区, ACC 得分较低, 只有 0.1 左右, 热带外 ACC 评分均对初值比较敏感。注意在 1997 年的强 El Nino 年, 全球降水 ACC 得分最高。

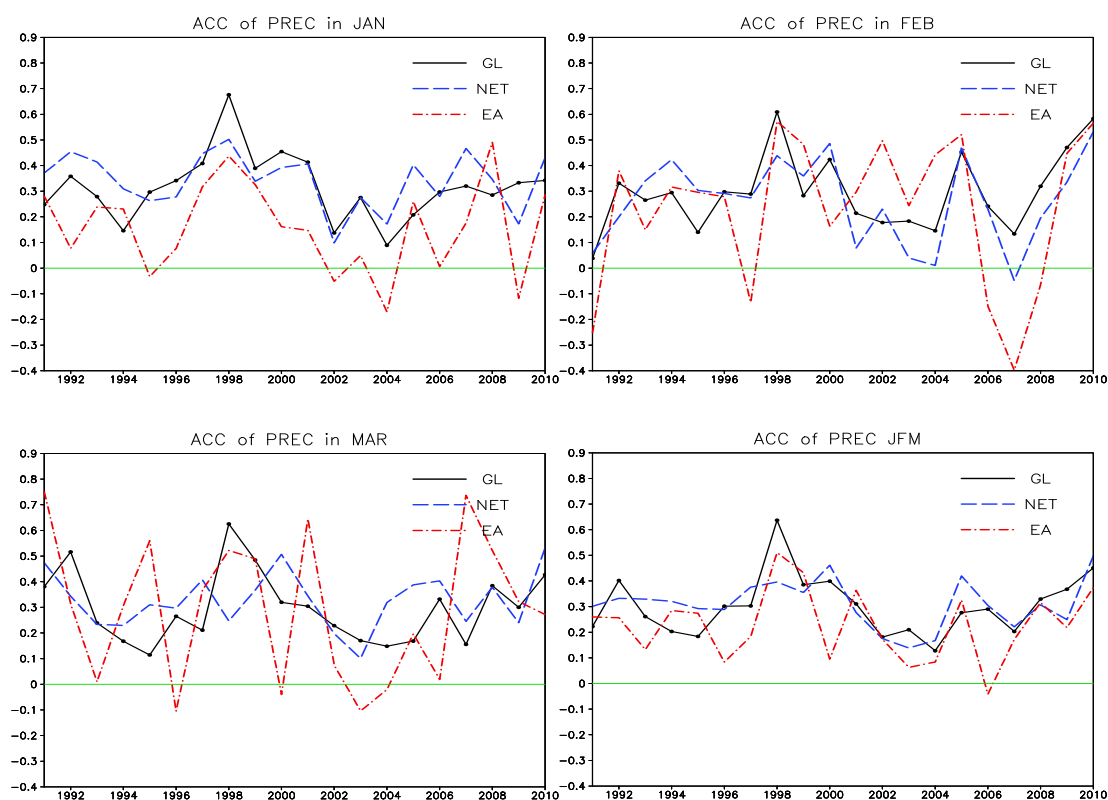


图 14 DERF2.0 提前 0 候起报的不同地区平均的（全球（GL）、北半球热带外（NET）、东亚（EA））陆地 1 月、2 月、3 月的降水（prec）与观测的空间距平相似系数（ACC）逐年演变

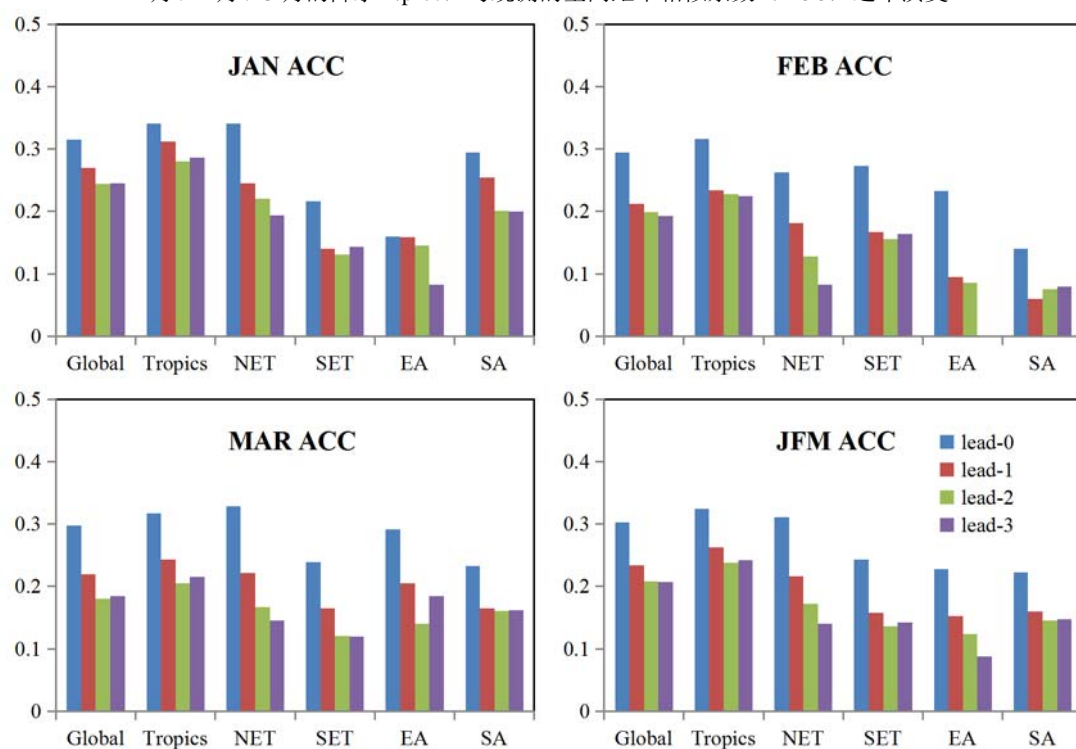


图 15 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的降水（prec）的距平相似系数在不同区域的多年平均，横坐标代表多个区域（见表 3）

3.2.3 距平一致率 (Pc)

从距平符号一致率上看（图 16），DERF2.0 对降水预报同号率约为 55%-60%，略低于气温预测，但仍能大体反映出旱涝趋势。与 ACC 不同，同号率在各个地区差别并不明显，热带地区的评分略高，对起报初值也不甚敏感，随着 lead time 的延长，同号率只是略有下降，模式起报的超前时间较长仍可为预测提供定性参考。

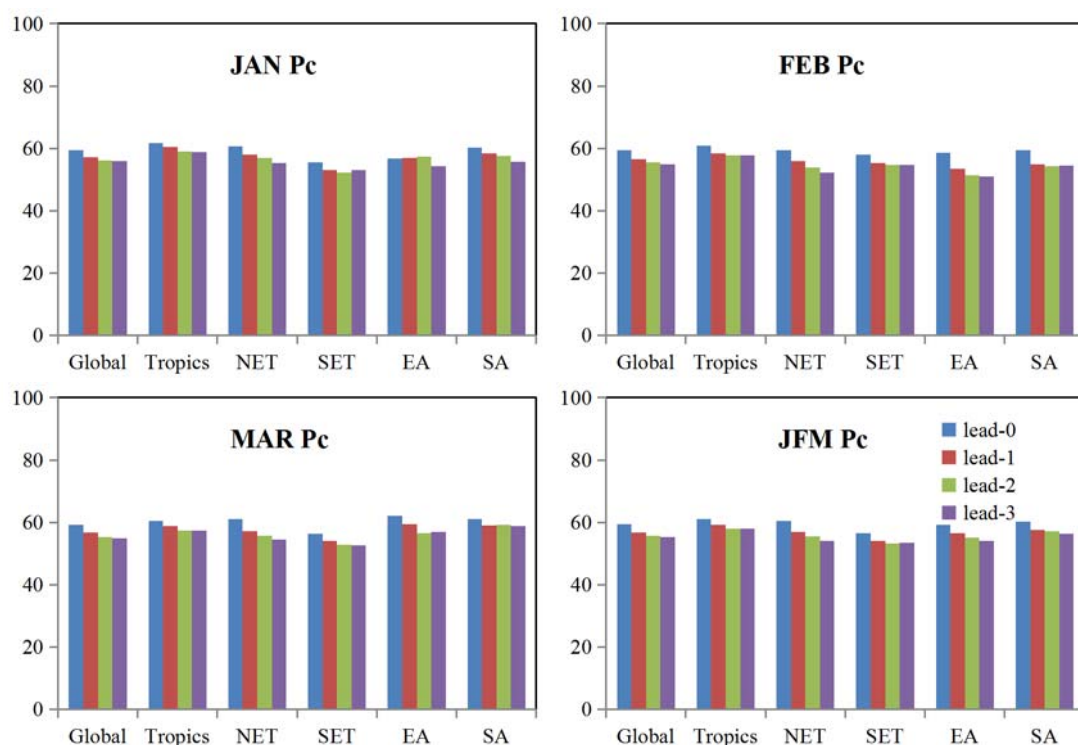
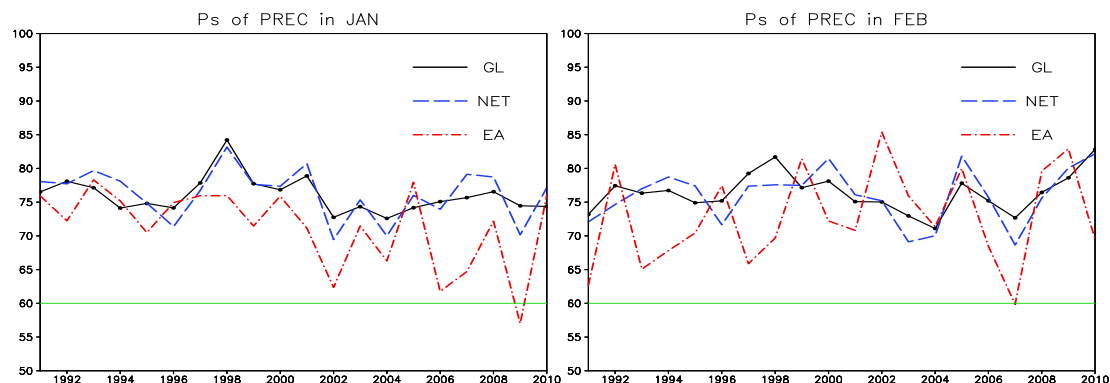


图 16 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的降水 (prec) 的距平符号一致率的在不同地区的多年平均

3.2.4 趋势异常综合检验 (Ps)

图 17 和图 18 分别给出降水 Ps 评分的年际变化和不同区域的多年平均。多年而言，降水的评分在 70-80 分之间，略低于对气温的预报，但年际变化大，特别是东亚地区，而热带和南亚地区得分较高。此外，随着 lead time 的延长，Ps 评分下降并不明显。



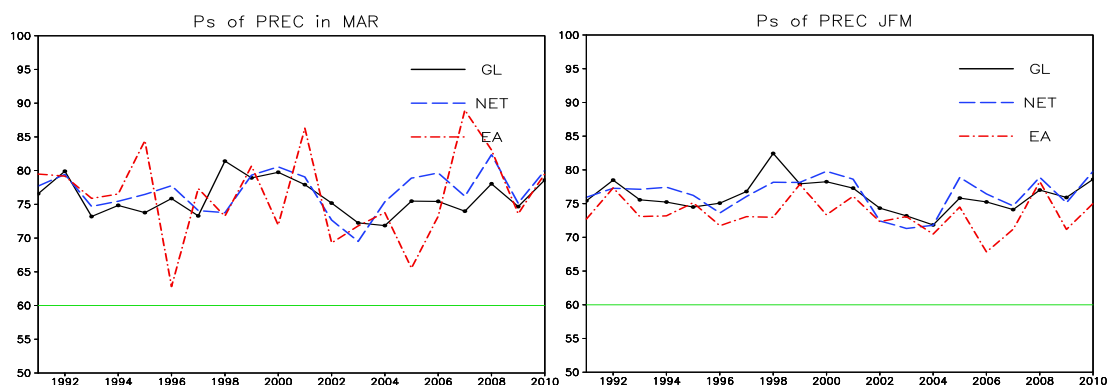


图 17 同图 14，但为趋势异常综合检验（Ps 评分）逐年演变

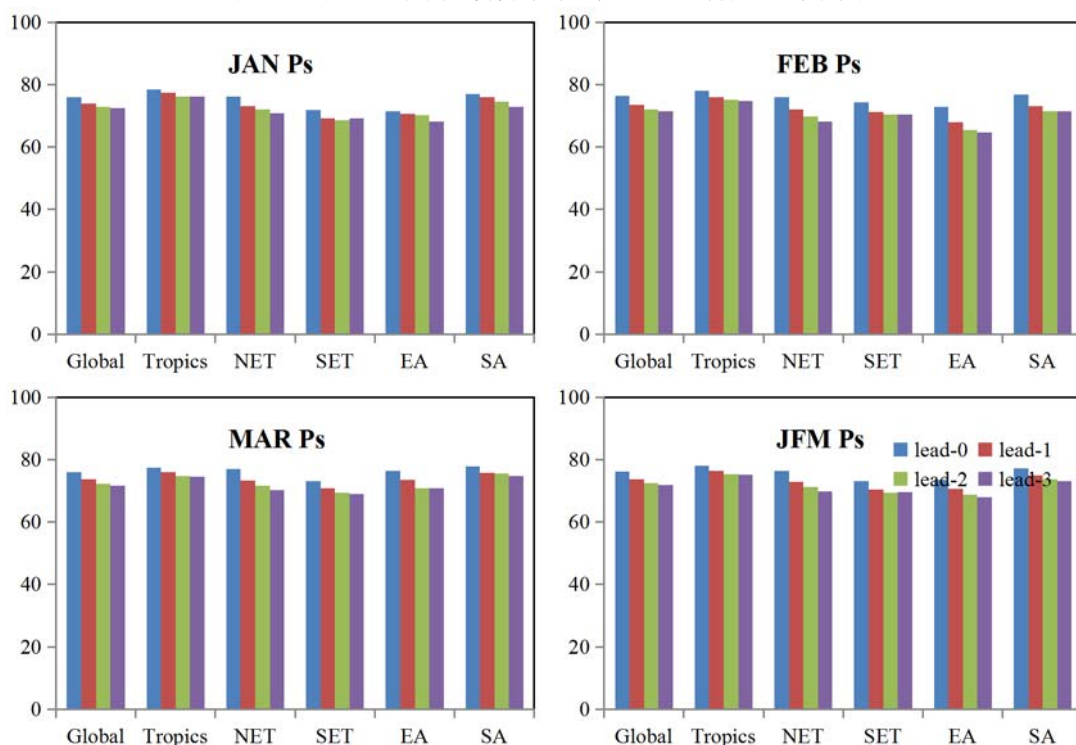


图 18 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的降水 (prec) 的趋势异常综合检验 (Ps 评分) 的在不同地区的多年平均

3.2.5 分级评分检验 (Pg)

DERF2.0 的 Pg 评分总体在 60-65 分之间，低于气温的 70-75 分，说明预测的降水距平百分率与实况符号相同时往往相差两个等级，或者同处正常级别但符号相反。对各个区域而言，模式对热带赤道附近（主要为太平洋上的 ITCZ）、南北半球中高纬西风带（60°N/S）地区的预测性能较好，在副热带地区、南太平洋秘鲁沿岸冷水区性能较差，对于我国而言，模式对我国西南地区降水异常等级把握较好。与同号率类似，Pg 对初值不敏感。

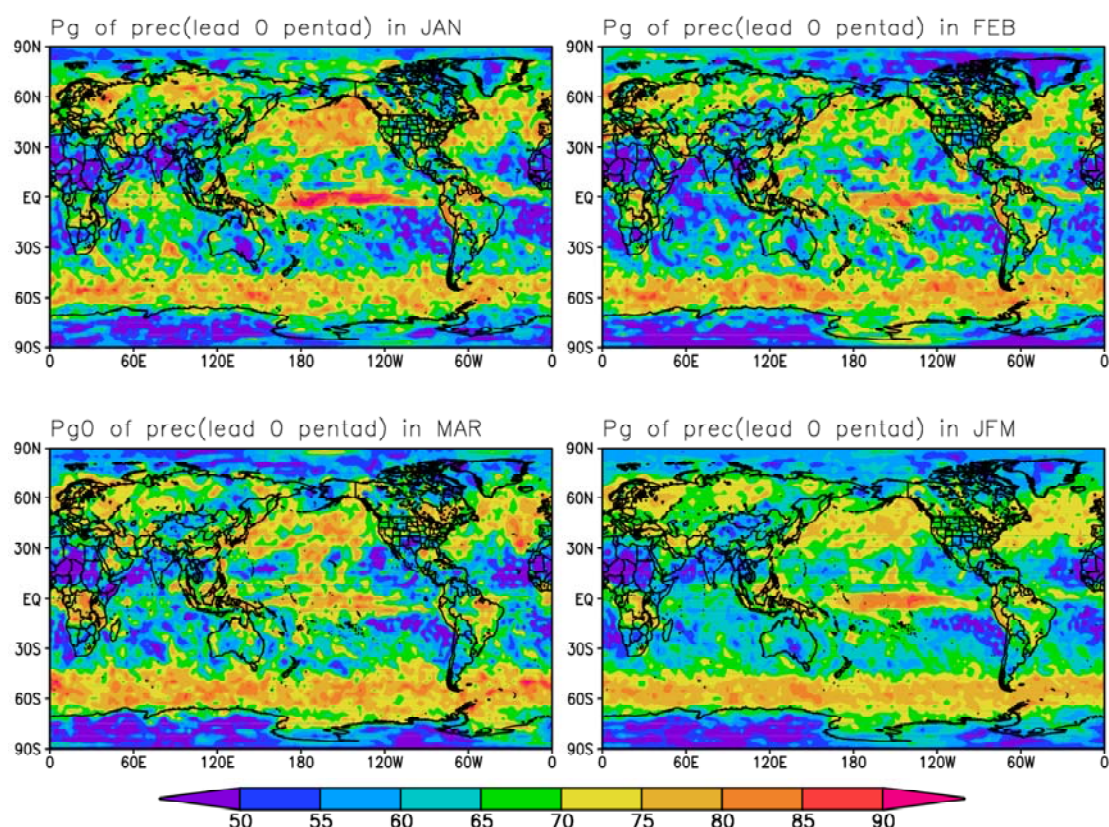


图 19 DERF2.0 超前 0 候起报的的全球 1 月、2 月、3 月和第 1 季度的降水 (prec) 的 Pg 评分多年平均分布

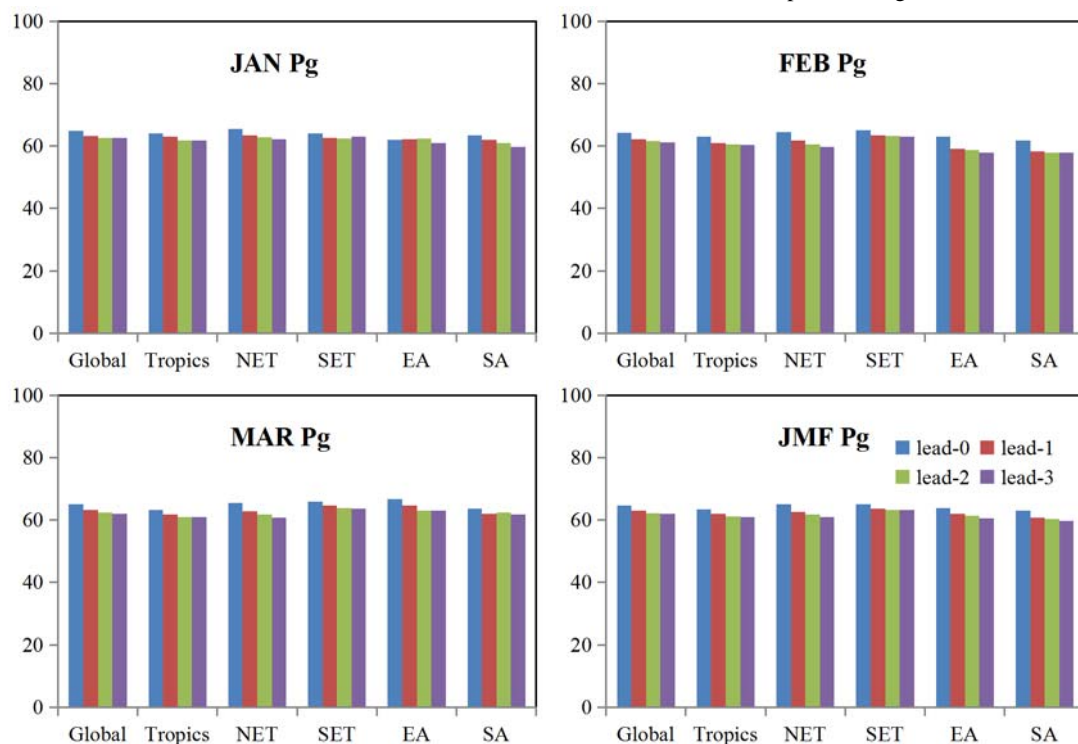


图 20 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的降水 (prec) 的 Pg 评分在不同地区的多年平均

