

DERF2.0 对全球气温、降水和环流的 预测性能评估

(第 3 季度)

国家气候中心气候研究开放实验室

吴捷 赵崇博 任宏利

2014 年 6 月 20 日

摘要

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（ACC，TCC，MSSS，Ps 评分等）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 3 季度评估结果，与上次报告相比，主要增加了概率预报 ROC 评估和对重要低频过程 MJO/BSISO 的预报技巧评估，主要得到以下结论：

1) DERF2.0 第 3 季度多年气候态在某些地区存在系统性漂移，如在欧亚大陆高纬度地区气温偏低，中纬度气温偏高，对应中纬度地区位势高度场偏高，高空副热带西风急流偏弱，而低层南亚西风偏弱，对应北印度洋和东亚副热带到中纬度地区降水偏弱，热带西太平洋降水偏强。

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；第 3 季度的预报技巧整体低于第 2 季度；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由于其他北半球热带外地区，特别对西太副高所在区域的环流预报技巧较高，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，ENSO 年较高，对我国东北和西南地区的温度和降水预报技巧略高。

4) 概率预报检验 ROC 的结果表明，模式对高/低异常出现时预报的技巧高于正常时的预报技巧，对气温场（T2m）在热带外地区的预报技巧较高，对我国东部地区降水异常的预报技巧也高于其他地区，即有相对较多的样本能够准确判断出气温高/低/常值事件出现与否，可为模式预报结果的校正提供参考。

5) 对 MJO 预报技巧检验表明，模式对 RMM 指数的预报技巧约为 14 天，与其他模式不同的是，DERF2.0 对秋夏 MJO 预报技巧较高而对春冬预报技巧较低；对两个 BSISO 指数预报技巧分别为 10 天和 8 天。

本次报告除了对第 3 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 DERF2.0 对季节内重要过程 MJO/BSISO 的预报技巧，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，增加对重要的大气环流系统，如东亚夏季风、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。

目录

1 概述	1
2 资料和方法（补充说明）	1
2.1 新增检验要素.....	1
2.2 新增检验方法.....	1
3 月尺度评估结果	3
3.1 2m 气温（t2m）.....	3
3.2 降水（prec）.....	11
3.3 500hPa 位势高度（HGT 500）.....	19
3.4 200hPa 和 700hPa 纬向风场.....	23
4 概率预报评估（ROC）	28
4.1 2m 气温（t2m）.....	28
4.2 降水（prec）.....	29
4.3 500hPa 位势高度.....	31
4.4 纬向风场.....	32
4.5 ROC 评估小结.....	33
5 重要季节内振荡现象评估	34
5.1 MJO.....	34
5.2 北半球夏季季节内振荡（BSISO）.....	36
5 全文小结	37
主要参考文献	38

DERF2.0 对全球气温、降水和环流的预测性能评估

(第 3 季度)

1 概述

根据科技业务处工作安排和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 3 季度 (7、8、9 月) 的评估结果，与第 2 季度报告相比，增加了概率预报检验 ROC 和对 MJO/BSISO 指数的技巧评估，以供参考。

2 资料和方法 (补充说明)

对 DERF2.0 回报资料的处理方式和评估区域与第 2 报告中所述基本一致，在此不再赘述，仅补充增加的检验要素 (纬向风场) 和检验方法，包括趋势异常综合检验 (Ps 评分)，Relative Operating Characteristics (ROC)、RMM 指数技巧评分 (Wheeler and Hendon, 2004) 和北半球夏季季节内振荡 (BSISO) 指数技巧评分等 (Lee, et al, 2013)

2.1 新增检验要素

本次报告检验的要素除降水 (prec)、2m 气温 (t2m)，位势高度场外 (z)，增加了对 200hPa 和 700hPa 纬向风的检验，以评估模式对高空急流异常和低空主要环流系统和水汽输送带的预报能力。

2.2 新增检验方法

本次报告保留之前的空间异常相似系数 (ACC)，距平一致率 (Pc)，分级评分 (Pg)、时间异常相关系数 (TCC) 和均方根技巧评分等 (MSSS)，在此不再赘述，仅补充新增的检验方法。

2.2.1 趋势异常综合检验 (Ps 评分)

月、季气候趋势预测采用六分类描述。Ps 评分针对确定性的分类预报进行评估，具体方法参考 <http://10.1.64.147/web/index.php?ChannelID=1>，使用的公式如下：

$$Ps = \frac{a * N0 + b * N1 + c * N2}{(N - N0) + a * N0 + b * N1 + c * N2 + M} * 100$$

a、b 和 c 分别为气候趋势项 (N0)、一级异常项 (N1) 和二级异常项 (N2) 的权重系数，分别取 a=2，b=2，c=4，漏报记为 M。

2.2.2 Relative Operating Characteristics(ROC)

ROC 用来衡量预报中区分两种可能性结果的能力。它对预报的偏差并不敏感，即使是有偏差的预报也可能得到较好的 ROC 曲线，因此，可对预报结果的校正提供参考。ROC 是基于表 1 中报

对和误报个数统计进行确定性预报或概率预报。这里事件被定义为二进制，即只有两种可能的结果，发生或不发生。

表1 用于定义事件概率预报参数的ROC列表

区间数	预测概率	观测发生	观测不发生
1	0-P ₂ %	O ₁	NO ₁
2	P ₂ -P ₃ %	O ₂	NO ₂
3	P ₃ -P ₄ %	O ₃	NO ₃
...
n	P _n -P _{n+1} %	O _n	NO _n
...
N	P _N -100%	O _N	NO _N

其中，n 用来定义概率区间，P_n 第n个区间的概率下限，P_{n+1}第n个区间的概率上限，N 为区间数。

$$O_n = \sum W_i (O)_i$$

$$NO_n = \sum W_i (NO)_i$$

$$W_i = \cos(\theta)_i$$

θ 为格点纬度

$$HR_n = \frac{\sum_{i=n}^N O_i}{\sum_{i=1}^N O_i}$$

$$FAR_n = \frac{\sum_{i=n}^N NO_i}{\sum_{i=1}^N NO_i}$$

其中，HR为报对率，FAR为误报率。

ROC曲线的横坐标为误报率，纵坐标为报对率，ROC曲线下的面积（ROCA）常用作代表预报技巧的统计指数，其值介于0-1之间，对角线（ROCA=0.5）代表无预测技巧。将各个变量分为高值、常值和低值三类，给定总区间为10，对超前0-4候的模式预报结果进行检验。

2.2.3 MJO-RMM 指数

按照（Wheeler and Hendon, 2004）方法提取 MJO 的主要模态（参见 5.1 小节），并计算得到两个 RMM 指数，计算的相关技巧（COR）、均方根误差（RMSE）和均方根技巧评分的公式如下（Lin, etal, 2009）：

$$COR(\tau) = \frac{\sum_{i=1}^N [a_{1i}(t)b_{1i}(t) + a_{2i}(t)b_{2i}(t)]}{\sqrt{\sum_{i=1}^N [a_{1i}^2(t) + a_{2i}^2(t)]} \sqrt{\sum_{i=1}^N [b_{1i}^2(t) + b_{2i}^2(t)']}}$$

$$RMSE(\tau) = \sqrt{\frac{1}{N} \sum_{i=1}^N \{ [a_{1i}(t) - b_{1i}(t)]^2 + [a_{2i}(t) - b_{2i}(t)]^2 \}}$$

$$MSE_f(\tau) = \frac{1}{N} \sum_{i=1}^N \{ [a_{1i}(t) - b_{1i}(t)]^2 + [a_{2i}(t) - b_{2i}(t)]^2 \}$$

$$MSE_c(\tau) = \frac{1}{N} \sum_{i=1}^N [a_{1i}^2(t) + a_{2i}^2(t)]^2$$

$$MSSS(\tau) = [1 - \frac{MSE_f(\tau)}{MSE_c}]$$

其中 $a_{1i}(t)$ 和 $a_{2i}(t)$ 代表观测的 RMM1 和 RMM2 指数， $b_{1i}(t)$ 和 $b_{2i}(t)$ 代表预测。

2.2.4 BSISO 指数

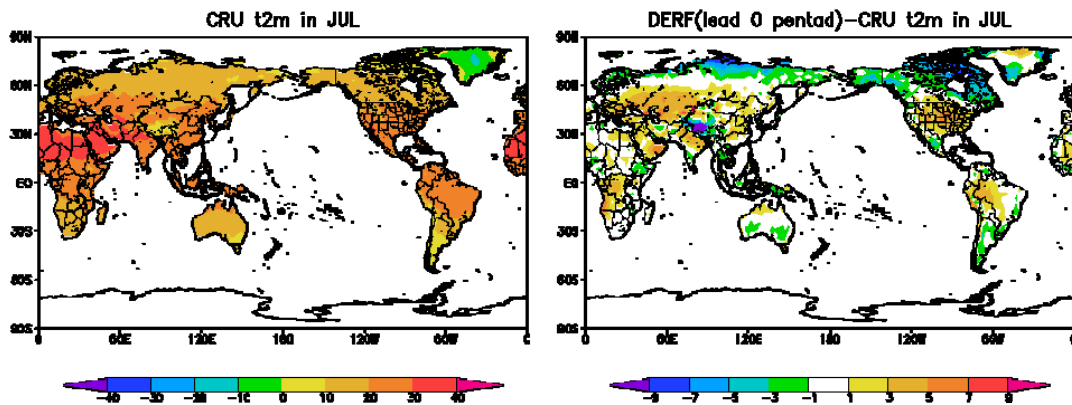
按照 (Lee et al, 2004) 方法提取 BSISO 的主要模态 (参见 5.2 小节), 并按计算得到四个 BSISO 指数, 上节所述公式分别计算的相关技巧 (COR)、均方根误差 (RMSE) 和均方根技巧评分。