3.2.6 时间距平相关系数 (TCC)

图 21 给出超前 0 候起报的全球各月降水的时间距平相关系数 (TCC) 分布, 总体而言,

相关技巧在 0.5 以上的区域与气温相比已显著减少，全球平均的 TCC 约为 0.1-0.25，表明模式对降水的预报技巧明显偏低。结合图 21，从不同区域来看，模式对热带地区，特别是热带太平洋暖池和 ENSO 发生地区降水的预报技巧较高，说明降水的异常信号主要来自前期海温强迫，初值信号延续时间较长，而其他地区则对初值十分敏感，对于我国而言，模式在第 1 季度对我国江淮地区的降水异常有一定预测能力。

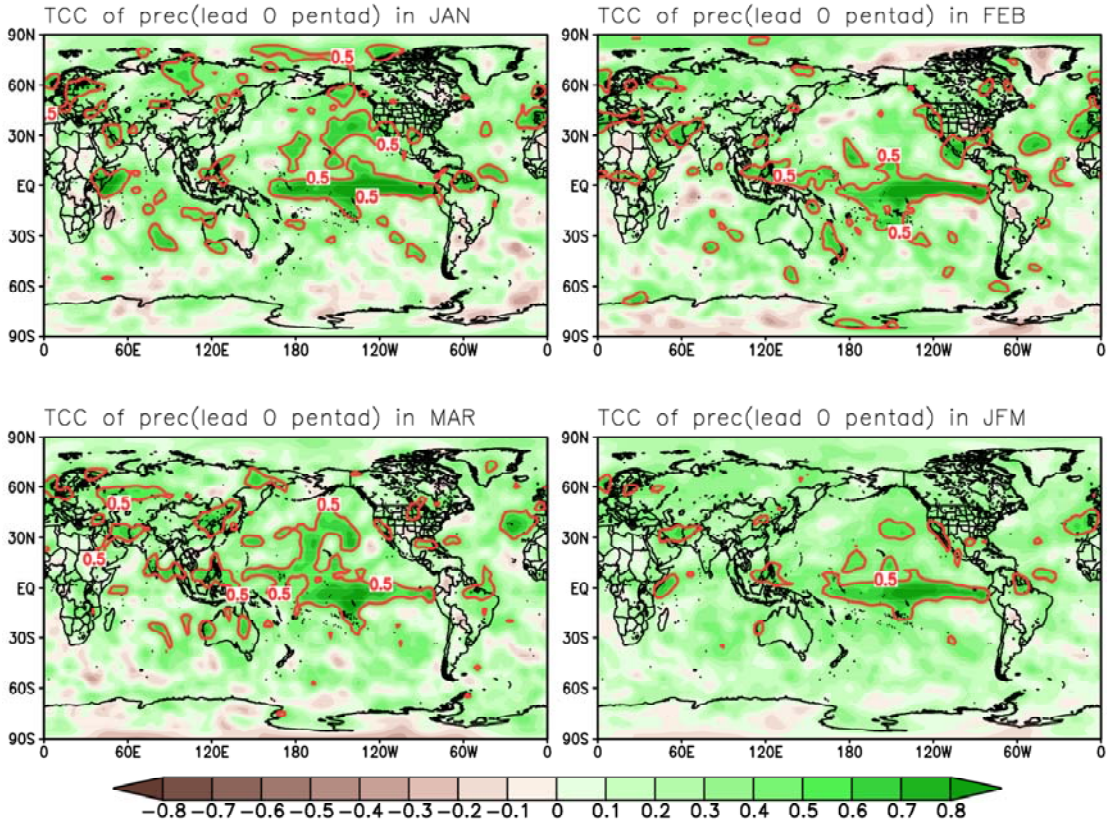
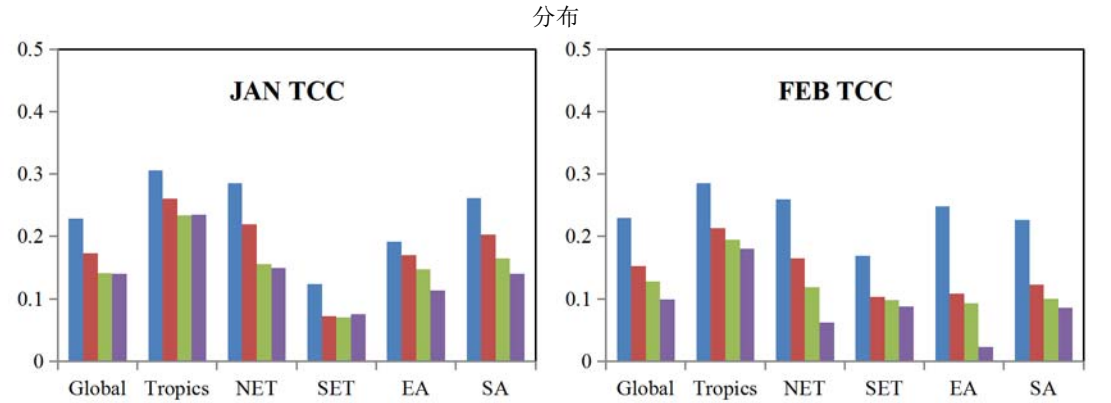


图 21 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的降水 (prec) 的时间距平相关系数 (TCC)



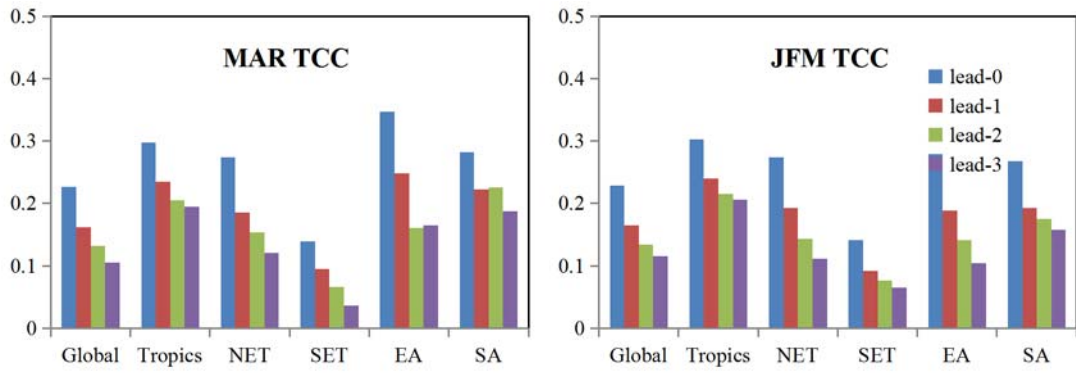


图 22 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的降水 (prec) 的时间距平相关系数 (TCC) 在不同地区的平均

3.2.7 平均方差技巧 (MSSS)

对降水均方差技巧 (MSSS) 评估表明, 模式对赤道中东太平洋 ENSO 发生的区域、北半球中高纬地区、南亚和我国江淮地区降水有一定的预测能力, 对热带东印度洋和热带西太平洋地区 and 我国西北大部的降水性能预报较差。

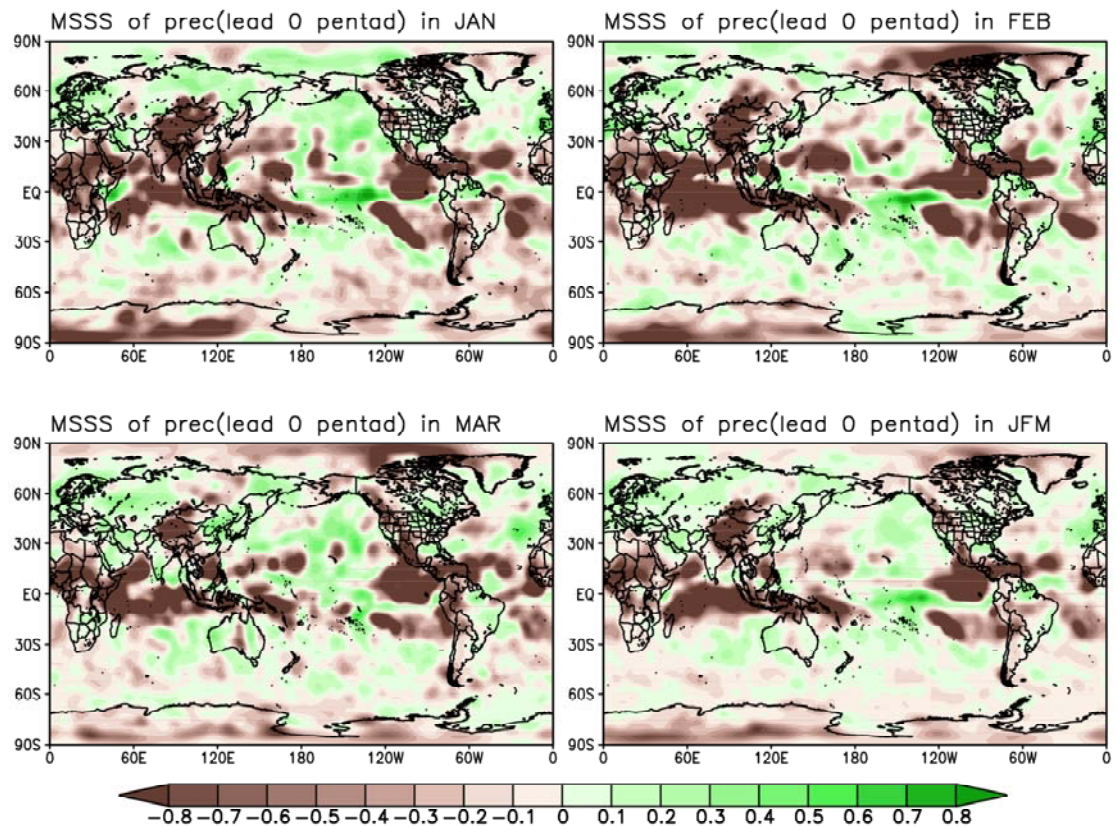


图 23 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的降水 (prec) 的平均方差技巧 (MSSS) 分布

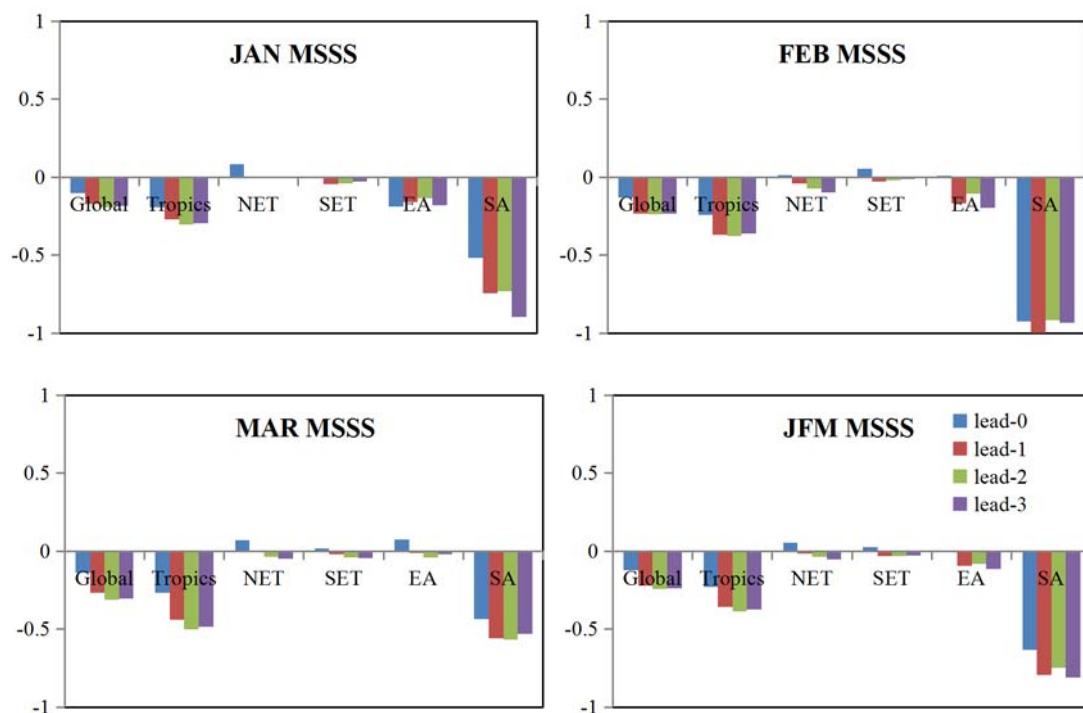


图 24 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的降水 (prec) 的平均方差技巧 (MSSS) 在不同地区的平均

3.2.8 降水 (prec) 评估小结

本部分从多个方面对 DERF2.0 对月尺度的降水预测性能进行检验和评估，主要得到以下几点结论：首先，在气候态，模式能够大致较好的把握各个月雨带分布，特别是热带太平洋的 ITCZ 和 SPCZ 的降水，但是，热带西太平洋暖池和 SPCZ 地区降水明显偏强，MC 地区降水偏弱；第二，整体而言，模式对降水有一定预测能力，lead 0 候时全球平均 TCC 在 0.25 附近，在非等级评估方法中，模式性能对初值十分敏感，整体预测水平明显低于气温预测，模式对第 1 季度降水的预测能力与第 4 季度相当；第三，使用 ACC, TCC 等评估方法表明，模式对南亚、热带太平洋、特别是 ENSO 地区的降水预测能力较强，对我国而言，对江淮地区降水预测能力相对较好。

3.3 500hPa 位势高度 (HGT 500)

由于模式性能的原因，预报员进行短期气候预测时，除了参考模式直接输出的气温和降水等要素外，更重要的是参考模式对整个环流形势的把握，对关键影响系统，如副高、阻塞、AO 等的预报，从而对未来降水和气温异常进行预测。因此，这部分对 DERF2.0 输出的 500hPa 位势高度场进行检验评估，定量给出模式对环流形势的预报能力。

3.3.1 气候态对比

与前面几部分相类似，这里仍然首先对模式预报的 1 月、2 月和 3 月环流的气候态进行对比分析。由图 25 所示，模式预报的气候态，在北半球欧亚大陆中高纬度整体偏低，特别在北大西洋上，而在 12 月份，模式的东亚大槽和北太平洋东部脊均偏西。

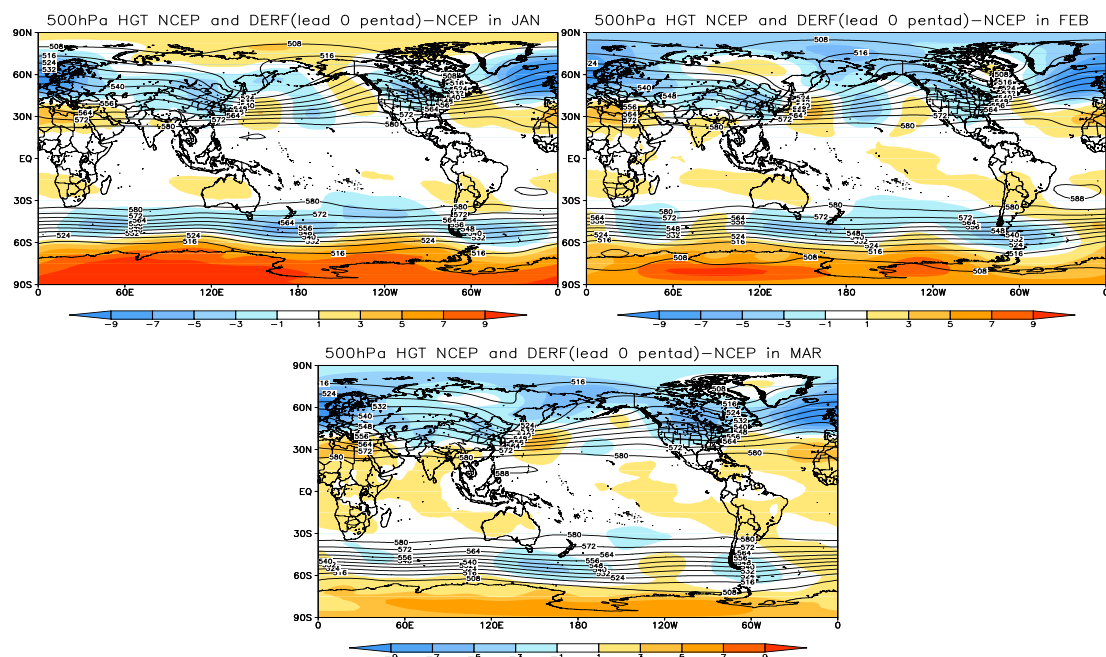


图 25 全球多年（20a）平均的 1 月、2 月、3 月 500hPa 位势高度（单位：位势什米），等值线代表 NCEP2 观测，阴影为 DERF2.0 超前 0 候预报与观测之差

3.3.2 时间距平相关系数（TCC）

图 26 给出了 DERF2.0 预测 500hPa 位势高度场的时间相关系数的空间分布。由图可知，在大部分地区模式预测的相关技巧都超过 0.5，远高于模式气温和降水的预测，其中模式对热带地区位势高度异常的预报能力要远高于其他地区，区域平均可达 0.7~0.8（图 26），且具有很好的持续性，北半球热带外的相关也高于南半球热带外。在我们更加关注东亚地区，模式预报的位势高度异常在东亚大陆和西北太平洋副热带高亚地区也具有较高的相关，模式对副高也有较好的把握能力，但对我国东北部地区的位势高度场预报技巧较低。需要注意的是，在热带外，模式对初值很敏感，随着 lead time 的延长模式预报性能显著下降。