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 3 季度气候态在欧亚大陆高纬度地区以及青藏高原均整体偏低 1°C 以上, 对夏季欧亚大陆中纬度地区 and 我国中东地区模拟的温度略偏高, 这种特征在 9 月有所减弱。



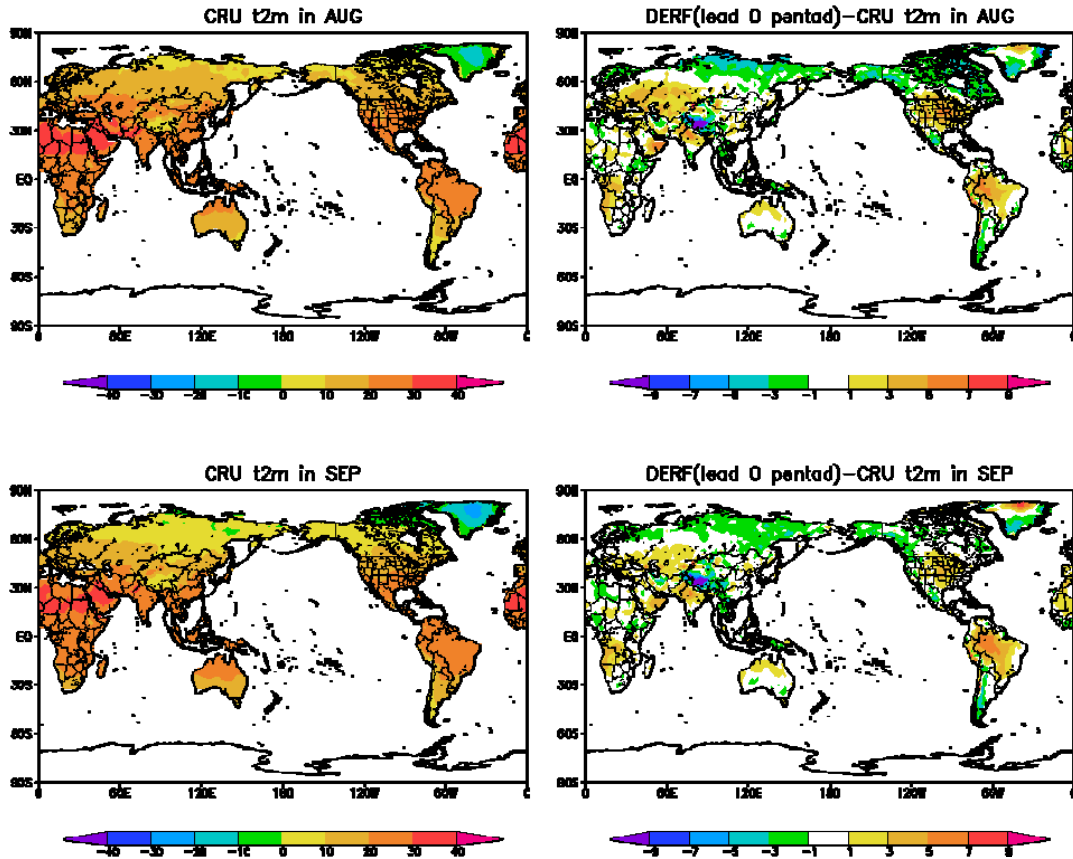
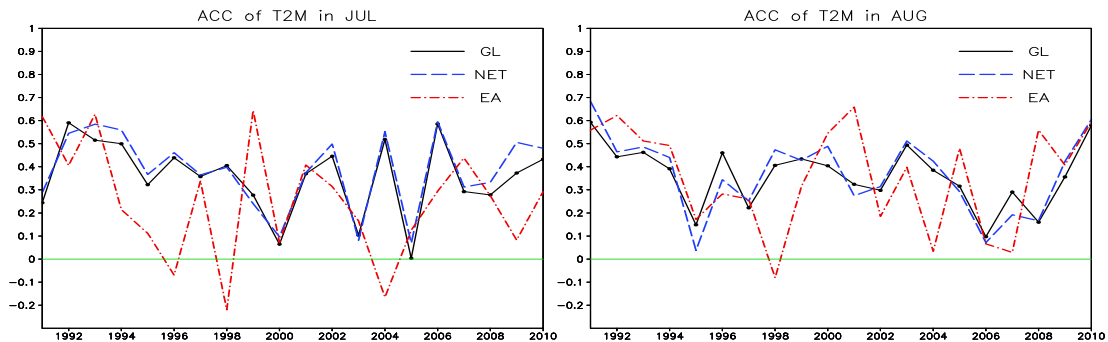


图1 全球多年 (20a) 平均的陆地 7 月、8 月、9 月的 2m 温度 (t2m), 左列为 CRU 观测, 右侧为 DERF (超前 0 候预报) 与观测的差值图

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, ACC 的年际变化很大, 特别是对东亚地区而言, 最高能达到 0.7, 最低则小于 0, 总体而言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 评分均已在 0.2 以下。



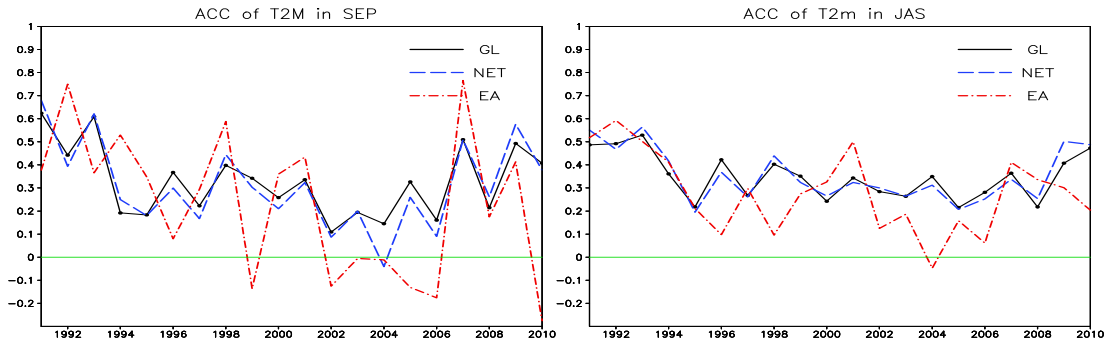


图2 DERF2.0 提前 0 候起报的不同地区平均的（全球（GL）、北半球热带外（NET）、东亚（EA））陆地 7 月、8 月、9 月的 2m 温度（t2m）与观测的空间距平相似系数（ACC）逐年演变

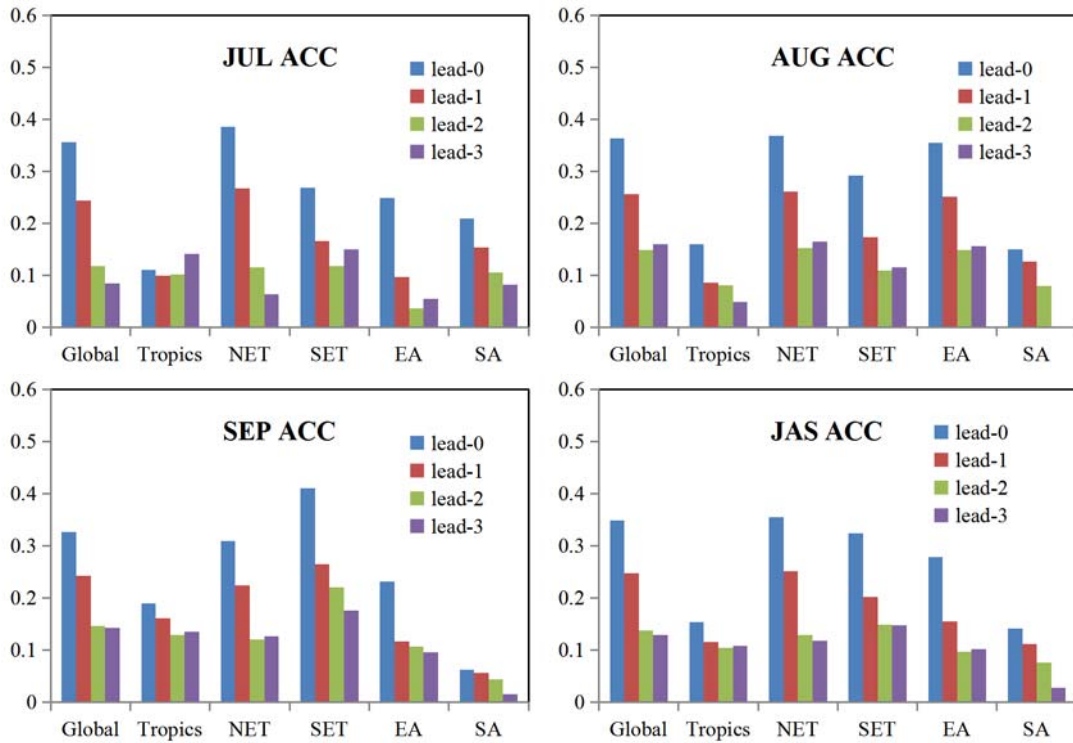


图3 DERF2.0 在不同 lead time 起报的 6 月、7 月、8 月和第 3 季度的 2m 温度（t2m）的空间距平相似系数在不同区域的多年平均，横坐标代表多个区域（见表 3）

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%~60%之间，北半球热带外则能够超过 60%，因此 DERF2.0 的预测总体上能够反映出 t2m 异常的主要趋势。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，这一点对气候预测中的定性预报十分重要。

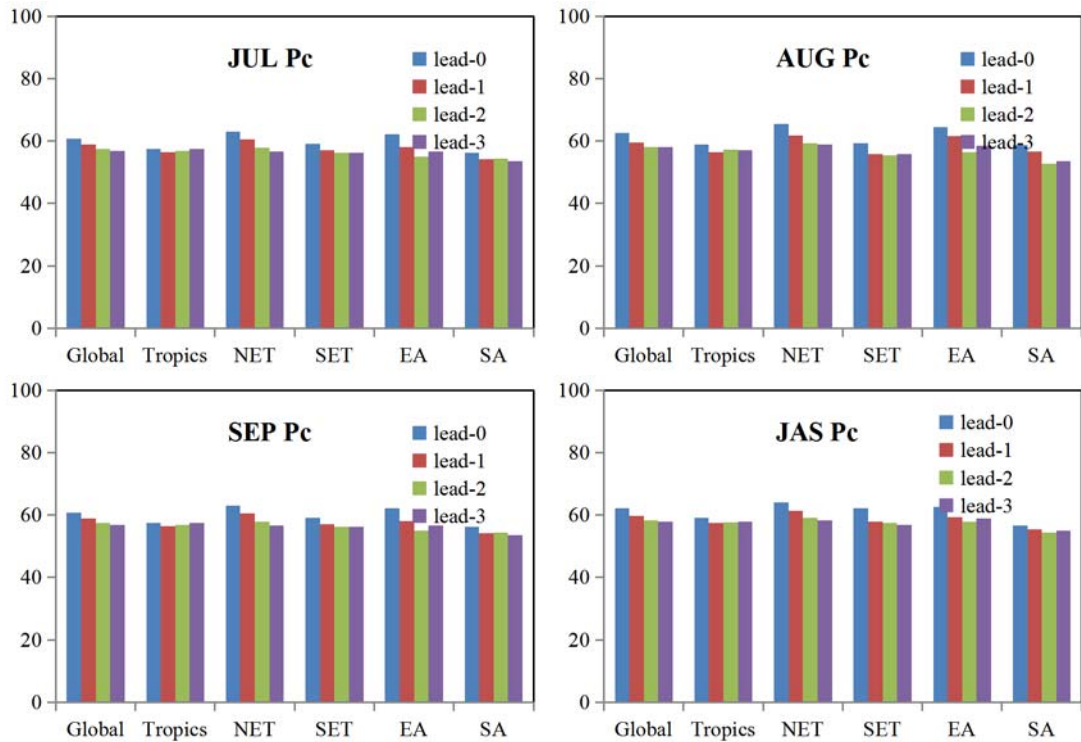


图 4 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的 2m 温度 (t2m) 的距平符号一致率 (Pc) 的在不同地的多年平均

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，总体评分在 70-80 分之间，基本都能高于 60 分，但同样在东亚地区的年际变化非常明显，但 Ps 评分对初值的敏感程度比 ACC 要小很多。

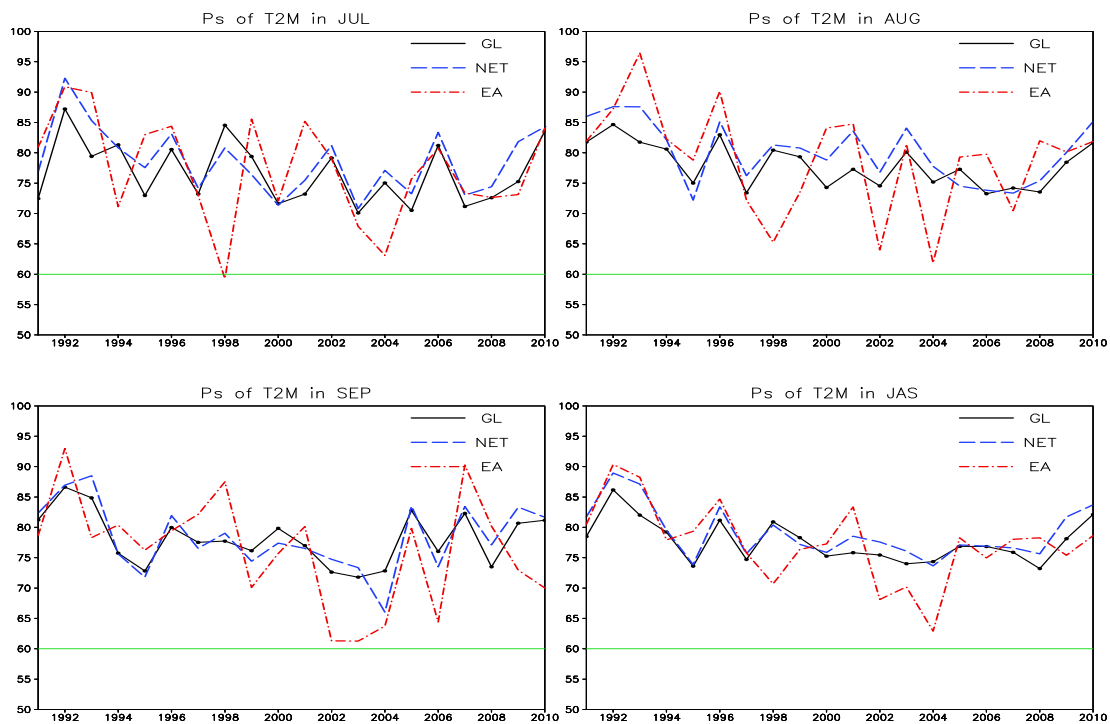


图 5 同图 2，但为趋势异常综合检验 (Ps 评分) 逐年演变

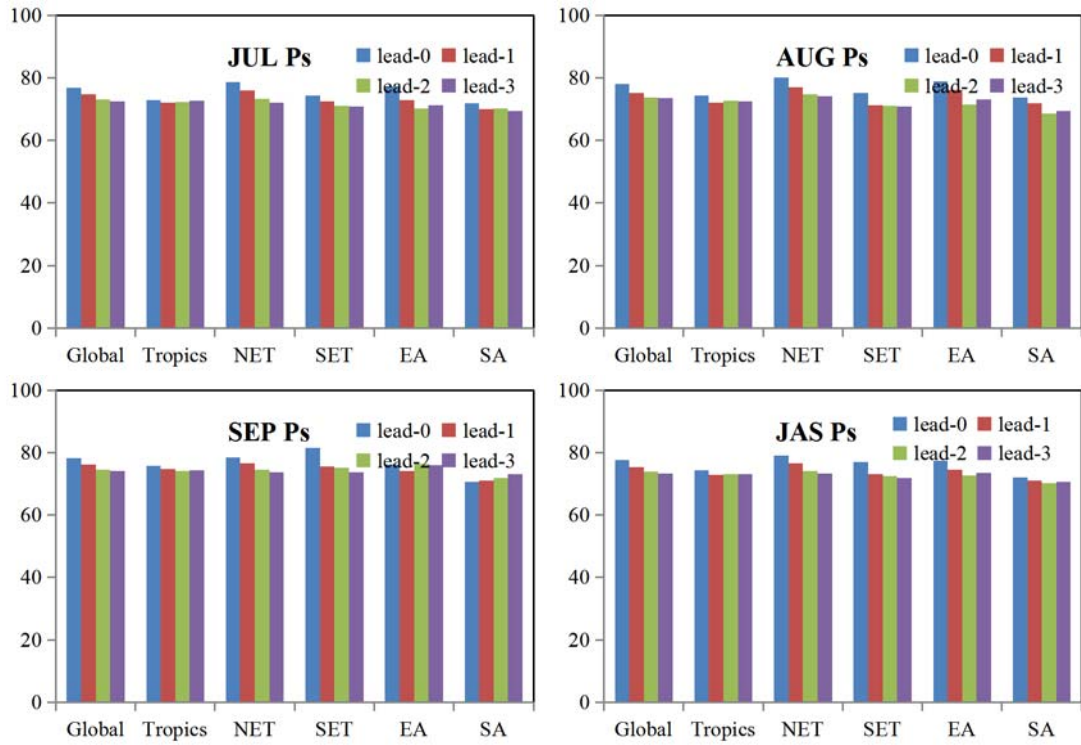


图6 DERF2.0在不同lead time起报的7月、8月、9月和第3季度的2m温度(t2m)的趋势异常综合检验(Ps评分)的在不同地的多年平均

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3季度各月气温预测的Pg评分均在75分附近，热带地区略高，我国西南部到中南半岛评分较高。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。