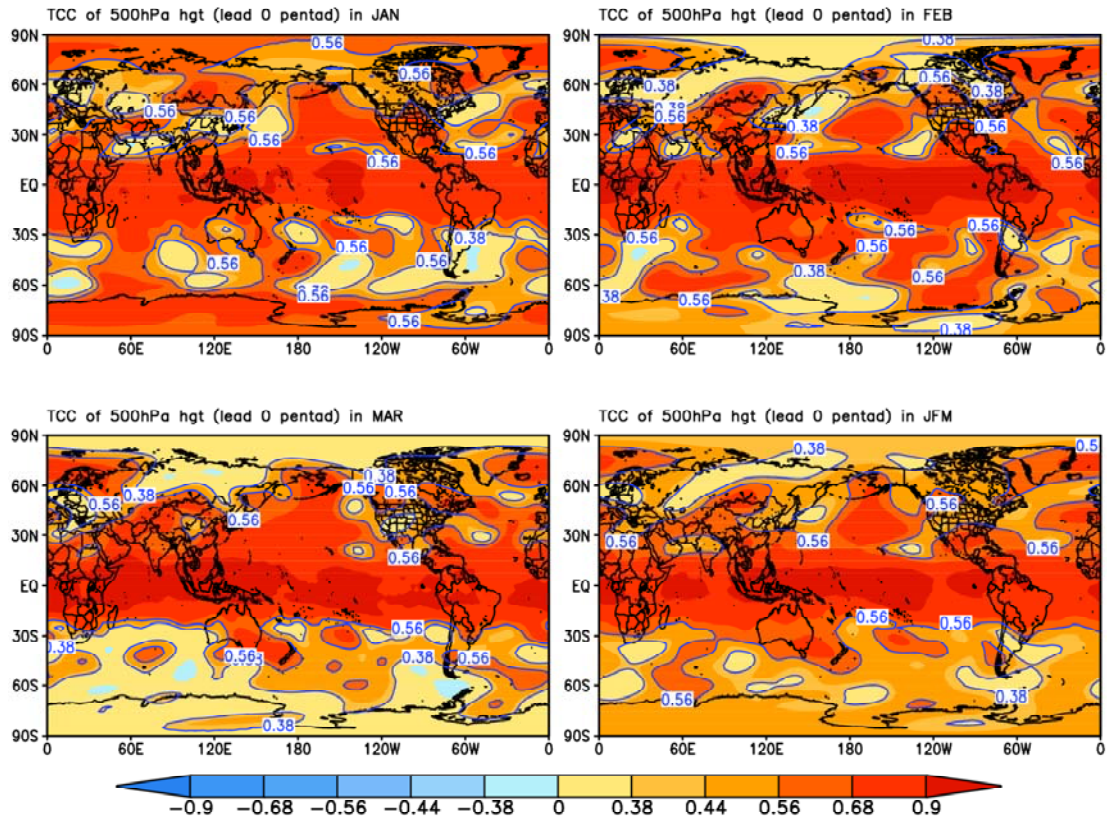


图 26 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的 500hPa 位势高度 (HGT) 的时间距平相关系数 (TCC) 分布, 等值线标定分别代表通过 0.1 和 0.01 显著性检验

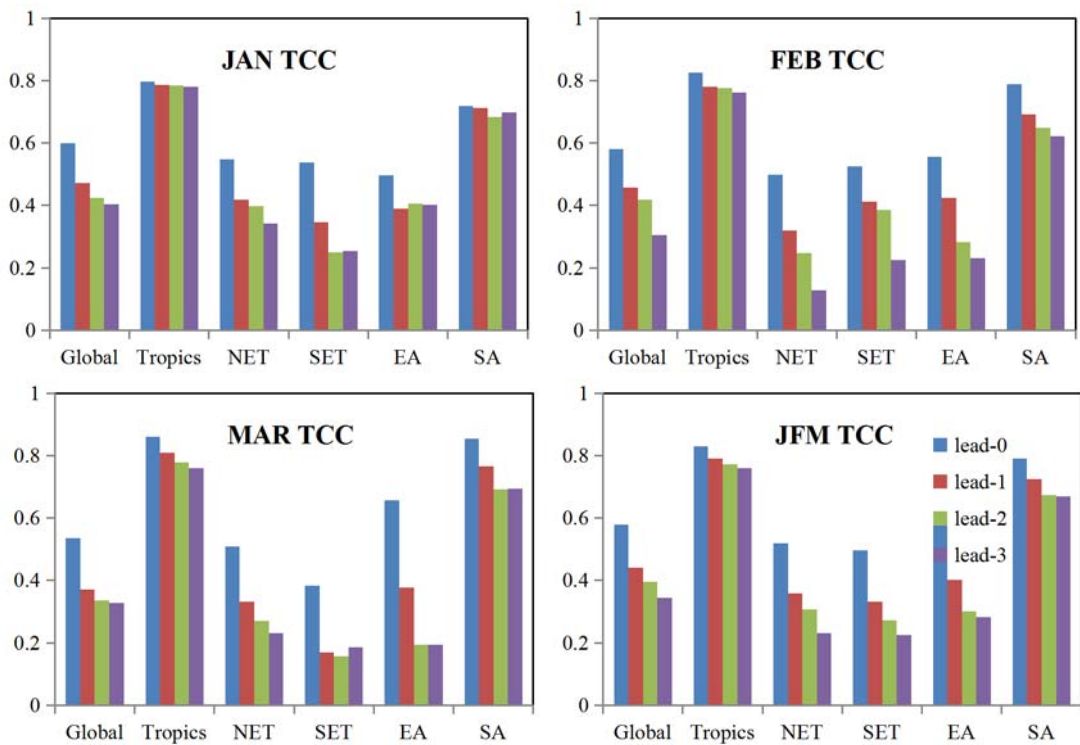


图 27 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 500hPa 位势高度 (HGT) 的时间距平相关系数 (TCC) 在不同地区的平均

3.3.3 平均方差技巧（MSSS）

对 DERF2.0 预报的 500hPa 位势高度的均方差技巧评估（MSSS）结果与 TCC 比较类似，均表现为模式对热带和北太平洋大部分地区的环流形式把握能力较强，具体到系统上来说，模式对东亚地区中纬度地区的环流有较一定的预测技巧，对我国北部预报技巧较低，此外对热带外的预报技巧均随着 lead time 的延长而迅速减小。

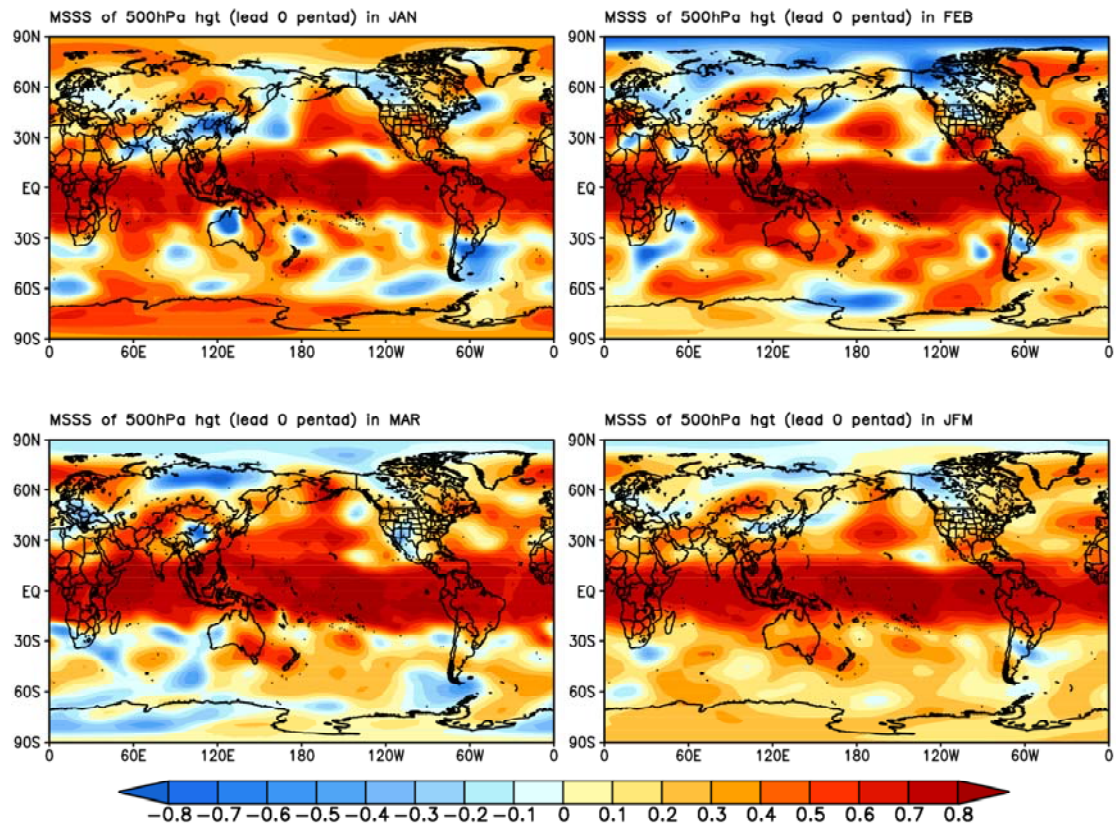
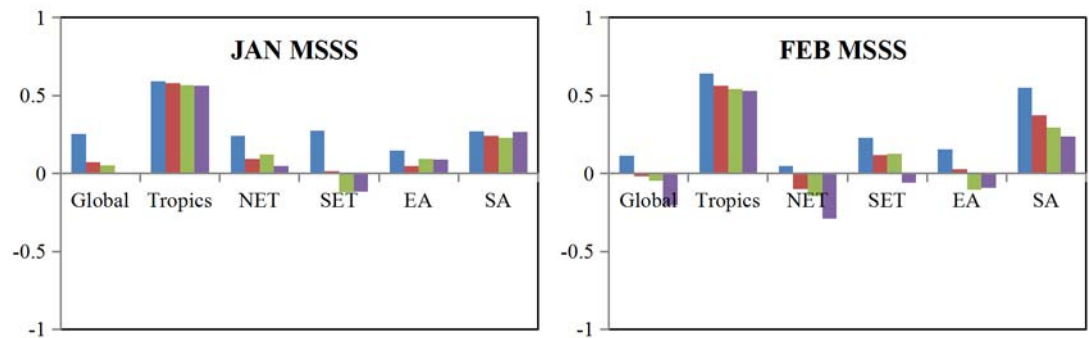


图 28 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的 500hPa 位势高度（HGT）的平均方差技巧（MSSS）分布



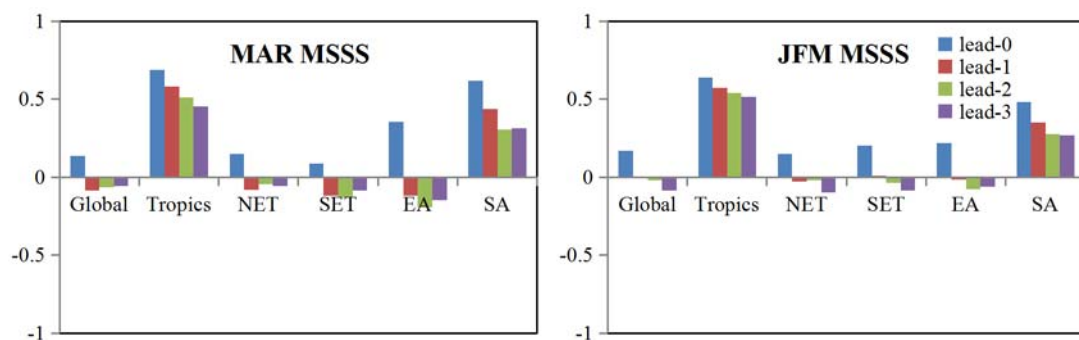


图 29 DERF2.0 在不同 lead time 起报的 1 月、2 月、3 月和第 1 季度的 500hPa 位势高度 (HGT) 的平均方差技巧 (MSSS) 在不同地区的平均

3.3.4 500hPa 位势高度评估小结

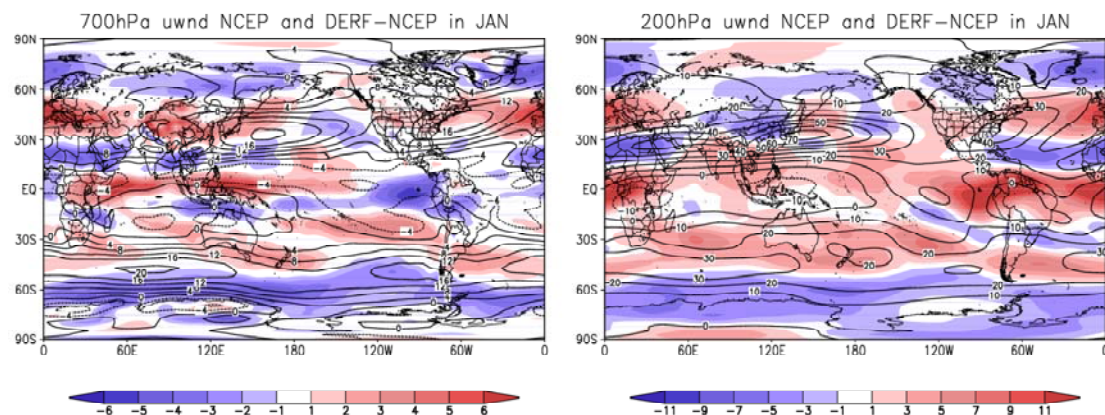
模式对 500hPa 位势高度的预测能力，代表着模式把握整体环流形势的性能，可以为预报员进行主观判断提供重要参考，对其预测能力的综合评估表明：首先，DERF2.0 对环流的把握能力较好，整体明显高于气温和降水；第二，模式主要对热带地区位势高度预报能力较强，特别对东亚中纬度地区预报技巧较高，有利于预报员通过对环流形势把握从主观上对降水和温度预报做出判断，但对我国东北部地区和欧亚大陆中高纬阻塞易发地区的位势高度场预报技巧偏低；最后，需要注意的是，模式在热带外地区，对初值十分敏感，随着 lead time 的延长，模式预报性能显著下降。

3.4 200hPa 和 700hPa 纬向风场

除 500hPa 位势高度的预测能力评估外，对 200hPa 和 700hPa 纬向风场的评估也是检验 DERF2.0 对高空急流和低空主要环流系统预报能力的关键要素。

3.4.1 气候态对比

同样首先对模式预报的 1 月、2 月和 3 月纬向风的气候态与观测进行对比分析。由图 30 所示，模式预报的气候态，在 700hPa 上从热带印度洋到东亚地区的赤道北侧有偏东风异常，在赤道附近的 MC 地区有西风异常；而在 200hPa 高层，北半球西风急流的强度略偏弱。此外，在低层赤道东太平洋的东风信风偏强，高层有异常西风。



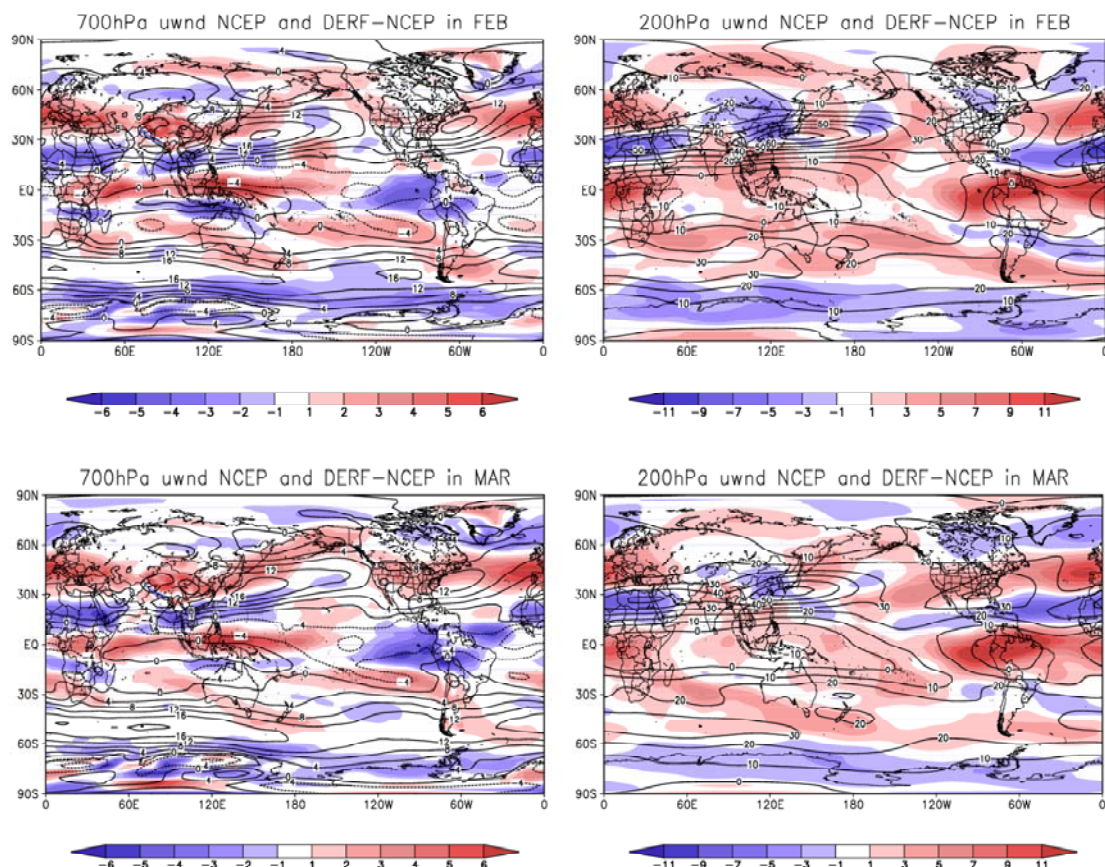


图 30 全球多年 (20a) 平均的 1 月、2 月、3 月 700hPa (左列) 和 200hPa (右列) 纬向风 (单位: m/s), 等值线代表 NCEP2 观测, 阴影为 DERF2.0 超前 0 候预报与观测之差

3.4.2 时间距平相关系数 (TCC)

图 31-图 33 给出高层和低层 DERF2.0 预报的纬向风的时间距平相关系数的空间分布和区域平均值。由图可见, 整体而言, 模式对高层 200hPa 风场的预报技巧更高, 在热带地区的预报技巧高于其他地区, 在 lead 0 pentad 时能够达到 0.6, 随 lead time 的延长技巧下降也较慢; 而在低层 700hPa, 模式对各个区域预报技巧类似, 特别的在热带和南亚地区预报技巧略高于其他地区。

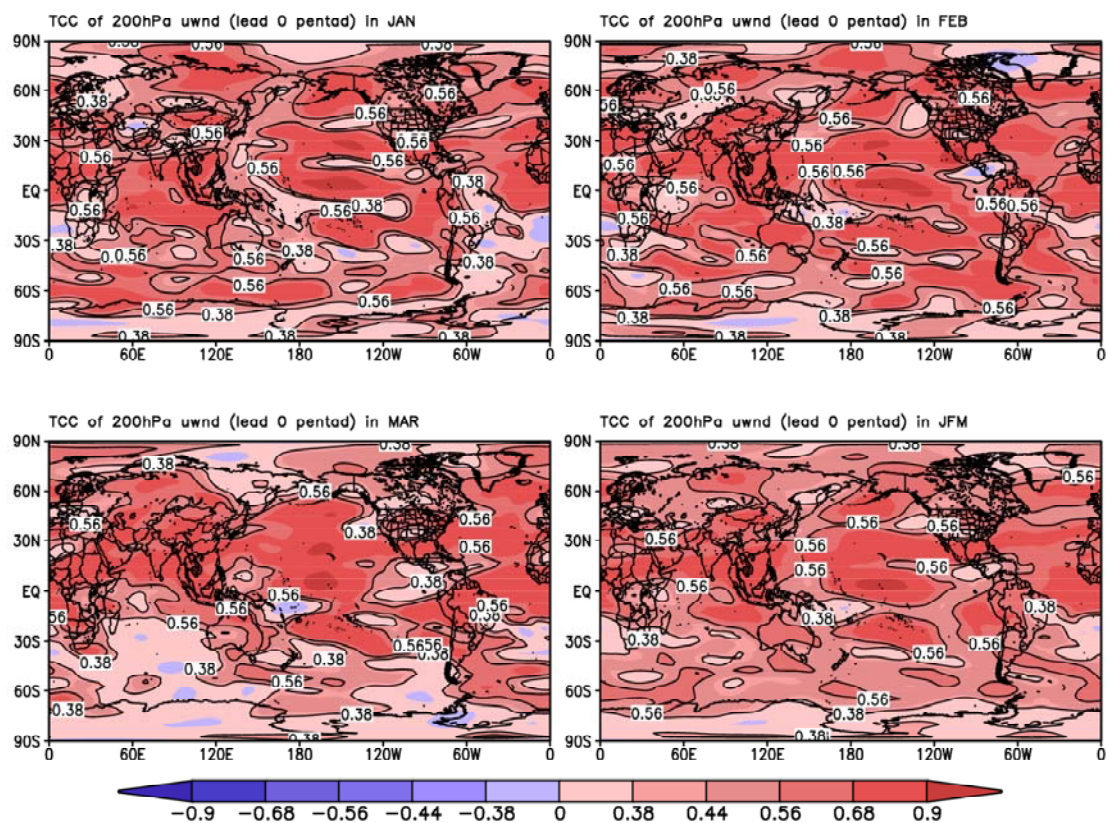


图 31 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的 200hPa 纬向风的时间距平相关系数(TCC) 分布，等值线标定分别代表通过 0.1 和 0.01 显著性检验