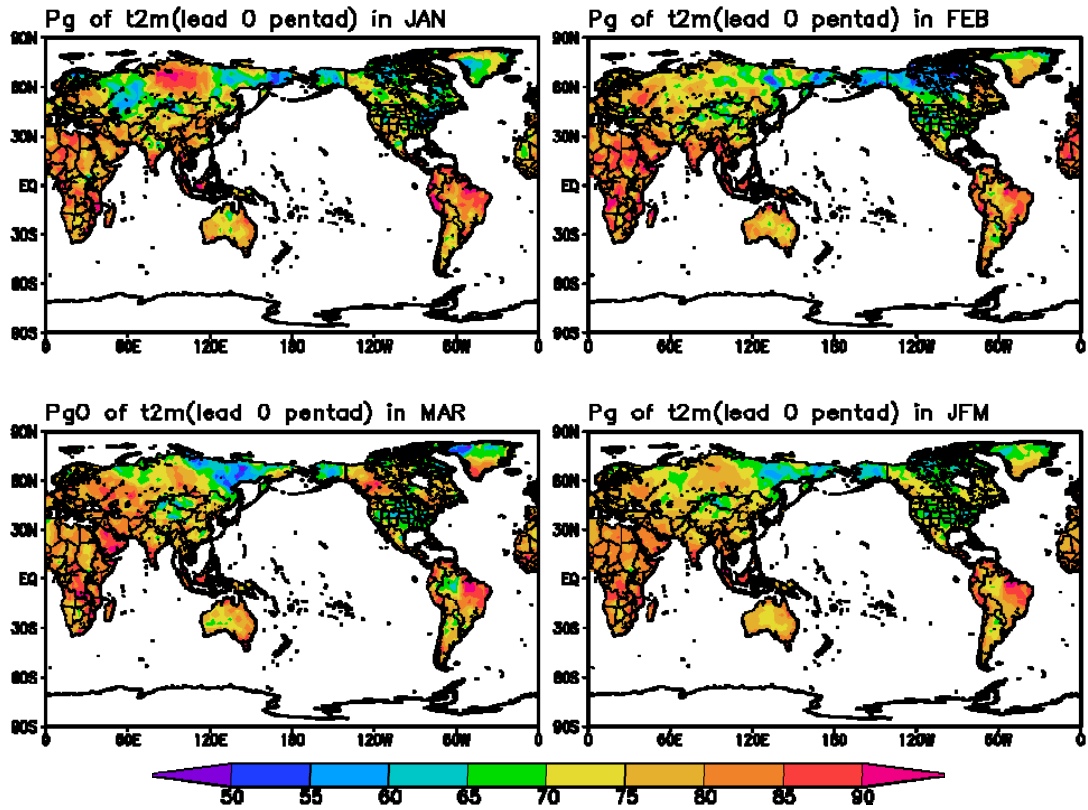


图7 DERF2.0 超前0 候起报的的全球7月、8月、9月和第3季度的2m 温度 (t2m) 的Pg 评分多年平均分布

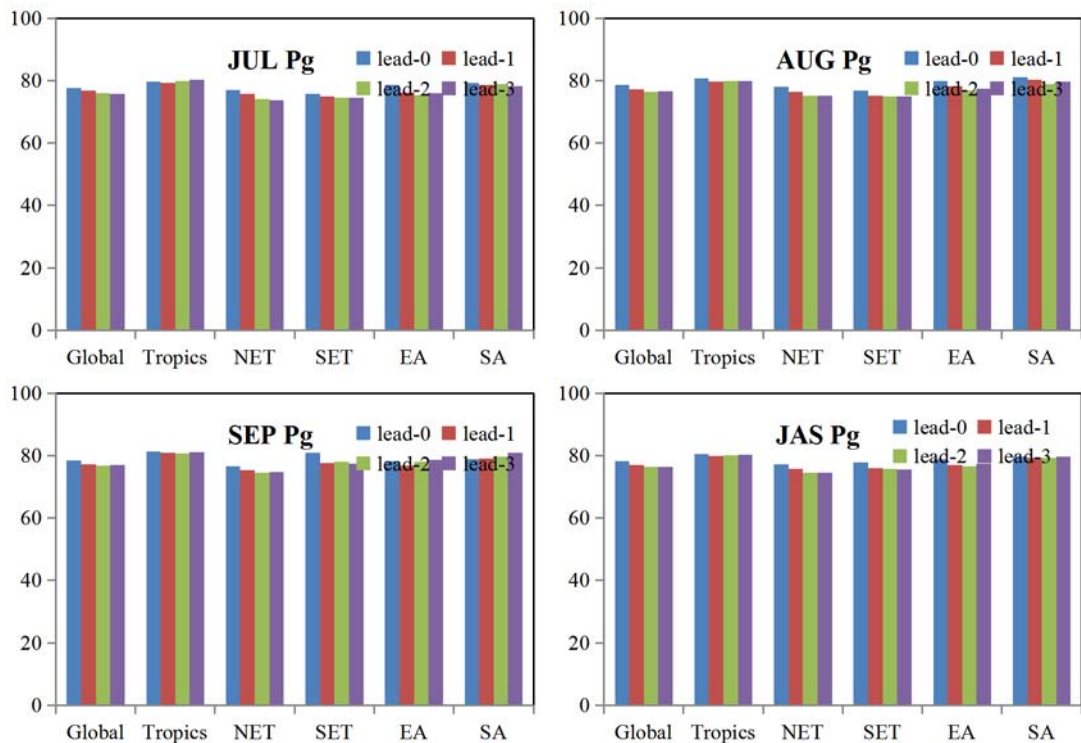


图8 DERF2.0 在不同 lead time 起报的7月、8月、9月和第3季度的2m 温度 (t2m) 的Pg 评分在不同地区的多年平均

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 2 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 2 季度相比第 3 季度的预报技巧均有显著下降，全球平均 TCC 在 0.3 左右；区域平均来看，北半球热带外的预报技巧高于其他地区，如欧亚中高纬度，我国东北地区等。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。

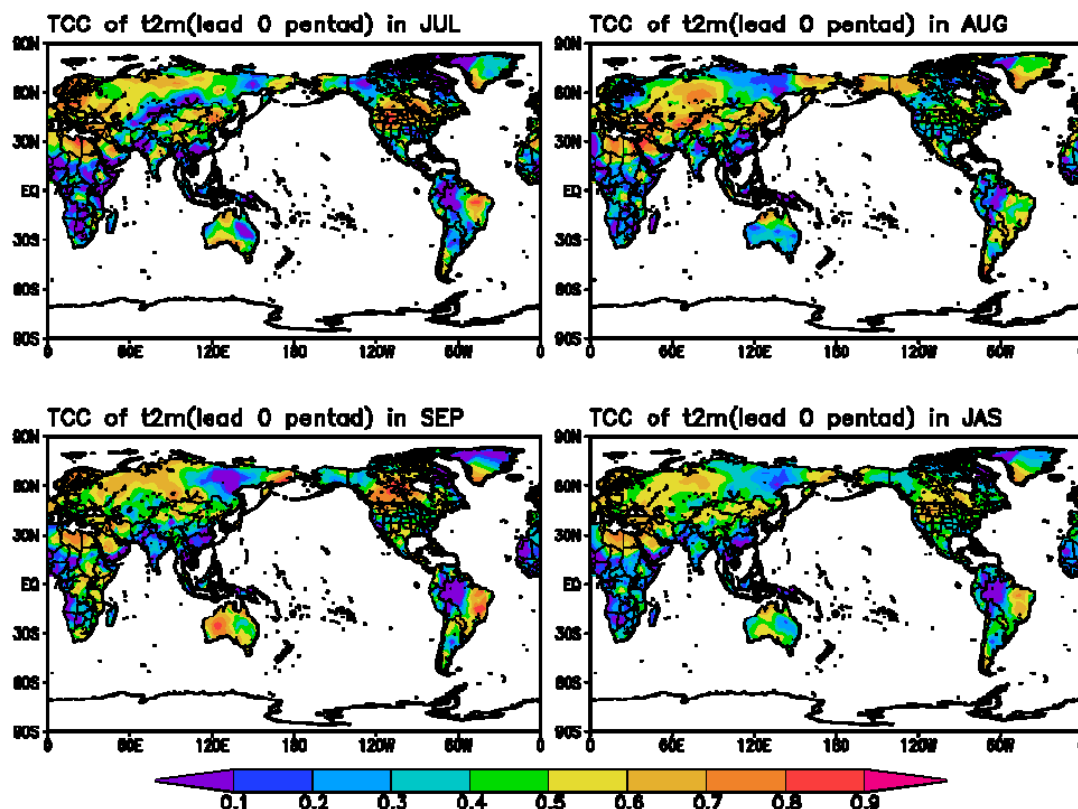
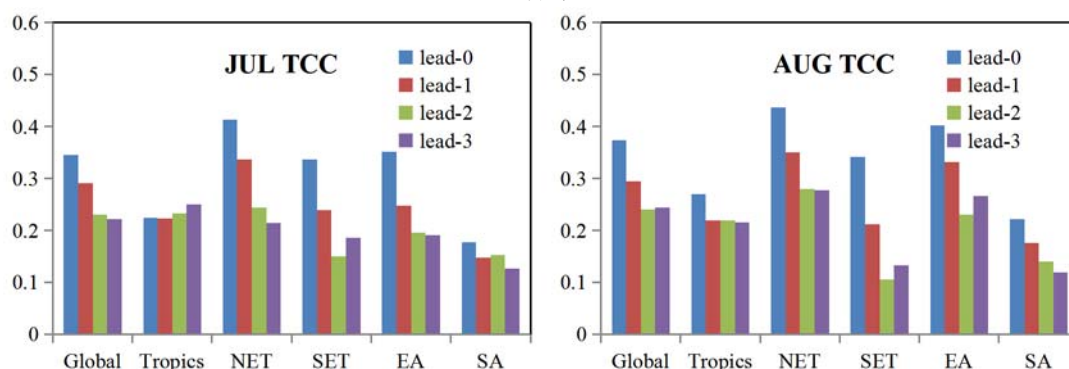