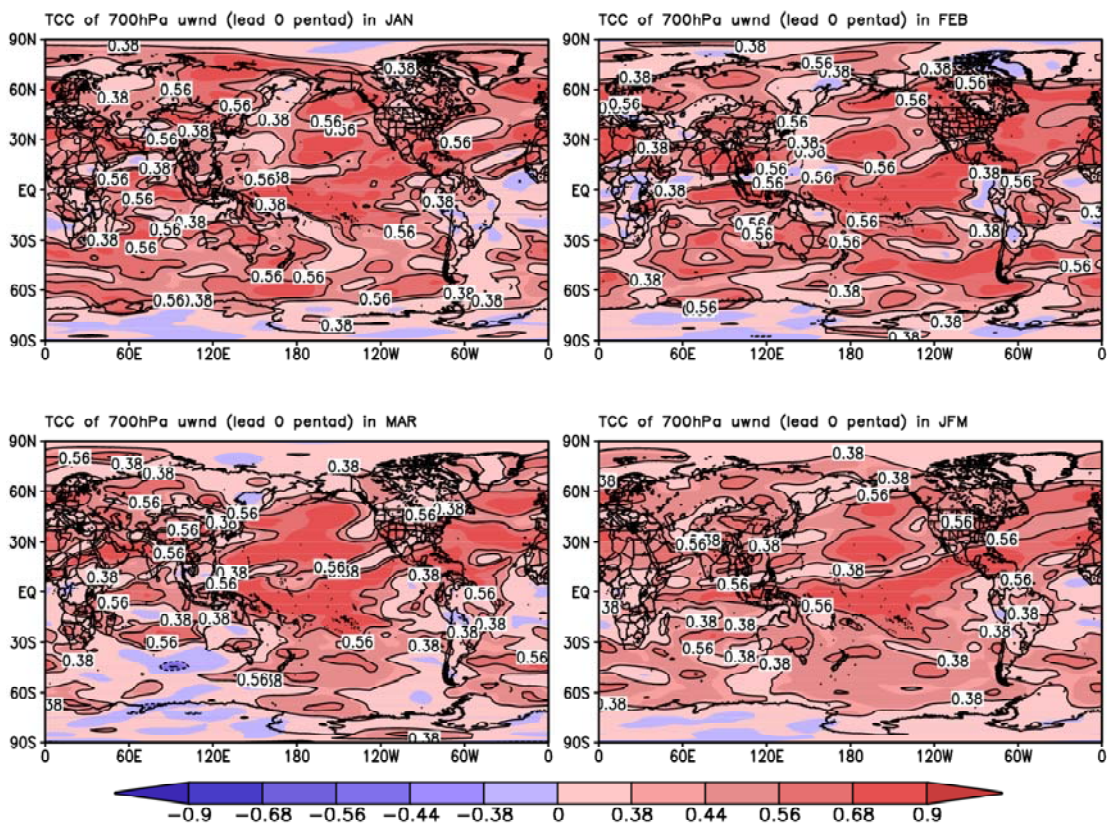


图 32 同图 31，但为 700hPa 纬向风的时间距平相关系数 (TCC) 分布

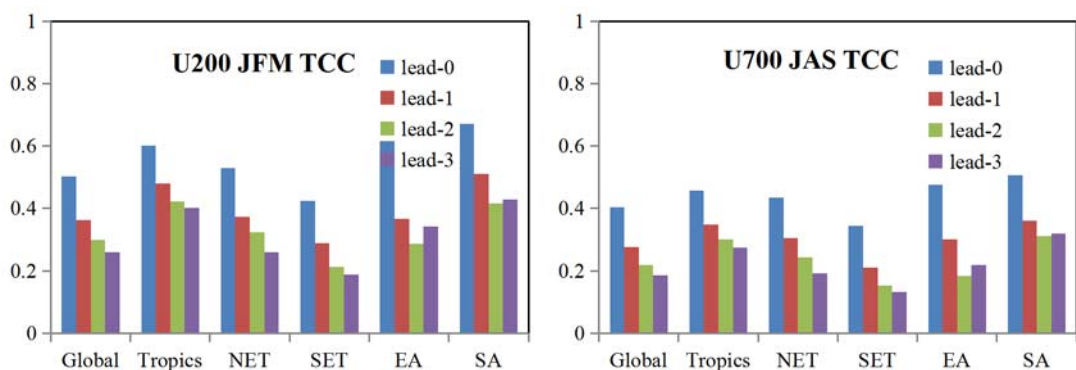


图 33 DERF2.0 在不同 lead time 起报的第 1 季度的 200hPa 和 700hPa 纬向风的时间距平相关系数 (TCC) 在不同地区的平均

3.4.3 均方技巧评分 (MSSS)

图 34-图 35 给出高层和低层 DERF2.0 预报的纬向风的均方技巧评分的空间分布和区域平均值。在北半球印度洋和东亚大陆地区，200hPa 纬向风的 MSSS 评分较高，且在太平洋大部分地区，200hPa 和 700hPa 纬向风的预报技巧均较高。

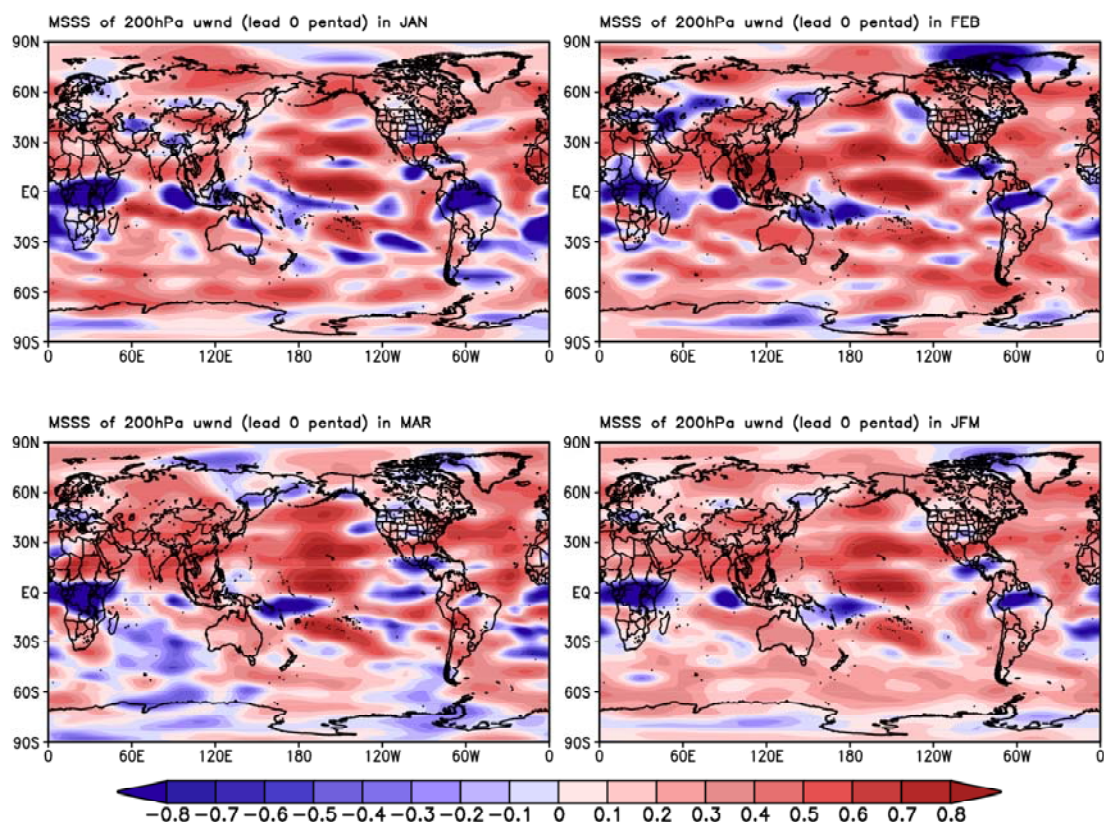


图 34 DERF2.0 超前 0 候起报的全球 1 月、2 月、3 月和第 1 季度的 200hPa 纬向风的平均方差技巧 (MSSS) 分布

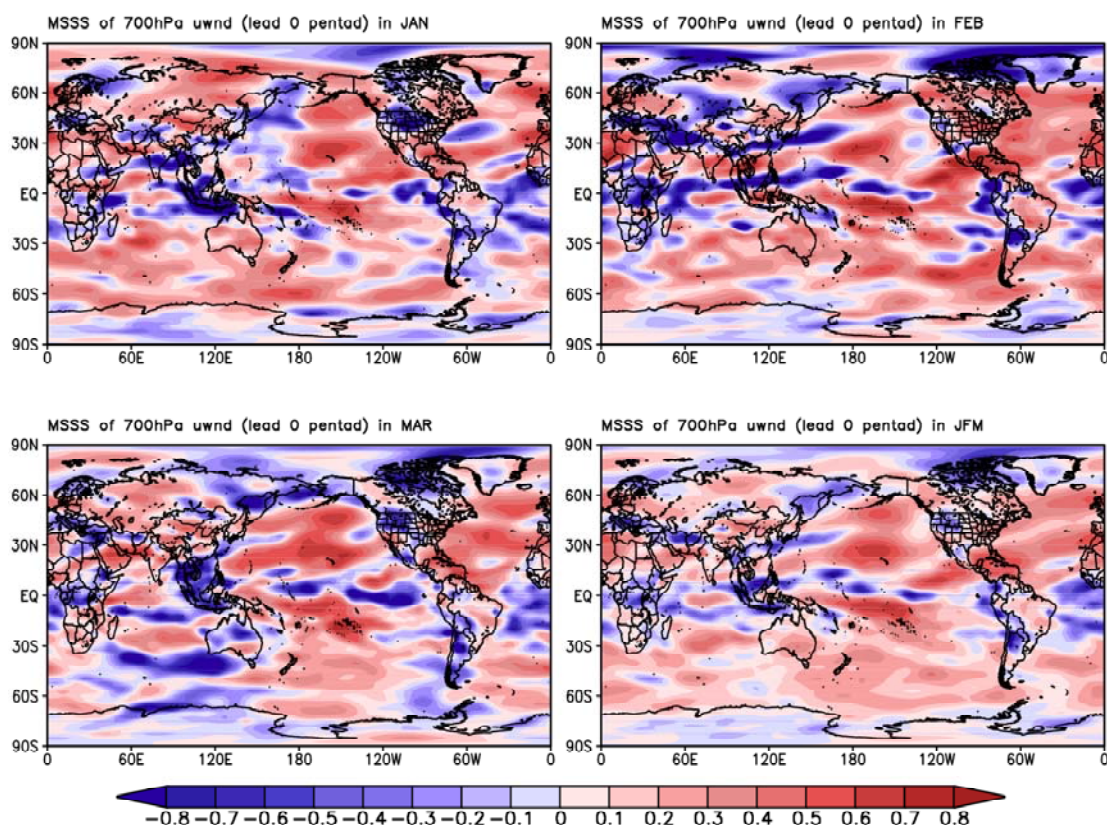


图 35 同图 34，但为 700hPa 纬向风的平均方差技巧（MSSS）分布

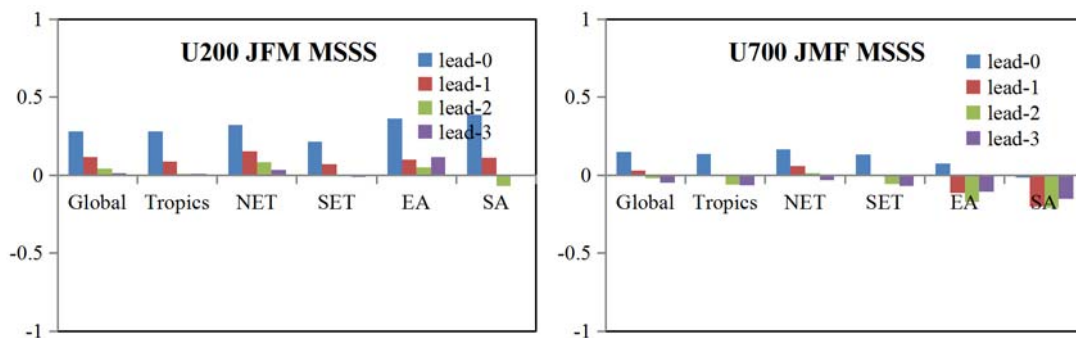


图 36 DERF2.0 在不同 lead time 起报的第 1 季度的 200hPa 和 700hPa 纬向风的平均方差技巧（MSSS）在不同地区的平均

3.4.4 纬向风场评估小结

模式对高低层纬向风场的预测，如对高空急流，低层西南季风的预测能为预报员主观判断降水异常的大小和区域提供重要参考，对其预测能力的综合评估表明：首先，DERF2.0 的气候态，在第 1 季度存在着一定的系统偏差，表现为高层西风急流较弱，低层北印度洋上有偏东风；第二，DERF2.0 对纬向风场的把握能力好于气温和降水，但要低于 500hPa 位势高度的预报，在高层模式主要对热带地区，特别是热带太平洋纬向风场预报能力较强。

4 概率预报评估（ROC）

在概率预报评估中，我们采用 ROC（Relative Operating Characteristics）方法，对各个

变量场的概率预报结果分别进行了检验。定义个变量场距平介于最大值和最小值之间范围 1/3 为负异常 (Below Normal, BN)、介于 1/3 至 2/3 之间为常值 (Near Normal, NN), 大于 2/3 为正异常 (Above Normal, AN), 分别给出三类事件的全球平均 ROC 曲线和 ROC 面积 (ROCA) 的水平分布图。当 ROC 曲线高于对角线以及 ROC 面积指数大于 0.5 时视为有预报技巧, 且数值越高预报技巧越高。

4.1 2m 气温 (t2m)

对 2m 气温场 (T2m) 的 ROC 评估表明, 模式在热带外中高纬地区的预报技巧最高 (图 38), 在赤道附近则较低, 且模式对高温和低温的预报技巧相对较高, 即有相对较多的样本能够准确判断出 2m 气温高/低事件出现与否, 特别是对高温异常出现与否把握能力较强。模式预报技巧对初值十分敏感, 随着 lead time 的延长, 模式预报的准确性逐渐下降 (图 37)。

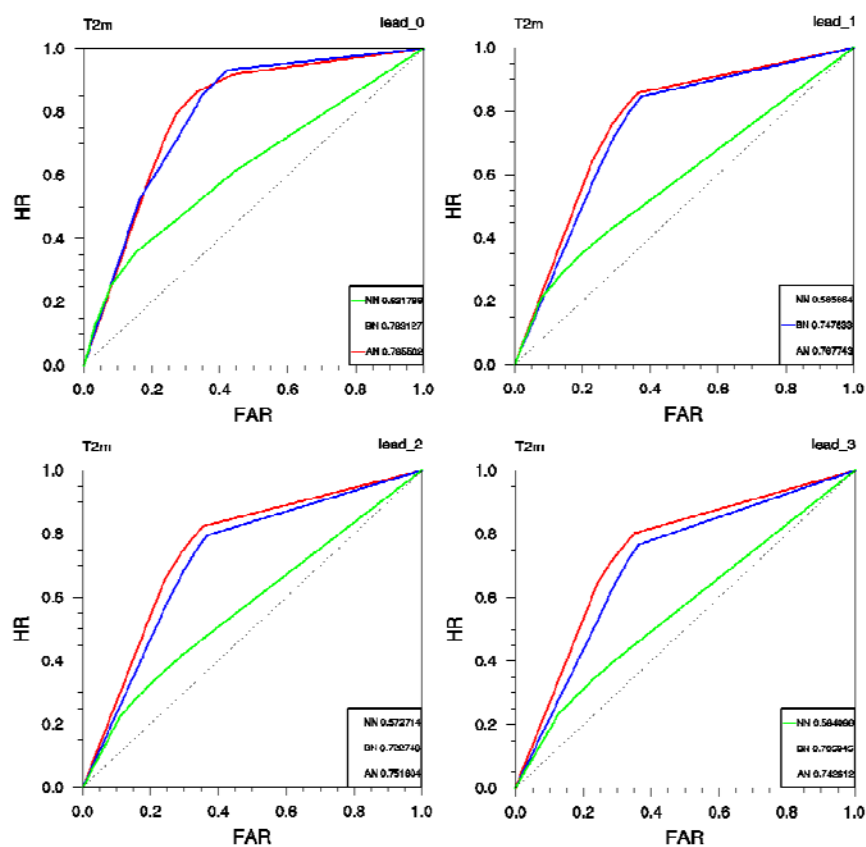


图 37 1991-2010 年 1-3 月不同 lead time 起报的 2m 温度 ROC 曲线

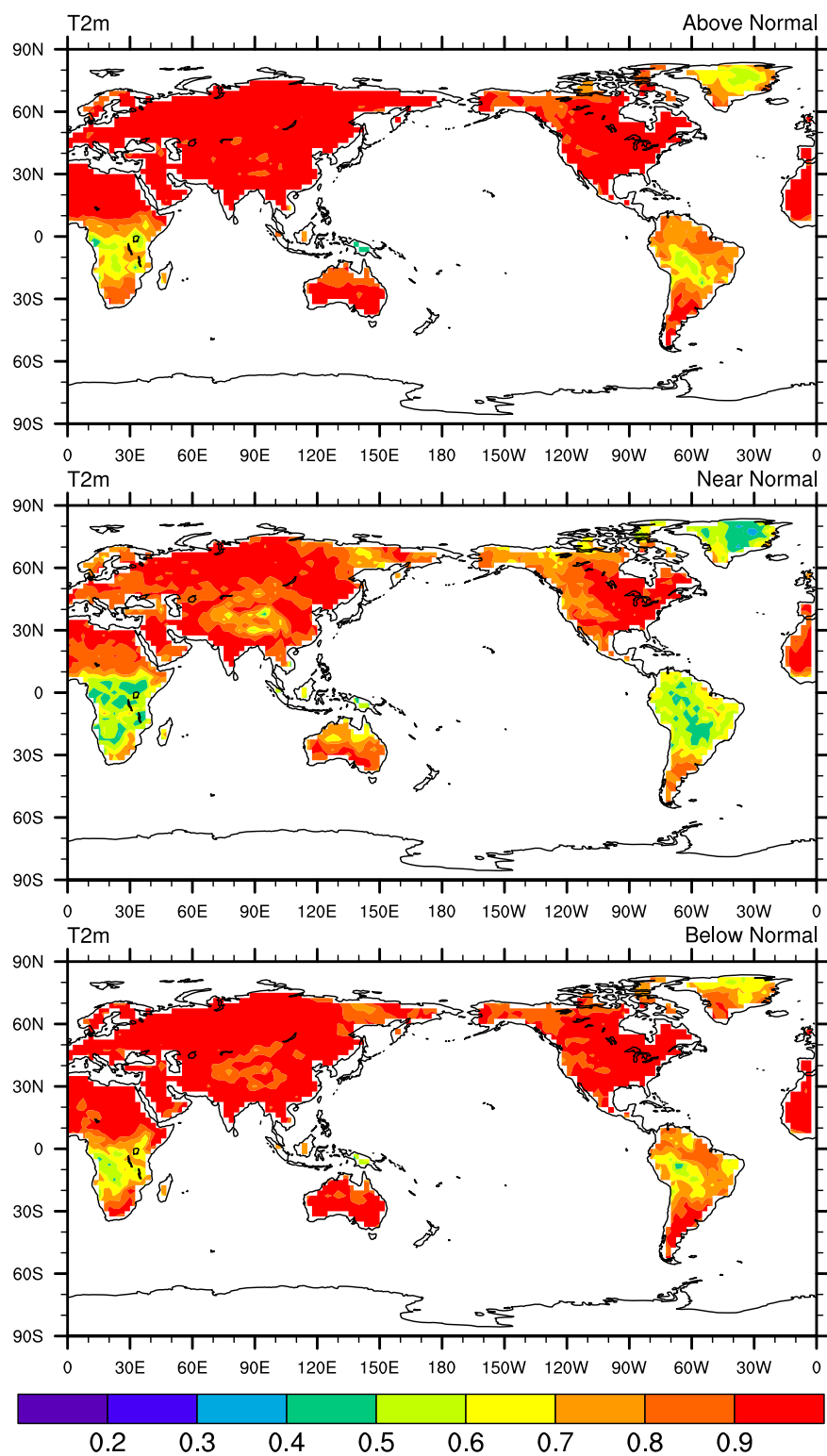


图 38 1991-2010 年 1-3 月超前 0 候的 2m 温度 ROCA 面积指数

4.2 降水 (prec)