图 9 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的 2m 温度 (t2m) 的时间距平相关系数 (TCC) 分布



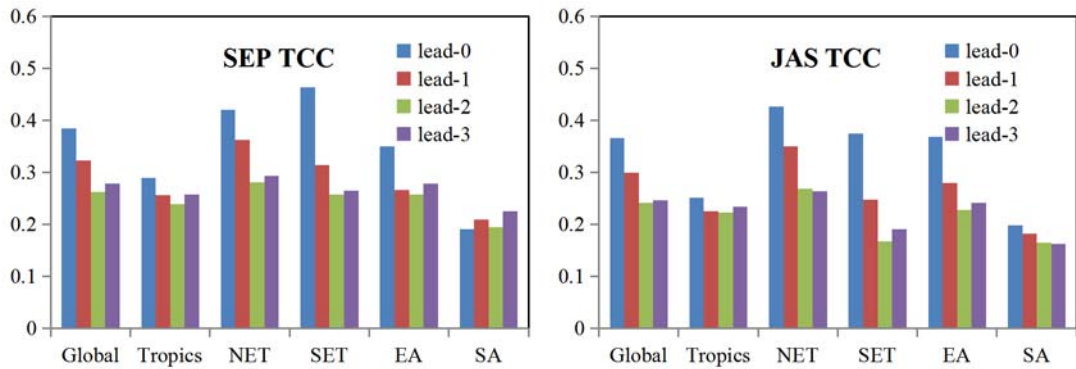


图 10 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的 2m 温度 (t2m) 的时间距相关系数 (TCC) 在不同地区的平均

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 总体得分小于 0, 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, 这主要是由于在某些地区, 模式的方差远大于观测, 造成 MSSS 评分较低, 在求 TCC 时进行了开方, 影响较小。

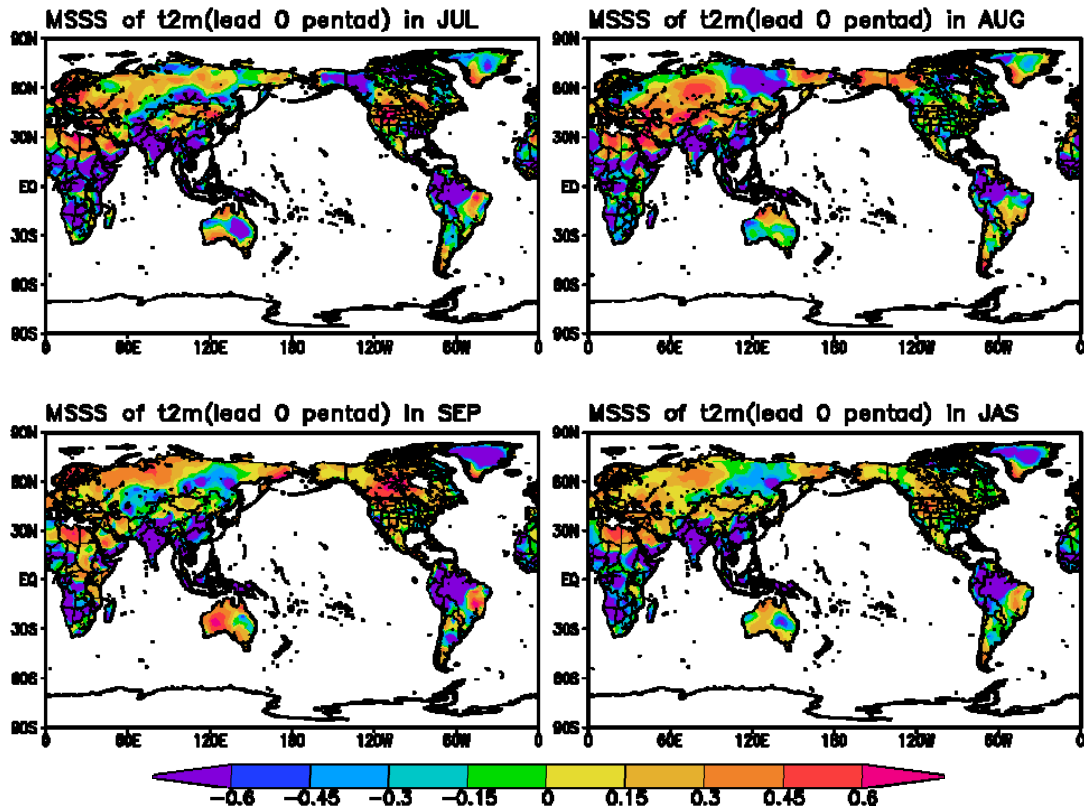


图 11 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的 2m 温度 (t2m) 的平均方差技巧 (MSSS) 分布

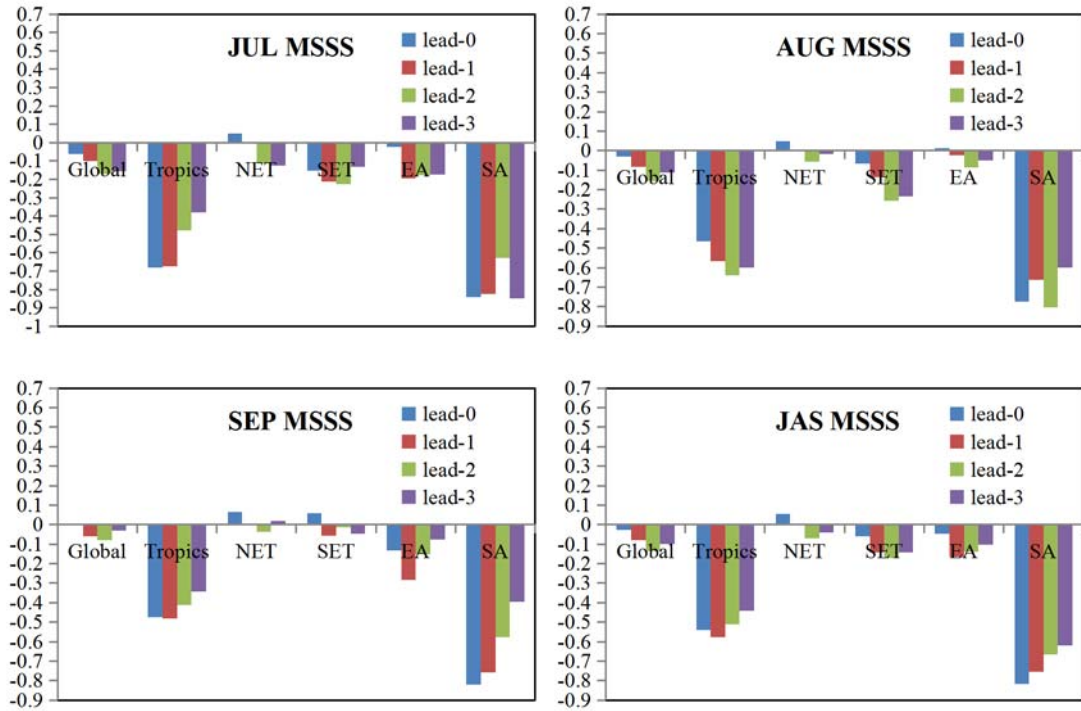


图 12 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的 2m 温度(t2m)的平均方差技巧(MSS)在不同地区的平均

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，表现为高纬和青藏高原偏冷中纬度地区偏暖；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，在热带的 MSS 技巧较低，这可能与评估方法及模式自身的方差偏大有关，对我国而言，模式对我国东北、西南地区预报技巧相对较高；第二，第 3 季度 JAS 三个月的技巧十分相似，与第 2 季度相比技巧有明显的下降；第三，在非等级预报评分中 (ACC、TCC、MSS)，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) 和南太辐合带 (SPCZ) 的主要特征，但热带西太平洋地区、热带西印度洋的降水明显偏强，而北印度洋和中纬度地区的降水则偏弱。

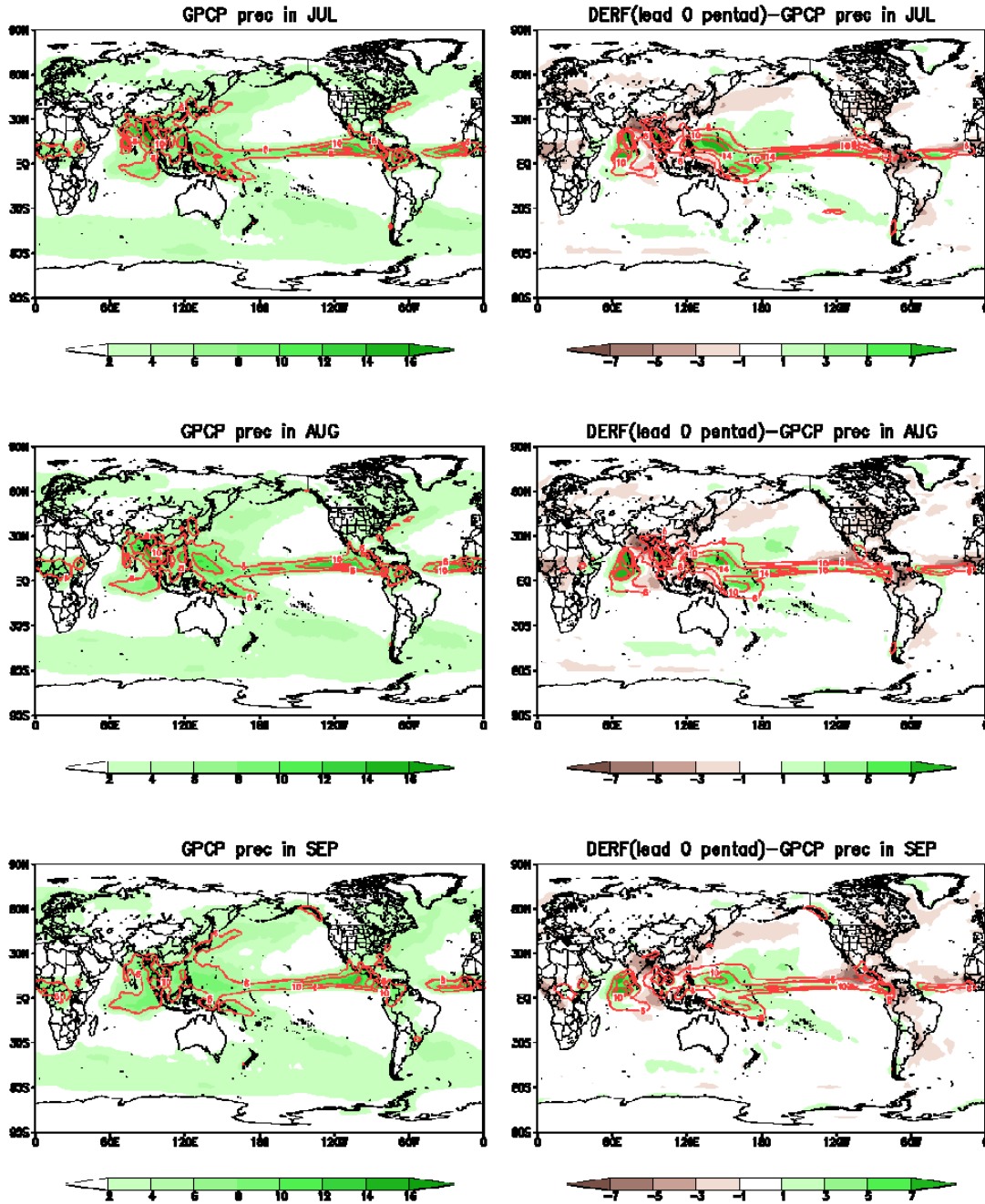


图 13 全球多年 (20a) 平均的 7 月、8 月、9 月降水量 (单位: mm/day), 左列为 GPCP 观测, 右侧阴影为 DERF (超前 0 候) 预报的气候态与 GPCP 观测之差, 红色等值线标出 DERF 降水量大于 6mm/day 的区域

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 0.2 左右, 也低于第 2 季度。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。此外在 ElNino 发展的 1997 年, 降水预报的 ACC 较高。

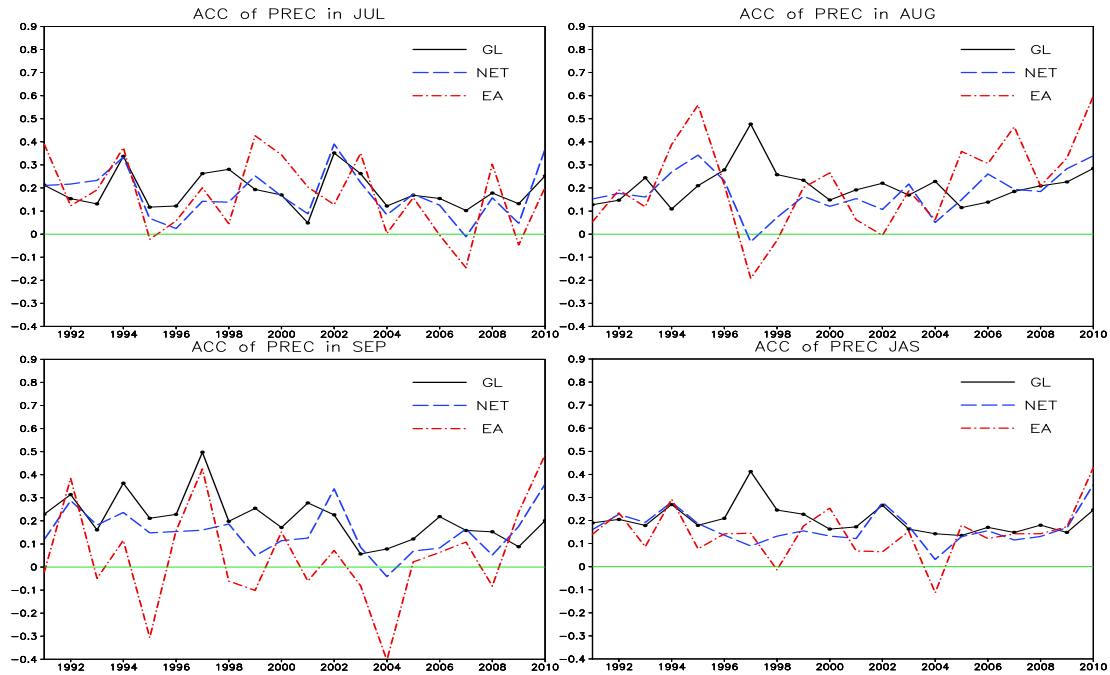


图 14 DERF2.0 提前 0 候起报的不同地区平均的（全球（GL）、北半球热带外（NET）、东亚（EA））陆地 7 月、8 月、9 月的降水（prec）与观测的空间距平相似系数（ACC）逐年演变

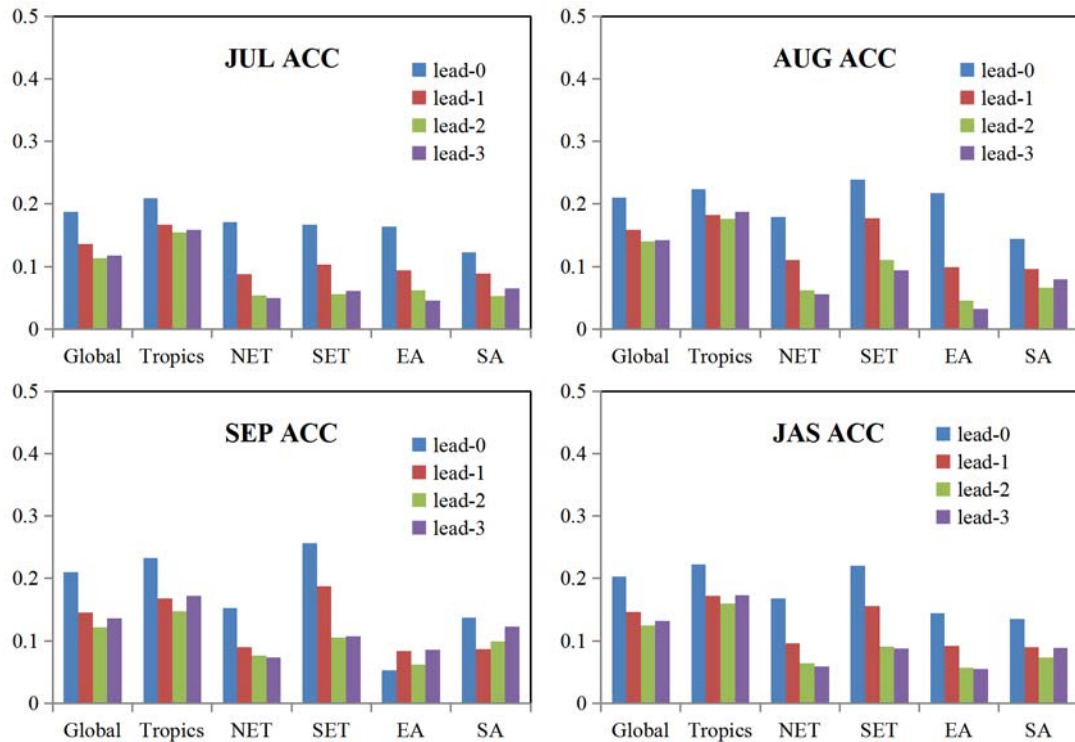


图 15 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的降水（prec）的距平相似系数在不同区域的多年平均，横坐标代表多个区域（见表 3）