模式对降水场预报技巧与 2m 温度相比较低 (图 39), 仅略高于随机预报, 对不同地区而言, 降水场的高/低值 ROCA 指数在赤道附近, 尤其是太平洋地区, 正负异常场的 ROCA 均可达到 0.8 以上 (图 40), 具有一定的预报技巧。

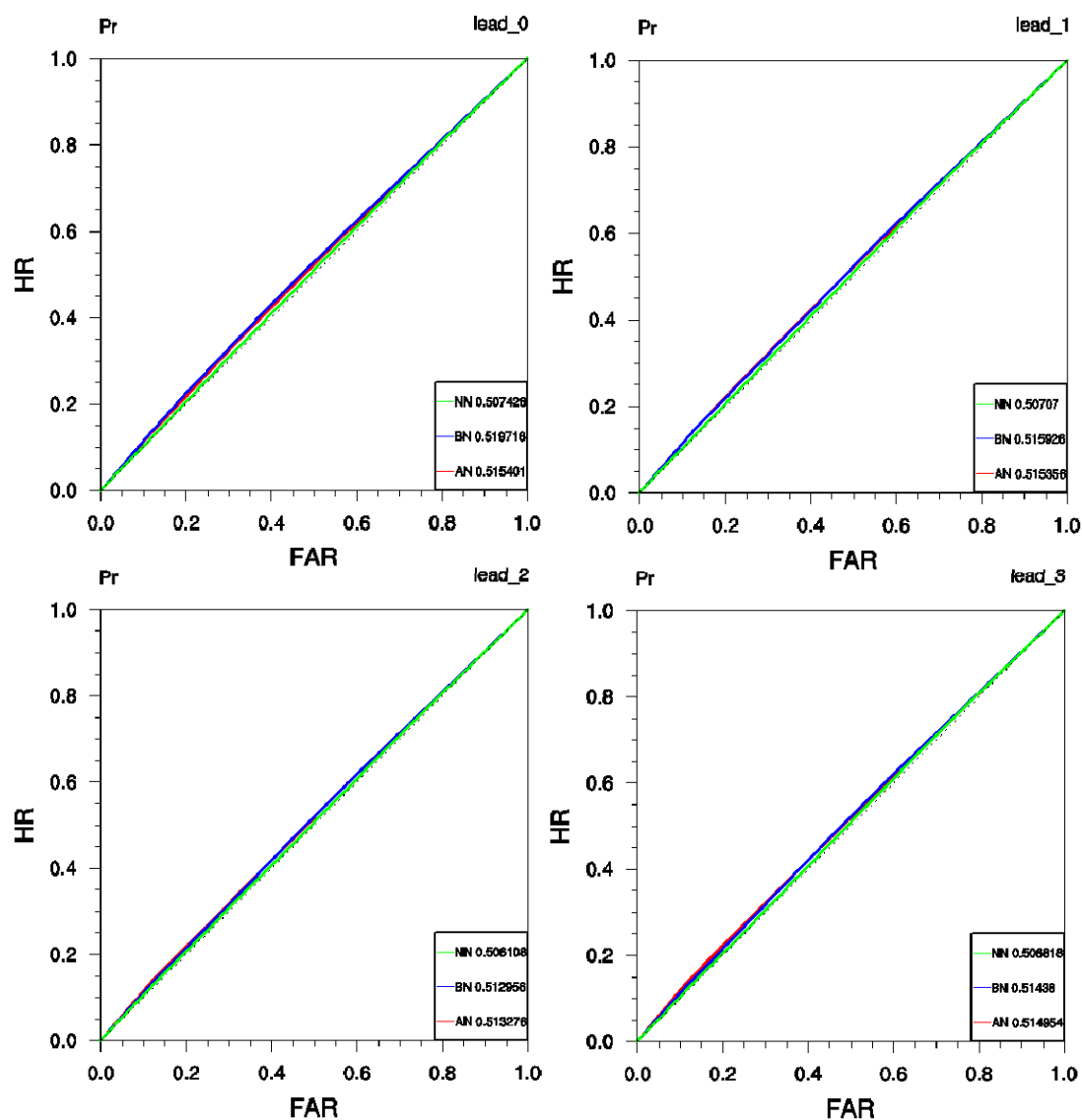


图 39 1991-2010 年 1-3 月不同 lead time 起报的降水 ROC 曲线

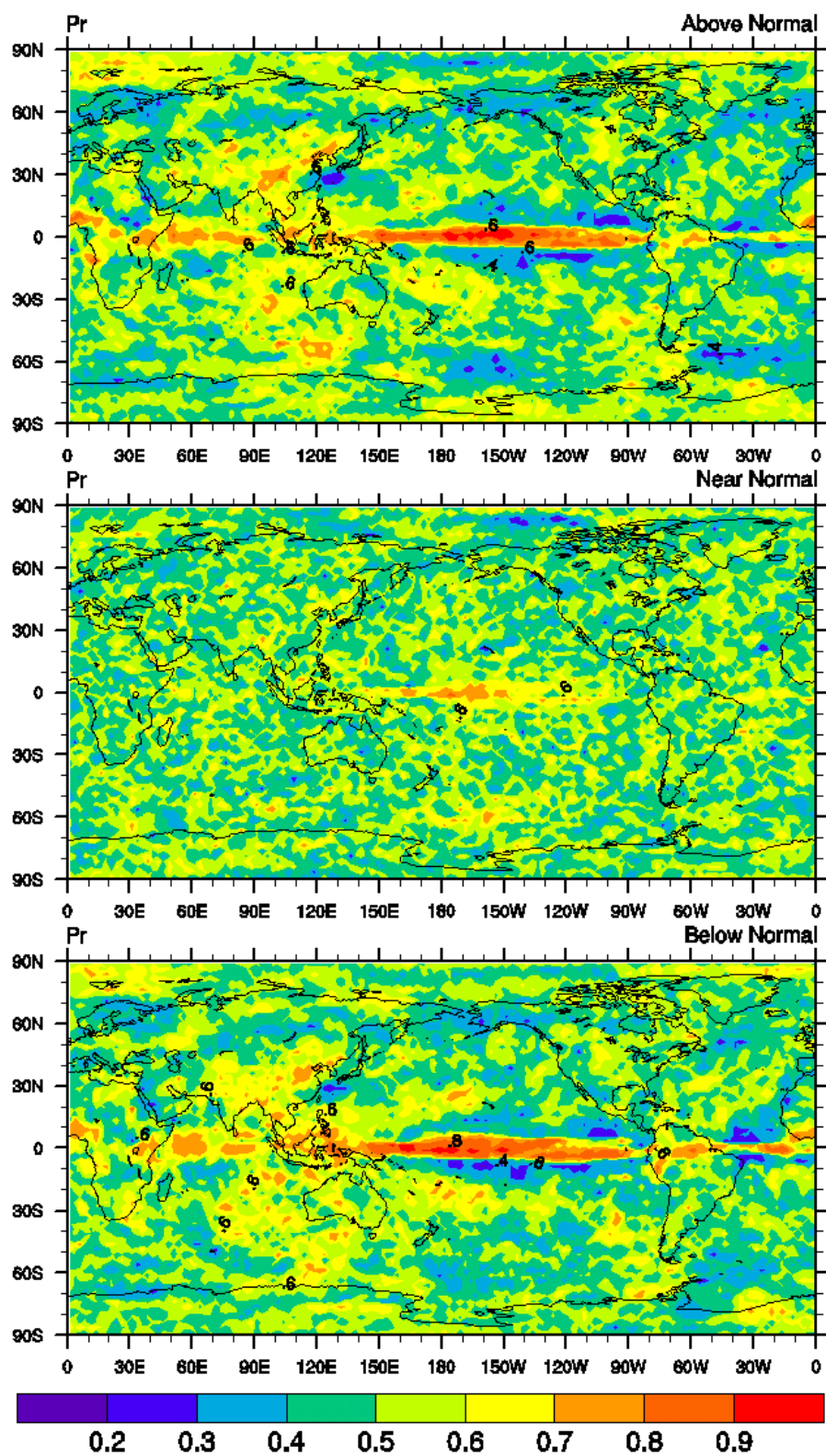


图 40 1991-2010 年 1-3 月超前 0 候的降水 ROCA 面积指数

4.3 500hPa 位势高度

如图41所示，在超前0-4候的模式预报结果中，全球平均的1991-2010年1-3月500hPa位势高度ROC曲线表明模式对500hPa位势高度的高低值的预报技巧比常值相对较高。热带地区的ROCA指数相对较大，即在热带地区模式具较高的预报技巧。在500hPa位势高度高值和低值ROCA水平分布图中，热带地区ROCA指数可达到0.6-0.8。在热带大西洋、海洋大陆、东太平洋、南美以及非洲大陆地区，ROCA指数均可达到0.7以上（图42）。

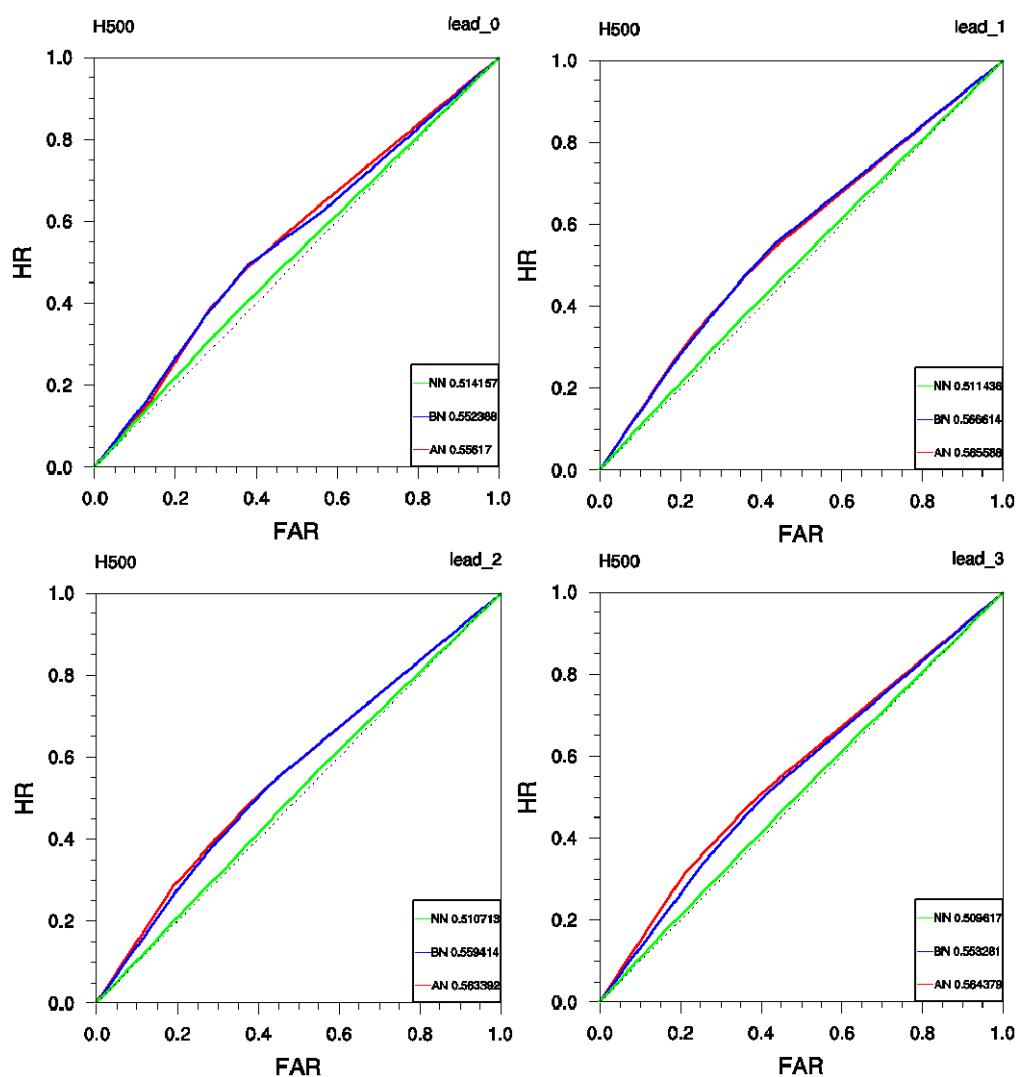


图 41 1991-2010 年 1-3 月不同 lead time 起报的 500hPa 位势高度 ROC 曲线

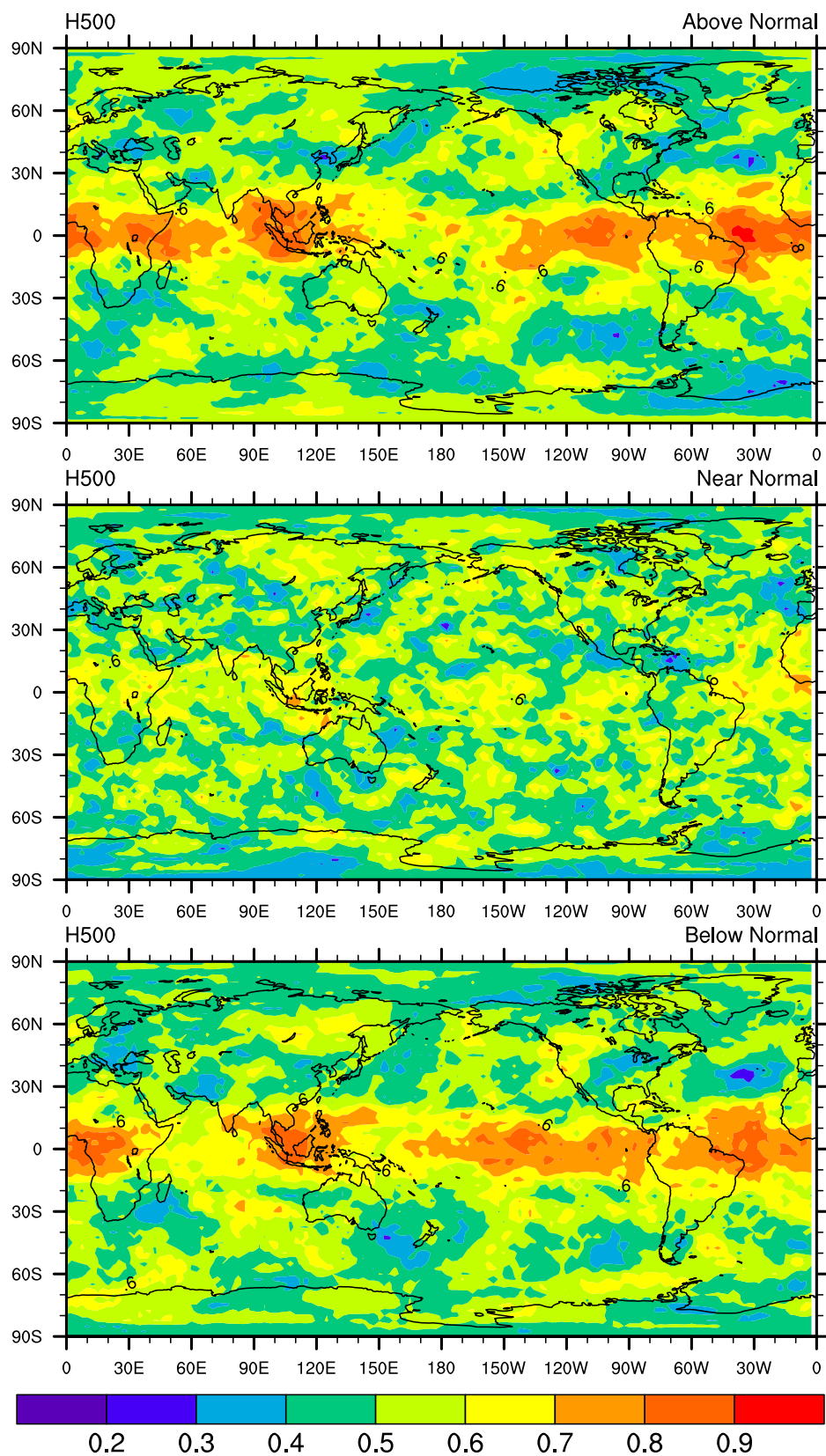


图 42 1991-2010 年 1-3 月超前 0 候的 500hPa 位势高度 ROCA 面积指数

4.4 纬向风场

1991-2010 年 1-3 月 200hPa 纬向风和 700hPa 温度场的全球平均结果表明，模式对环流场的预报技巧略高于降水场，但仍然远低于 500hPa 位势高度场，因此其预报价值较低（图 43）。如图 44 所示，在 ROCA 指数的水平分布图中，赤道太平洋地区的 200hPa 纬向风的预报技巧均相对其地区较高，可以达到 0.6-0.7 以上。700hPa 纬向风的预报技巧低于 200hPa 纬向风，其 ROCA 指数仅在赤道西太平洋略高（图 46）。

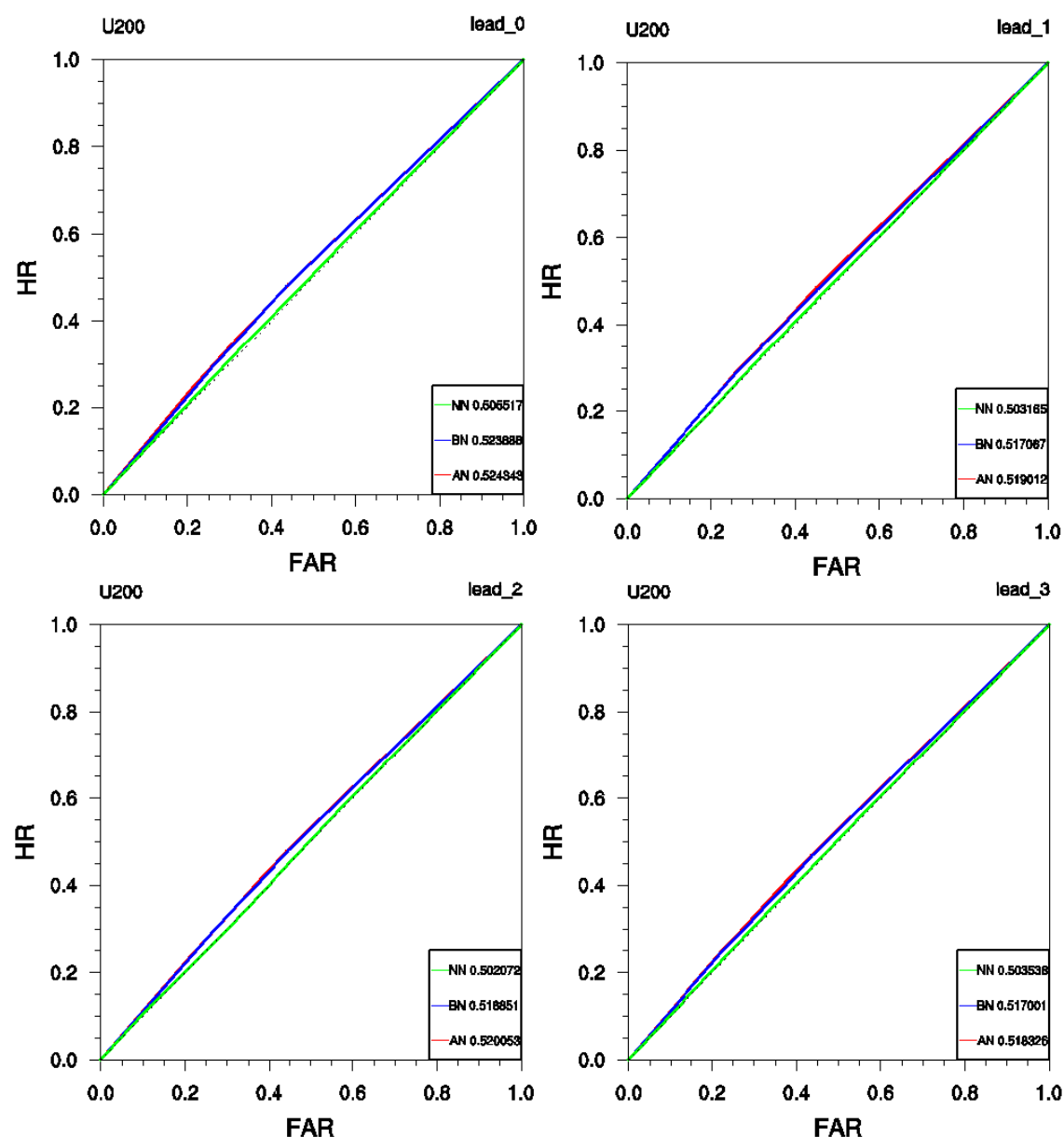


图 43 1991-2010 年 1-3 月不同 lead time 起报的 200hPa 纬向风 ROC 曲线

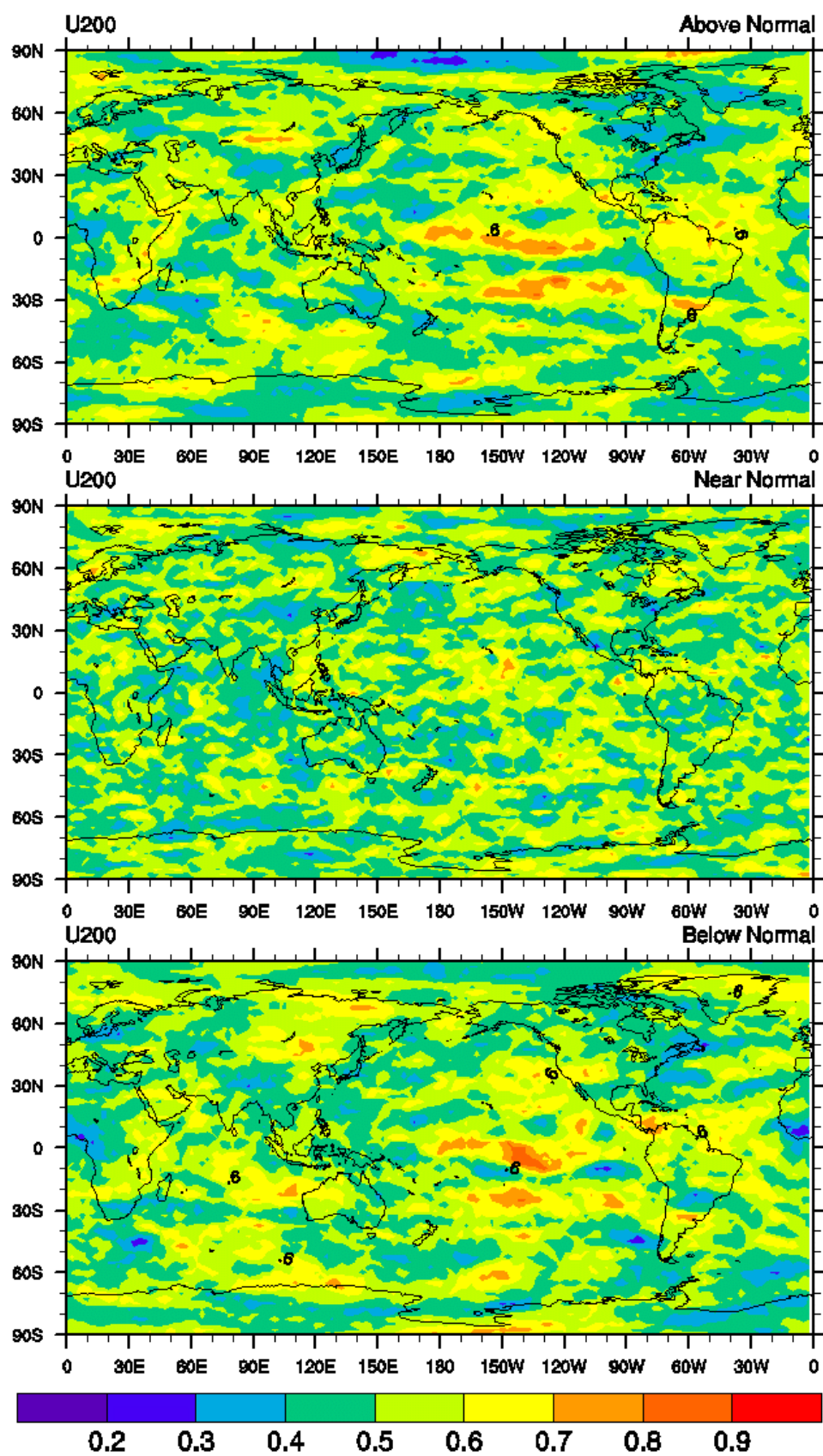


图 44 1991-2010 年 1-3 月超前 0 候的 200hPa 纬向风 ROCA 面积指数
超前 0 候的 200hPa 纬向风 ROCA 面积指数

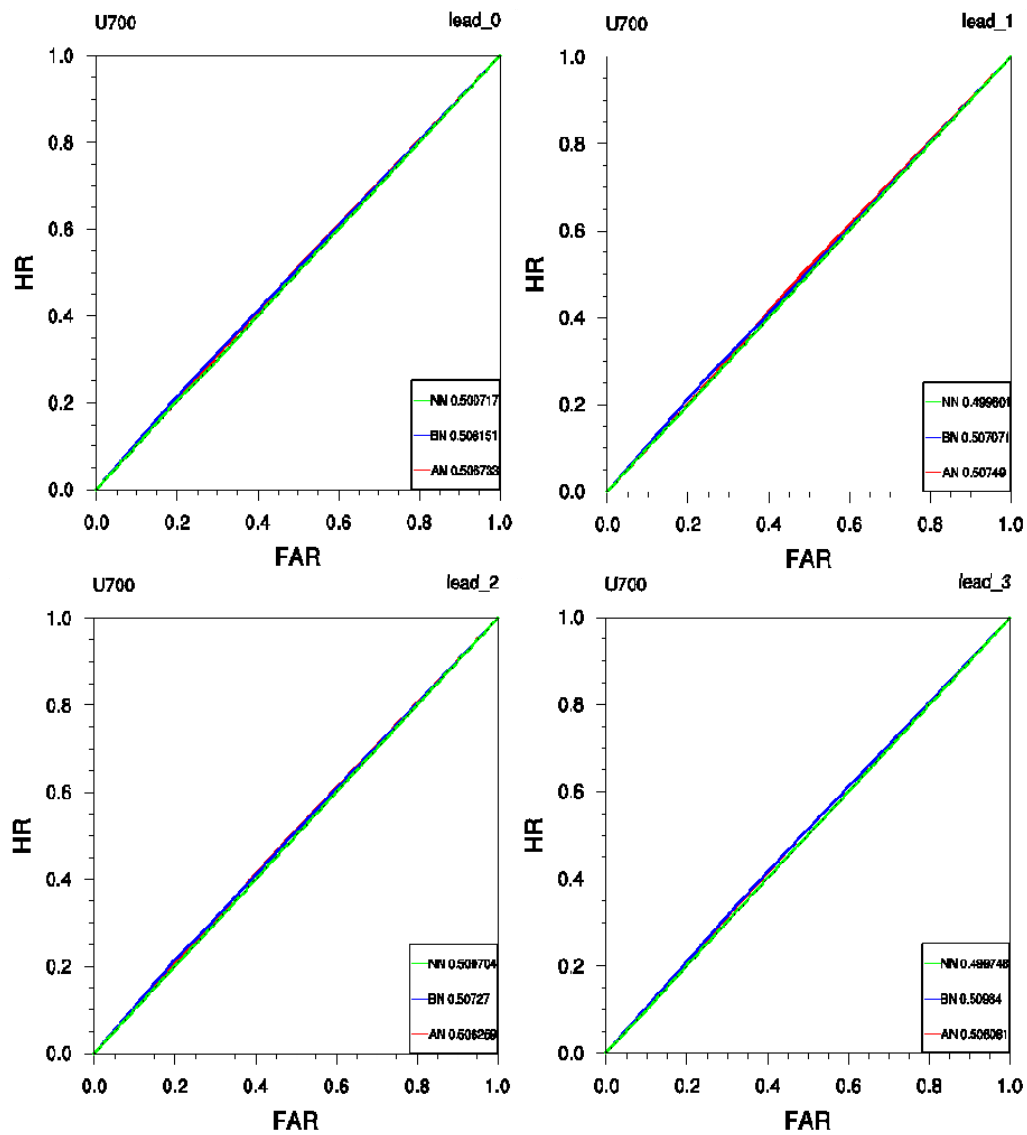


图 45 1991-2010 年 1-3 月不同 lead time 起报的 700hPa 纬向风 ROC 曲线

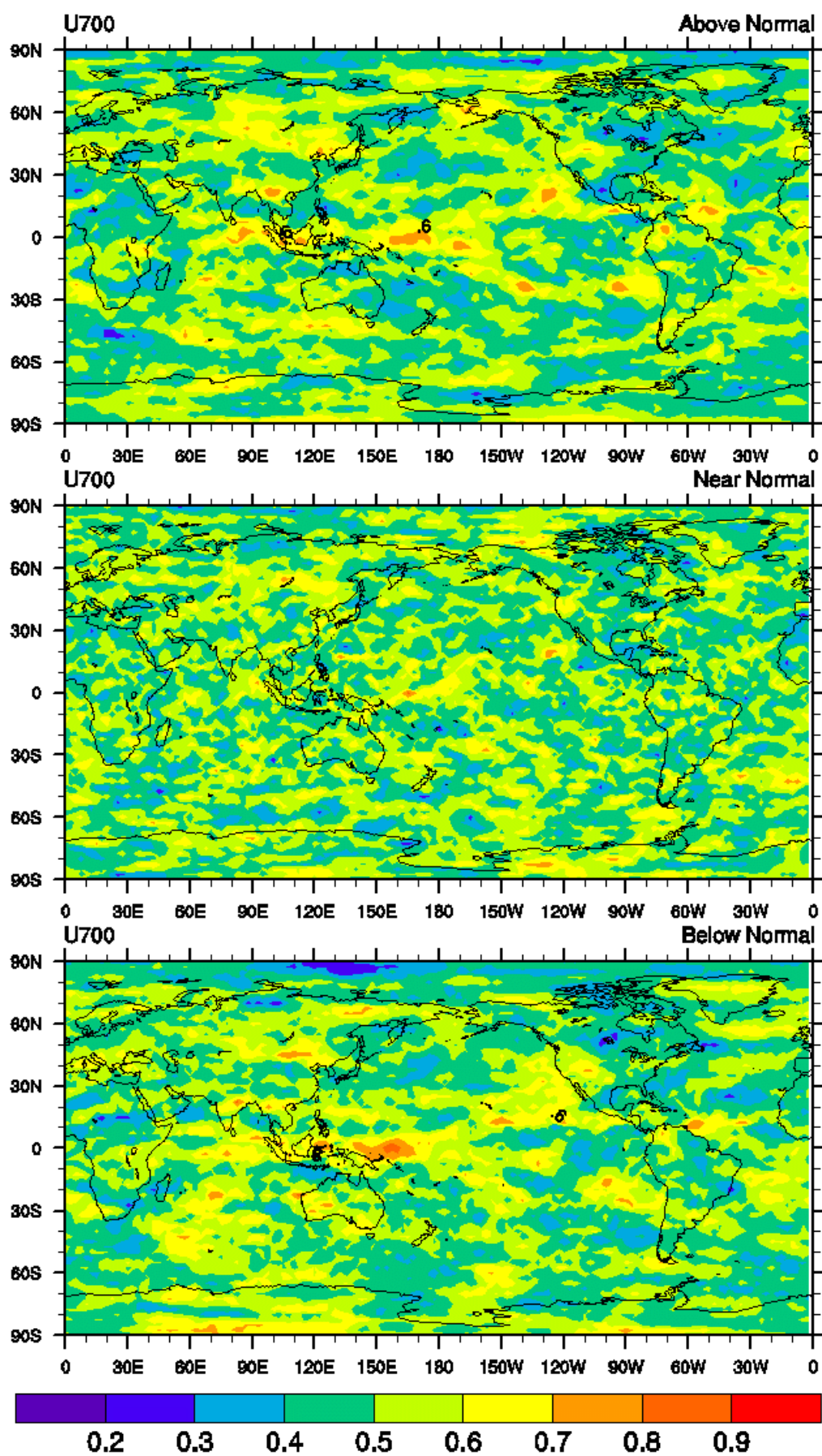


图 46 1991-2010 年 1-3 月超前 0 候的 700hPa 纬向风 ROCA 面积指数

4.5 ROC 评估小结

利用 ROC 方法对超前 0-4 候模式预报结果的检验表明，气温场（T2m）在热带外地区的预报技巧较高，即有相对较多的样本能够准确判断出热带外地区气温的高/低/常值事件出现与否。模式对环流场的高值和低值预报技巧在热带地区相对于其它纬度较高；而降水场的高值和低值预报技巧，除了在赤道附近相对较高外，在我国中部地区也具有较高的预报技巧，可为模式预报结果的校正提供参考。

5 对延伸期降水预报的检验评估

5.1 确定性预报检验

5.1.1 时间距平相关系数（TCC）

与对月尺度降水的评估类似，这里对第 1 季度期间模式对逐候降水计算时间距平相关系数，定量分析模式对不同地区延伸期降水的预报能力。如图 47 所示，在第 1 候，全球平均的距平相关系数约为 0.4，相关技巧较高的区域主要位于赤道中太平洋、北大西洋和印度洋西北地区，此外，欧亚大陆中纬度地区的降水的预报技巧也较高；到预报的第 3 候，热带外降水的相关系数已经迅速下降到 0.1，高相关区域仍然维持在赤道中太平洋地区，不同区域比较而言，模式对东亚地区的降水预报性能较差。

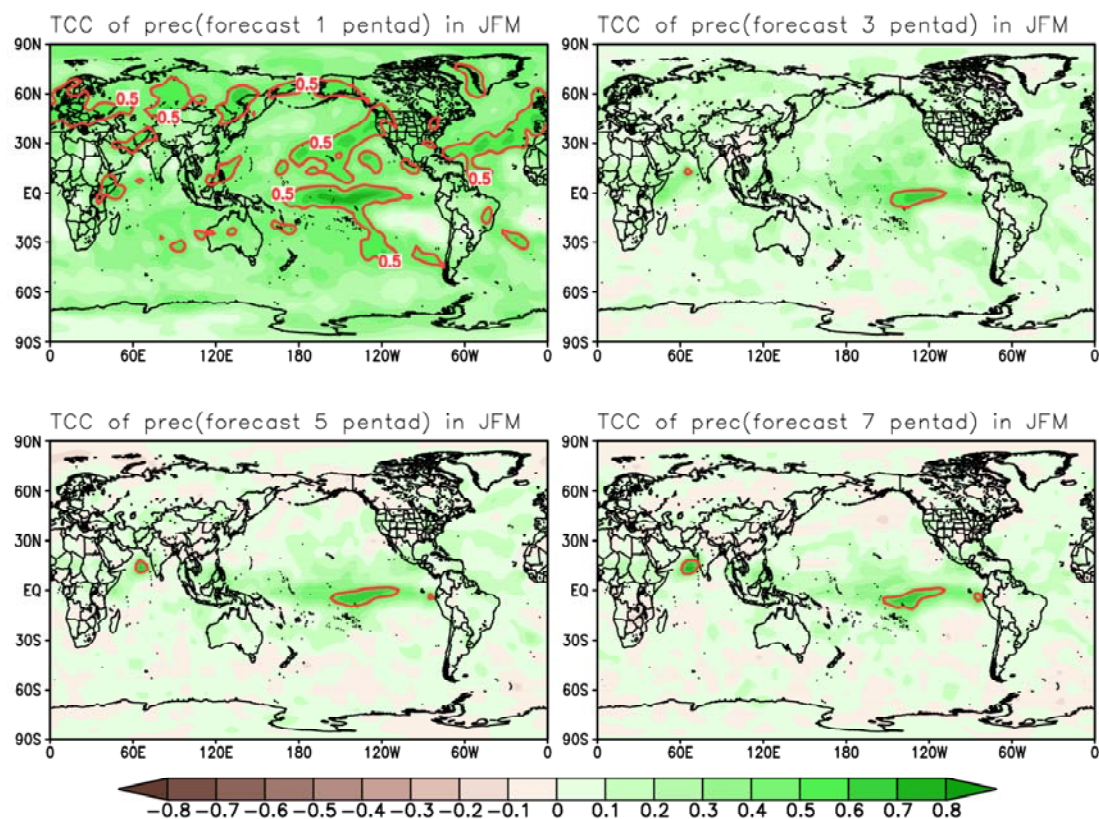


图 47 1991-2010 年第 1 季度预报第 1 候、第 3 候、第 5 候和第 7 候降水异常时间相关系数指数

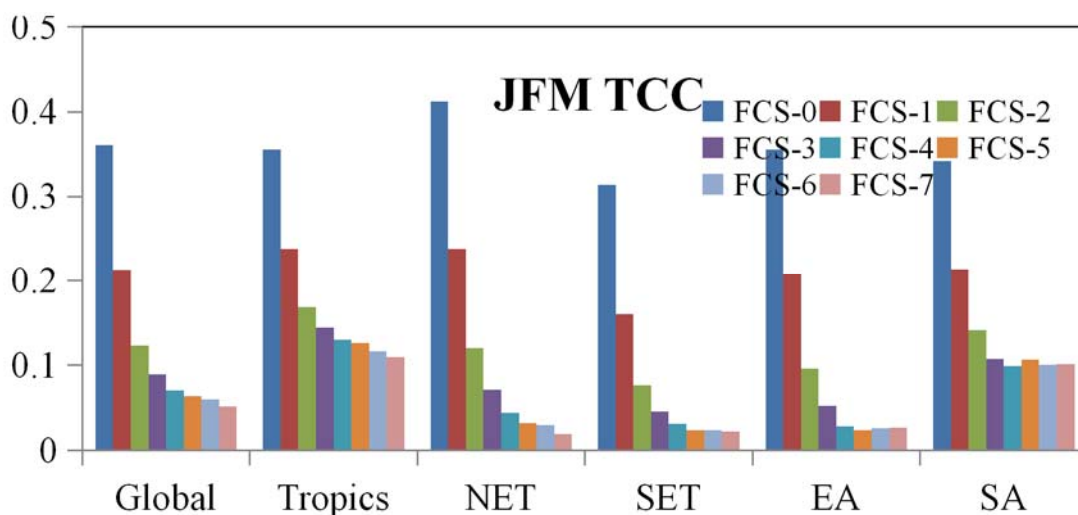


图 48 1991-2010 年第 1 季度预报第 1 候、第 3 候、第 5 候和第 7 候不同区域的降水异常时间相关系数指数

5.1.2 均方技巧评分 (MSSS)