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。

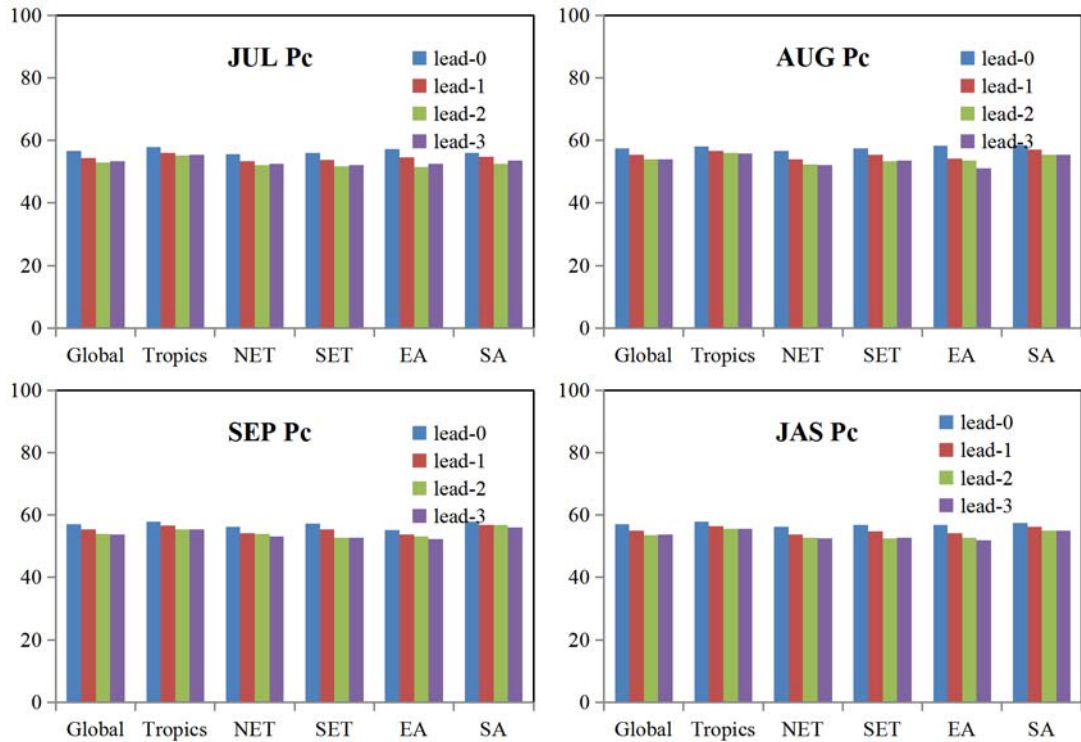


图 16 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的降水 (prec) 的距平符号一致率在 不同地区的多年平均

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 70-75 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。此外，随着 lead time 的延长，Ps 评分下降并不明显。

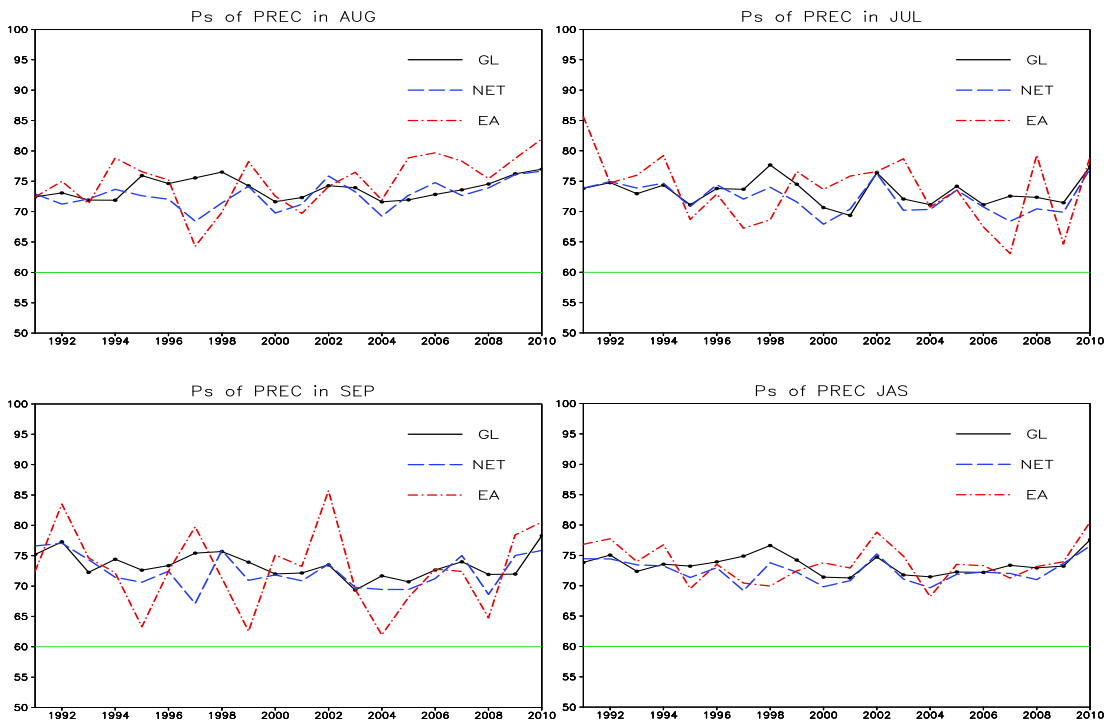


图 17 同图 14，但为趋势异常综合检验 (Ps 评分) 逐年演变

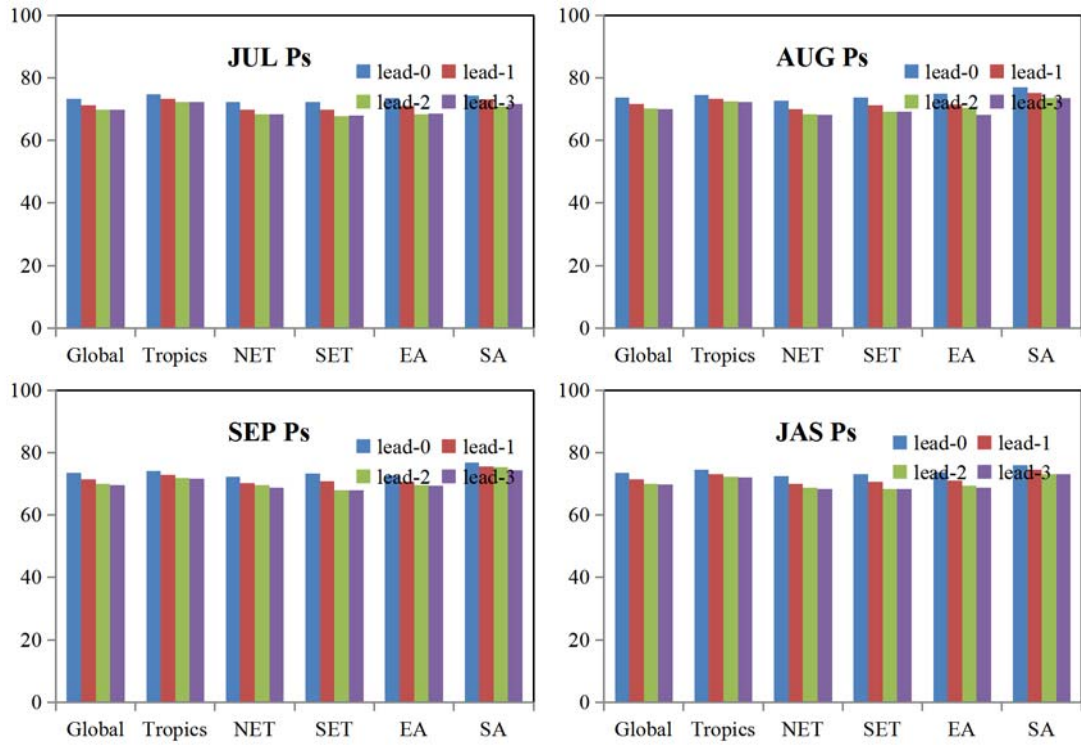


图 18 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的降水 (prec) 的趋势异常综合检验 (Ps 评分) 的在不同地区的多年平均