均方技巧评分与 TCC 结果较为接近，模式对欧亚大陆中高纬降水的预报技巧较高，在 3 候之后，模式对大部分地区的降水均失去预报技巧。此外，模式对我国西部降水的预报技巧很低。

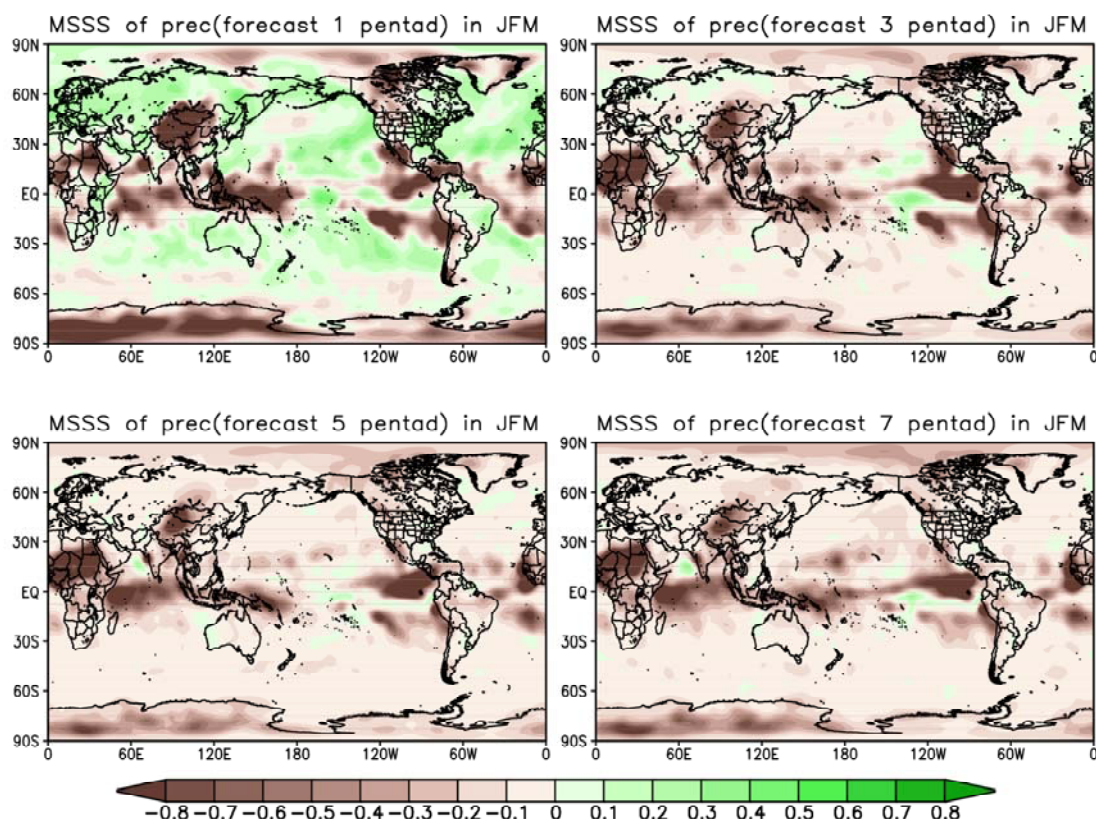


图 49 同 47，但为均方根技巧评分 (MSSS)

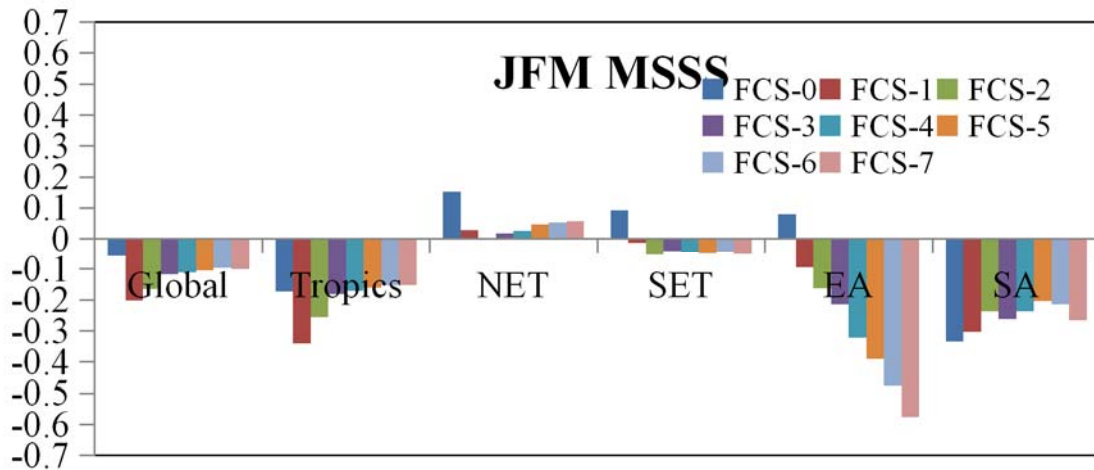


图 50 同 48，但为均方根技巧评分（MSSS）

5.2 概率预报检验（ROC/RD）

下面使用 ROC（Relative Operating Characteristics）和 RD（Reliability diagrams）对逐候降水概率预报进行评估检验，定义降水距平百分率大于 20% 为降水偏多（Above Normal, AN）、介于正负 20% 之间为正常（Near Normal, NN），小于 20% 为偏少（Below Normal），分别给出三类事件的 ROC 面积（ROCA）和 ROC/RD 折线图。当 ROC 面积指数大于 0.5 时视为有预报技巧，ROC 面积指数越高、RD 折线越接近对角线，则预报技巧越高。

如图 51-图 56 所示，模式对 AN 和 BN 事件的预报技巧十分接近，均高于 NN 事件的预报技巧，特别对热带和赤道太平洋地区，在预报的第 1 候，全球大部分地区的 ROC 面积指数均超过 0.6，赤道中太平洋超过 0.8，在第 3 候，ROCA 迅速减小，模式对我国东部地区的降水预报技巧较高，在 5 候之后，具有较高的预报技巧只停留在赤道中东太平洋、西印度洋和 MC 地区，对我国而言，模式对我国东部降水的预报技巧较低，对 AN 事件而言，随着模式预报时间的增长，模式较大概率的预报反而更加可信（图 52）。

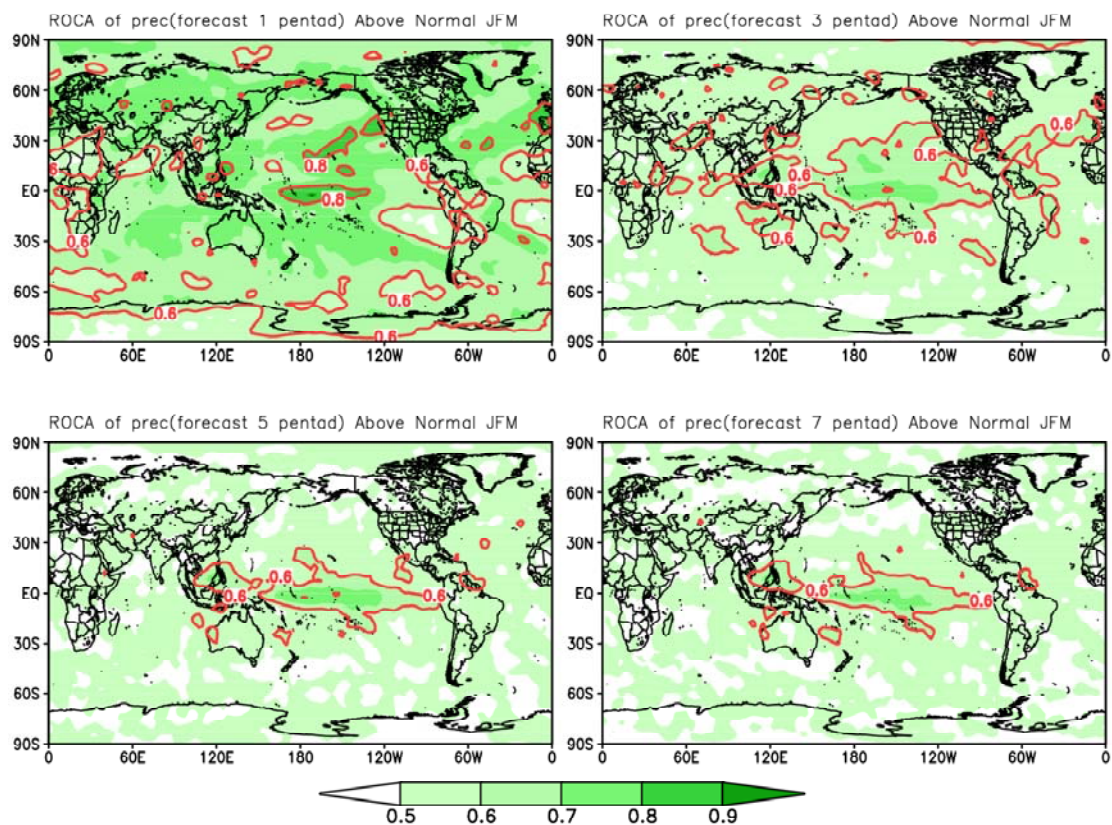


图 51 1991-2010 年第 1 季度预报第 1 候、第 3 候、第 5 候和第 7 候降水偏多事件的 ROCA 面积指数

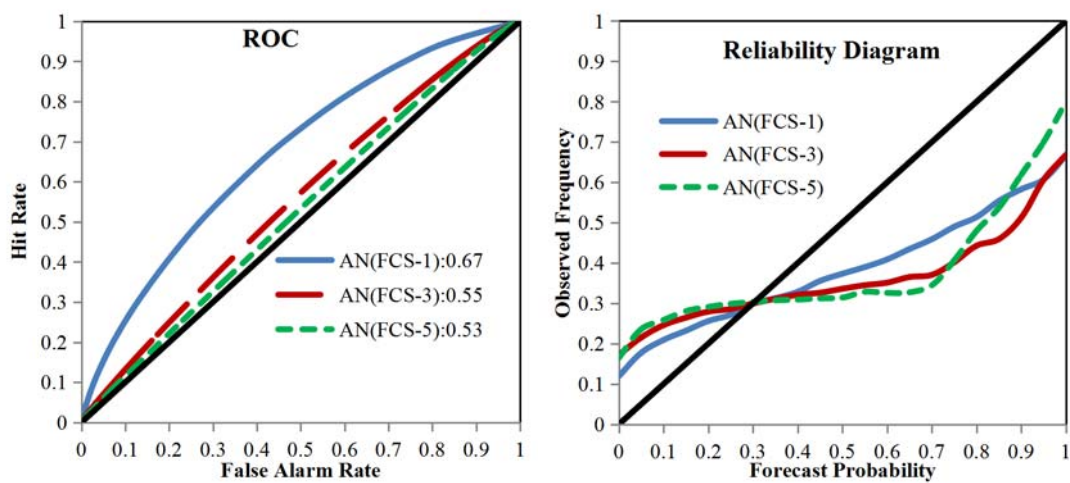


图 52 1991-2010 年第 1 季度预报第 1 候、第 3 候和第 5 候降水偏多事件的 ROC 和 Reliability Diagram 曲线

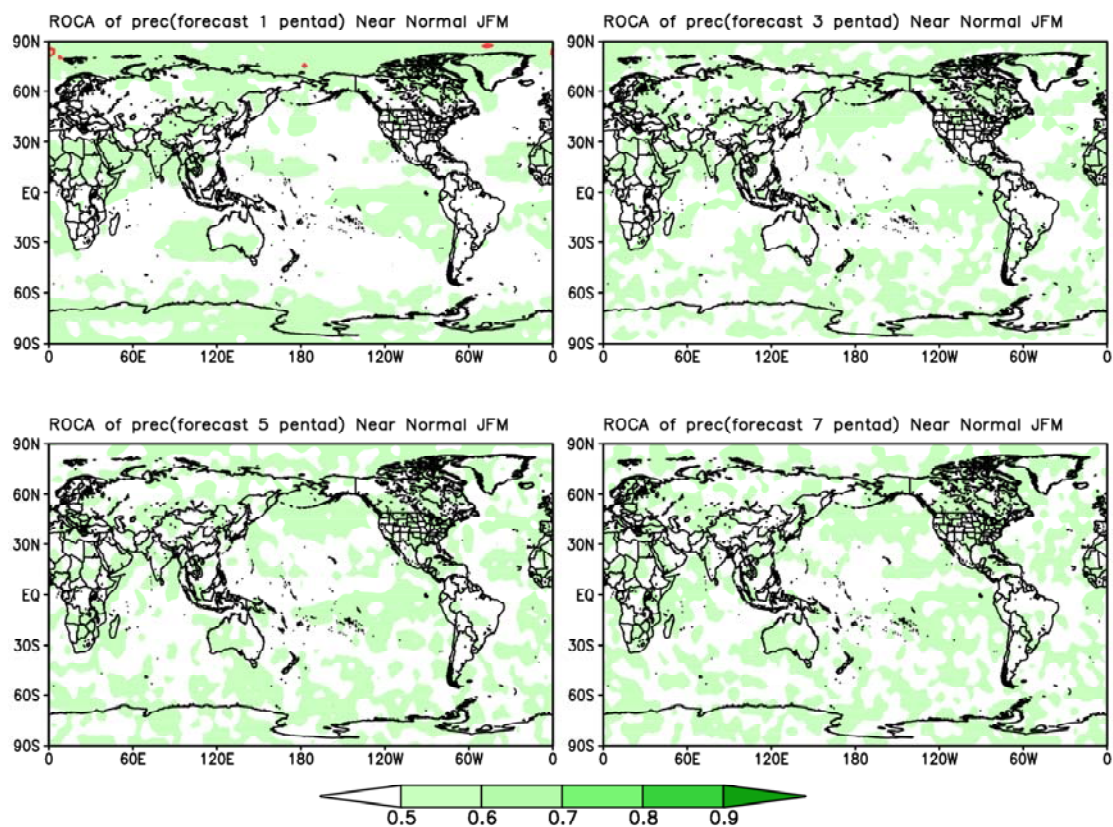


图 53 同图 51，但为预报降水正常事件

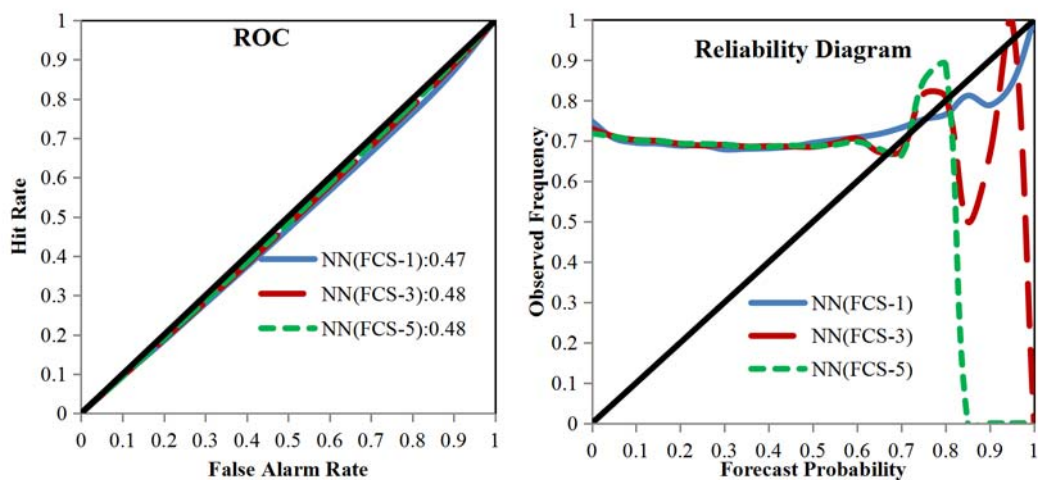


图 54 同图 52，但为预报降水正常事件

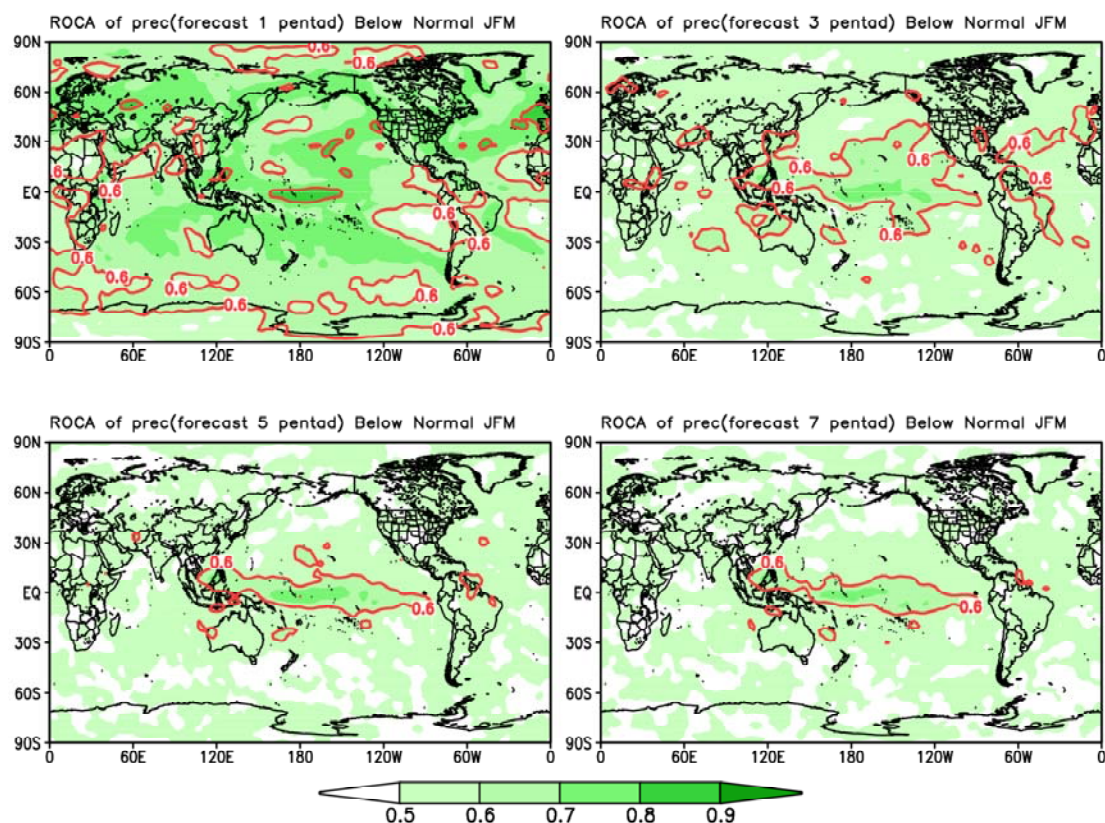


图 55 同图 51，但为预报降水偏少事件

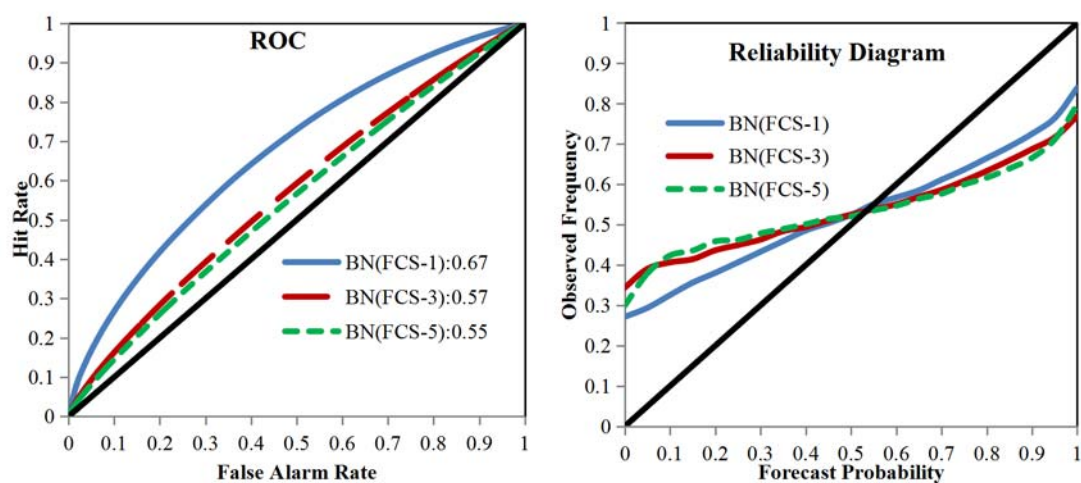


图 56 同图 52，但为预报降水偏少事件

5.3 延伸期降水预报小结

本小节分别使用确定性预报评估方法（TCC/MSSS）和概率预报评估方法（ROC/RD）对降水延伸期预报（1-8 候）技巧进行评估，结果表明，首先，模式对赤道中东太平洋降水的预报技巧较高，可持续 7-8 候，而对热带外预报技巧较低，预报技巧随时间的延长很快减小；第二，具体到我国而言，在第 3-4 候，模式对我国东部地区的预报技巧相对较高；第三，从概率预报上看，模式对 AN/BN 事件的预报技巧类似，均高于对 NN 事件的预报技巧。

6 基于 DERF2.0 实时预报产品

6.1 MJO 监测预测实时产品

国家气候中心第二代模式 DERF2.0 已经于 2013 年开始业务运行，利用其实时输出产品以及实时获取的全球分析和再分析资料（使用 NCEP1 及气象中心提供的 T639 资料，包括全球 850 百帕纬向风、200 百帕纬向风、OLR 等变量），采用 Wheeler 和 Hendon（2004）提出的 MJO 实时信号提取方法，发展出适用于 DERF2.0 系统的 MJO 的实时监测和预测技术，并建立自动化运行业务系统。运用这一系统，可实现针对 MJO 现象的实时监测诊断和预测，提供过去一个月的 MJO 指数监测和未来一个月 MJO 预测指数相限图及数据，并能提供与 MJO 相关联的重建逐日实时纬向风和 OLR 监测预测时间演变图。

图 57 为近期演变和未来预测（过去 30 天实时监测与未来 30 天预报）的 RMM 指数空间相图，提供了包含过去 30 天实时监测的和未来 30 天预报的 RMM 指数相限图，为预报员提供连贯的 MJO 演变和预报参考。图中 START 代表监测的 RMM 指数起始点，FCST 代表预报起始点。由于计算预报的 RMM 指数时需减去用观测资料计算的前 120 天平均，因此基于 NCEP1 和 T639 资料计算的预报期间的 RMM 指数略有不同。

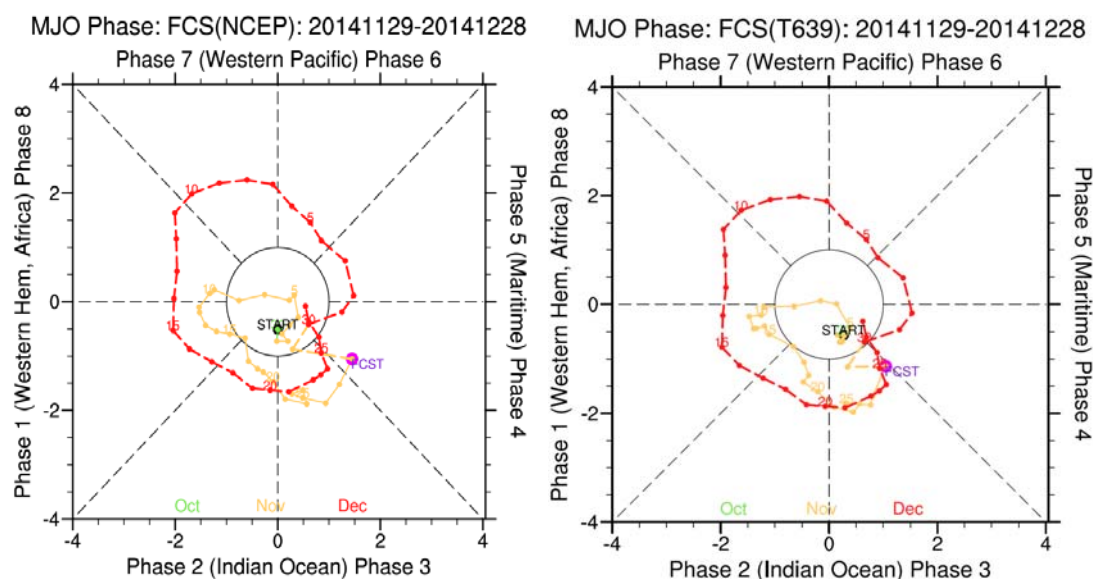


图 57 近期 30d 演变（左边是基于 NCEP1 资料、右边是基于 T639 资料）和未来 30d 预报的 RMM 指数相限图

图 58 为 RMM 指数近期演变及前期起报指数相限图，提供过去 45 天 RMM 指数实时监测（图中的绿色线和黄色线，以区分不同的月）及提前时效为 10 天的 30 天 RMM 指数预测（基于 DERF2.0 模式结果，图中蓝色点线标识），以便对模式预报进行直观检验。这里分别给出基于 NCEP1 和 T639 资料的监测和预报结果。

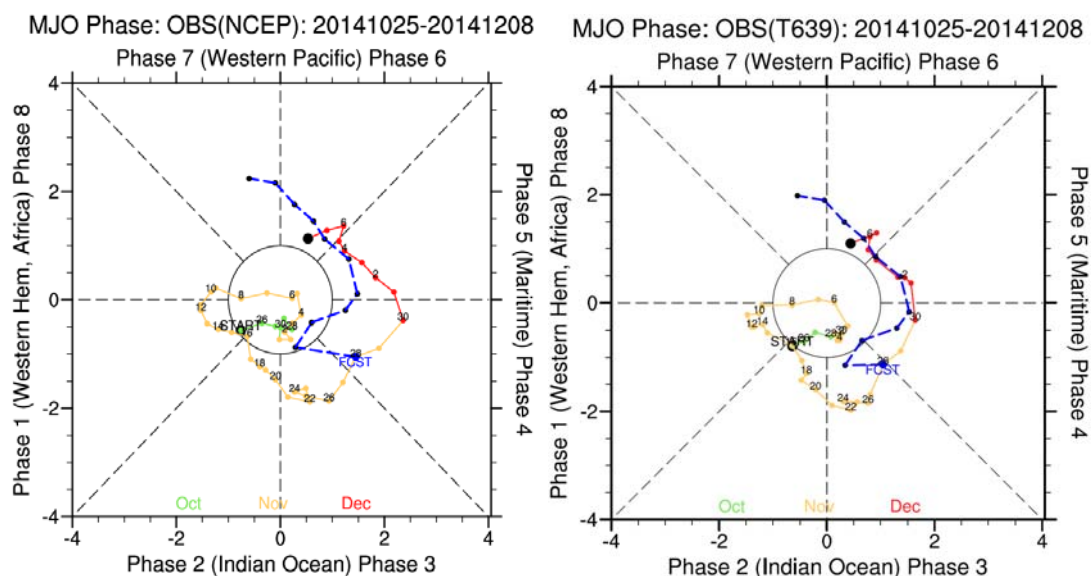
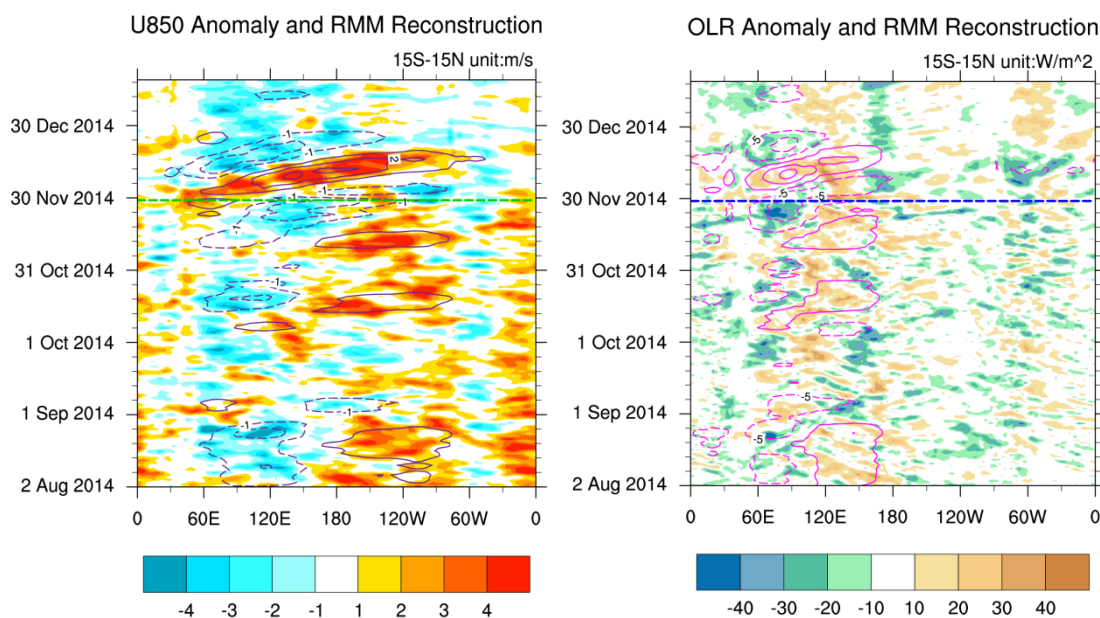


图 58 RMM 指数近期演变及 10 天前起报的指数相限图

左边是基于 NCEP1 资料、右边是基于 T639 资料

图 59 为实时重建的过去 120 天和预报的未来 50 天 U850 和 OLR 异常场的纬向-时间和经向-时间剖面图。在指数监测预测基础上，提供了与 MJO 相关联的纬向风和 OLR 的异常演变图，实时重建的过去 120 天和预报的未来 50 天 U850 和 OLR 异常场的纬向-时间和经向-时间剖面图，为预报员提供实时 MJO 现象时空演变和预报信息以及对我国东部的可能影响。



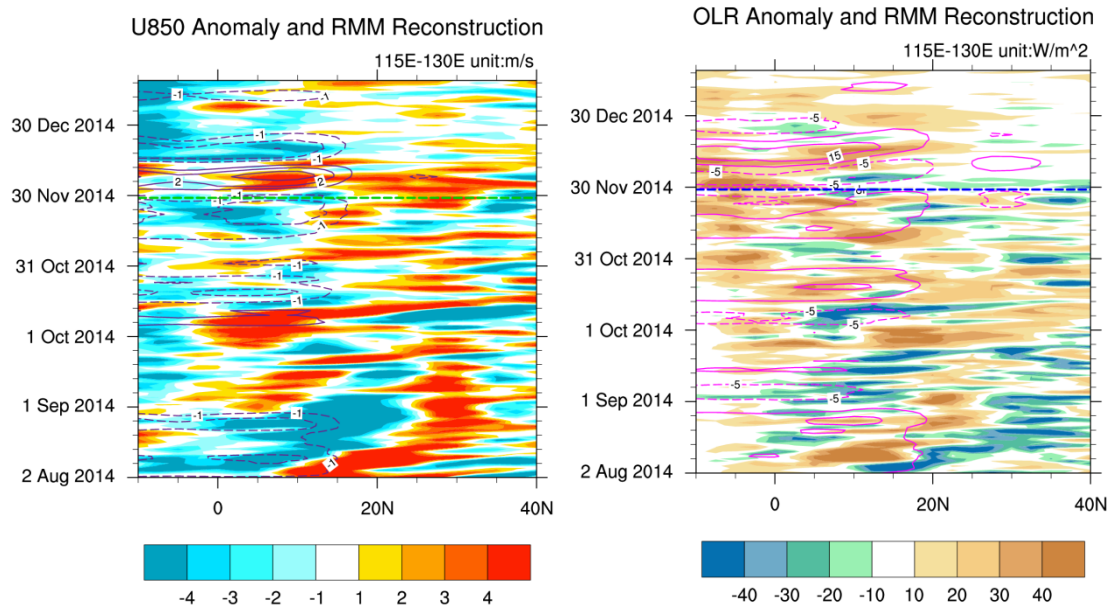
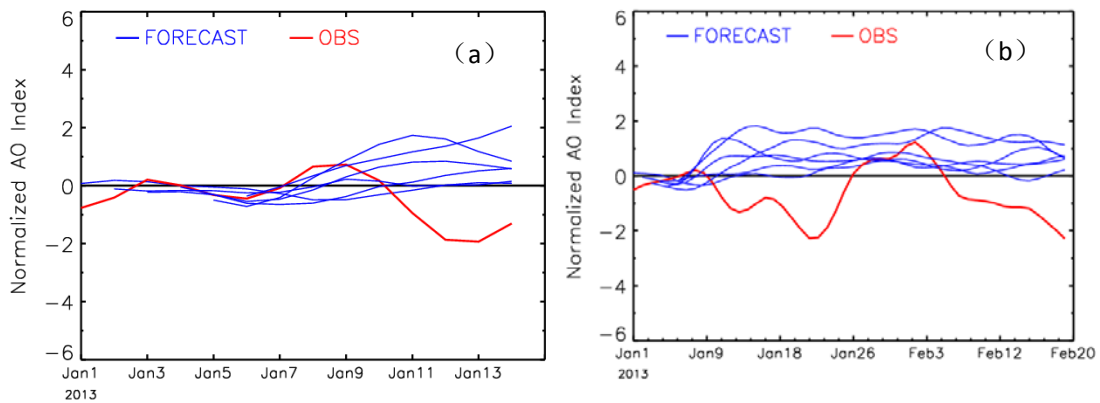


图 59 实时重建的过去 120 天和预报的未来 50 天 U850 和 OLR 异常场

上图是纬向-时间剖面；下图是经向-时间剖面图。其中，等值线和阴影分别代表与 MJO 相关联的和原始的异常场，粗虚线代表预报初始时刻。这里是基于 NCEP 近期演变资料，基于 T639 资料的结果差别微小（图未给出）

6.2 AO 预测实时产品和实时检验

同样的，利用 DERF2.0 实时输出产品以及实时获取的全球分析和再分析资料，实现了 AO 指数的实时监测和预报，图 60 给出了一个 2013 年 1 月的预报个例和与观测的对比。从 3.3 节对 500hPa 位势高度场评估结果来看，DERF2.0 对低纬度环流的预报能力远高于高纬度地区，因此模式对 AO 指数的预报技巧较低，预报时效通常不超过 10 天，但滑动平均的结果能够在一定程度上反映出 AO 指数的后期震荡趋势。



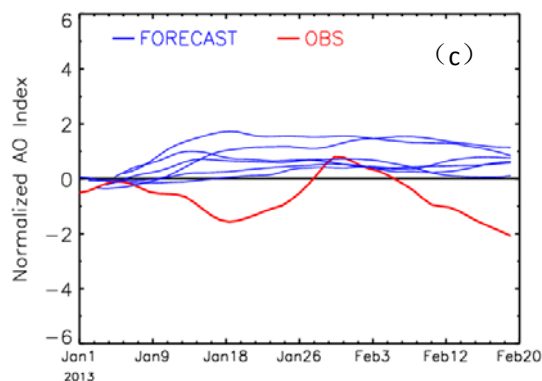


图 60 一个 DERF2.0 对逐日 AO 指数的预报个例及与观测的对比，其中（a）为原始逐日指数，（b）为 5 天滑动平均，（c）为 11 天滑动平均

7 全文小结

本次报告依据 WMO 推荐中国气象局业务标准，分别使用确定性预报评估方法，如空间相似系数（ACC）、时间相关系数（TCC）、平均误差方差（MSSS）等定量评估以及 Ps 评分、Pg 检验等定性评估方法，和概率预报评估方法 ROC，对 DERF2.0 月尺度的气温(t2m)、降水和 500hPa 位势高度和高低层纬向风场的预报性能进行检验评估，按照业务处时间节点安排，本次报告主要提供第 1 季度评估结果，与上次报告相比，主要增加了基于 DERF2.0 模式信息的实时 MJO 监测预测产品评估检验，主要得到以下结论：

1) DERF2.0 第 1 季度多年的气候态在某些地区存在系统性的漂移，如在欧亚大陆中高纬度东部地区气温偏高，我国大部分地区气温偏低，热带西太平洋暖池和 SPCZ 地区降水偏强。

2) 对要素确定性预报技巧评估表现出很多通性，如对 500hPa 高度场预报技巧最高，纬向风场次之，2m 气温场的技巧略低于环流但高于降水；除气温外，模式对热带地区的预报性能好于热带外，特别对降水而言；非等级的定量评估（ACC、TCC、MSSS）等表明，模式的预报技巧对初值非常敏感，随着 lead time 的延长模式预报性能迅速下降，但等级评估（同号率、Ps 评分等）则对初值的敏感性较弱。由于预报员进行短期气候预测时，往往不能及时获取最新初值的预报结果，因此在参考模式较长 lead time 的定量预报结果时需多加注意。

3) 具体到东亚地区而言，模式对东亚地区的位势高度场预报优于其他北半球热带外地区，但对降水和温度的预报技巧要低于其他地区，同时预报性能的年际变化很大，对我国东部的降水和气温预报技巧较高，对我国西部预报技巧明显偏低。

4) 概率预报检验 ROC 的结果表明，模式对高/低异常值出现概率的预报技巧高于常值的预报技巧，对气温场（T2m）在热带外地区的预报技巧相对较高，对我国中东部地区降水异常的预报技巧也高于其他地区，另外模式在热带地区对 500hPa 高度场的概率预报技巧最高，对赤道中太平洋的纬向风场的预报技巧也高于其他地区。模式较高的概率预报技巧表明模式具有相对较多的样本能够准确判断出各个变量场的高/低值事件出现与否，可为模式预报结果的校正提供参考。

5) 对延伸期降水预报技巧检验表明，模式对赤道中东太平洋降水的预报技巧较高，可持续 7-8 候，而对热带外预报技巧较低，预报技巧随时间的延长很快减小，在第 3 候时，模式对我国长江以南地区的预报技巧相对较高，对我国西部的预报技巧很低。从概率预报上看，

模式对 AN/BN 事件的预报技巧类似，均高于对 NN 事件的预报技巧。

本次报告除了对第 1 季度的各个要素月尺度预报结果进行检验外，也初步探讨了基于 DERF2.0 模式信息的实时 MJO 监测预测产品评估检验，这部分工作还需要继续深入下去。同时，我们希望在下次报告中，增加对东亚夏季风、AO、NAO 等的检验评估，同时考虑利用 5 分法，检验模式对极端异常的预报能力。

主要参考文献

- WMO. 1996. The eleventh WMO working group session, Italy.
- WMO. 1999. Abridged final report with resolutions and recommendations. Secretariat of the WMO, Geneva, Switzerland, No.896: 154-157.
- WMO. 2006. Standardised Verification System (SVS) for Long-Range Forecasts (LRF): New attachment II-8 to the manual on the GDPFS (WMO-No. 485), Volume I.
- 何慧根, 李巧萍, 吴统文等. 2014. 月动力延伸预测模式业务系统 DERF2.0 对中国气温和降水的预测性能评估. 大气科学. 待刊
- 李焕连, 李清泉, 王兰宁等. 2010. 跨季度汛期气候数值预测试验. 气象科学. 30(6): 745-753
- 李清泉, 丁一汇, 张培群. 2004. 一个全球海—气耦合模式跨季度汛期预测能力的初步检验和评估. 气象学报, 62(6): 740-751.
- 刘一鸣, 丁一汇, 李清泉. 2005. 区域气候模式对中国夏季降水的 10 年回报试验及其评估分析. 应用气象学报. 16 (增刊): 41-47