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。

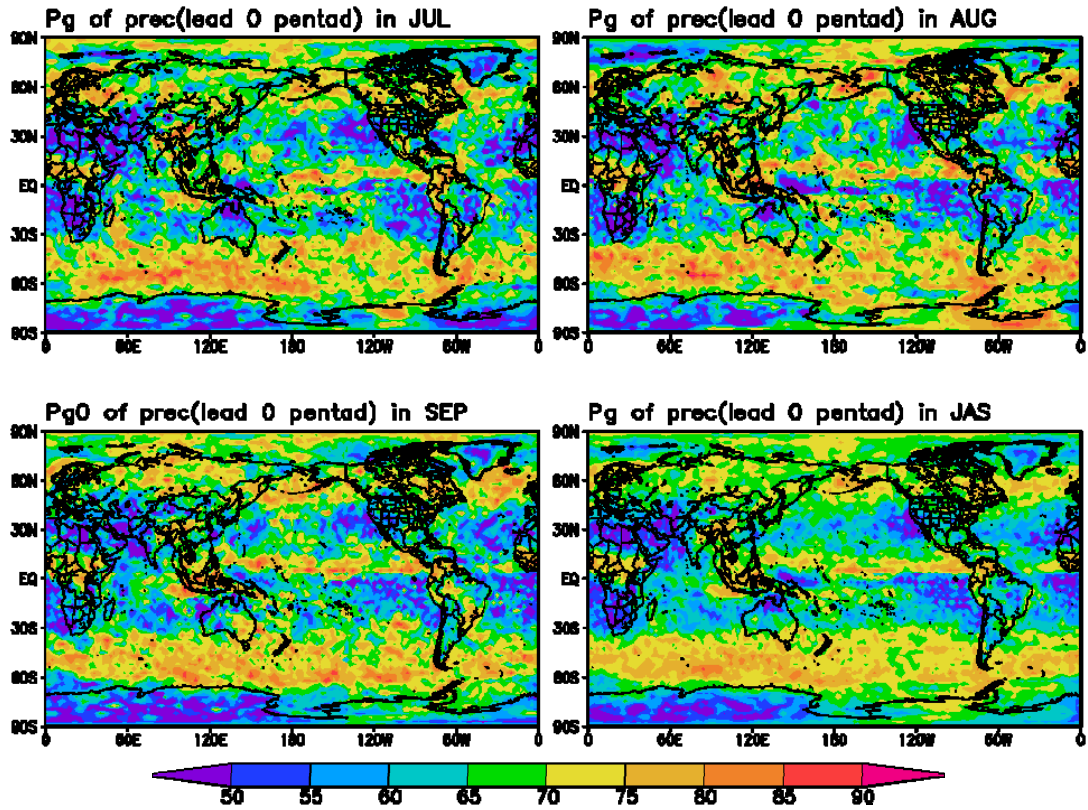


图 19 DERF2.0 超前 0 候起报的的全球 7 月、8 月、9 月和第 3 季度的降水 (prec) 的 Pg 评分多年平均分布

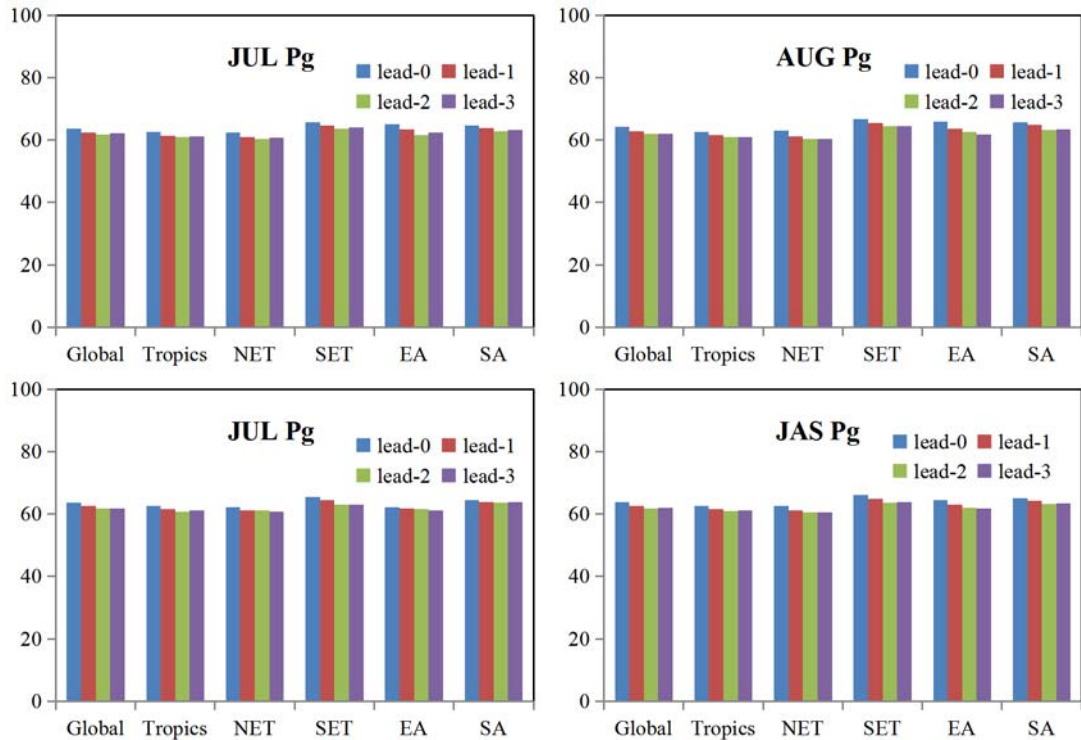


图 20 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的降水 (prec) 的 Pg 评分在不同地区的多年平均

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 小于 0.2，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，而其他对其则对初值十分敏感，对于我国而言，模式同样对西南地区的降水异常有一定预测能力，对西北地区降水预测性能较差。

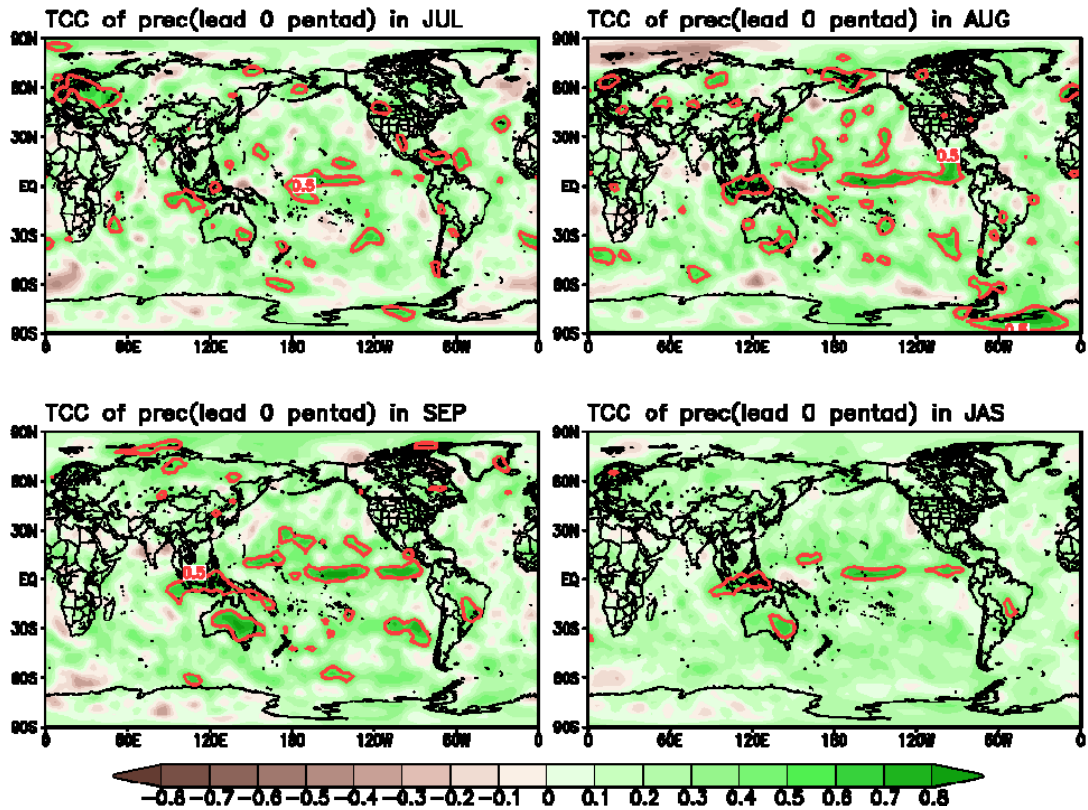
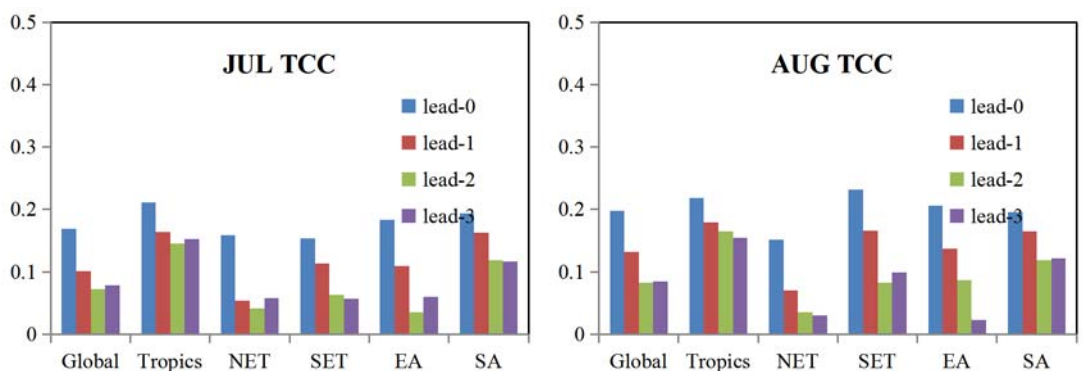


图 21 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的降水 (prec) 的时间距平相关系数 (TCC) 分布



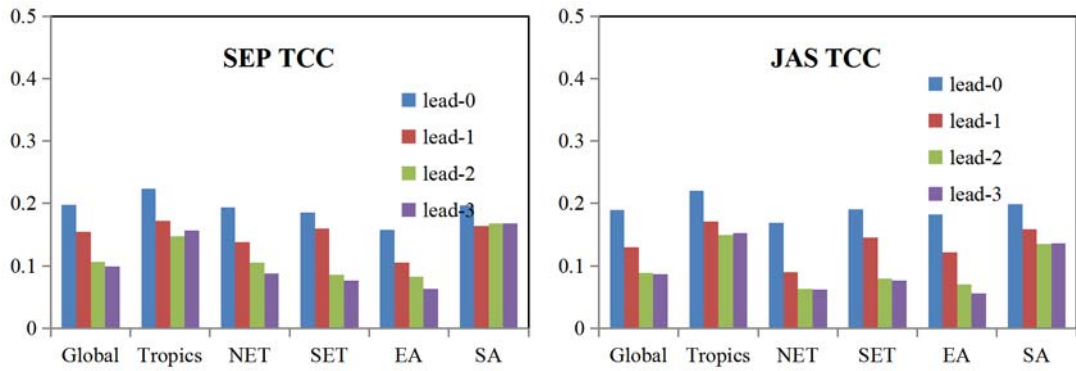


图 22 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的降水 (prec) 的时间距平相关系数 (TCC) 在不同地区的平均

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、我国西南和东北的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的降水性能预报较差。

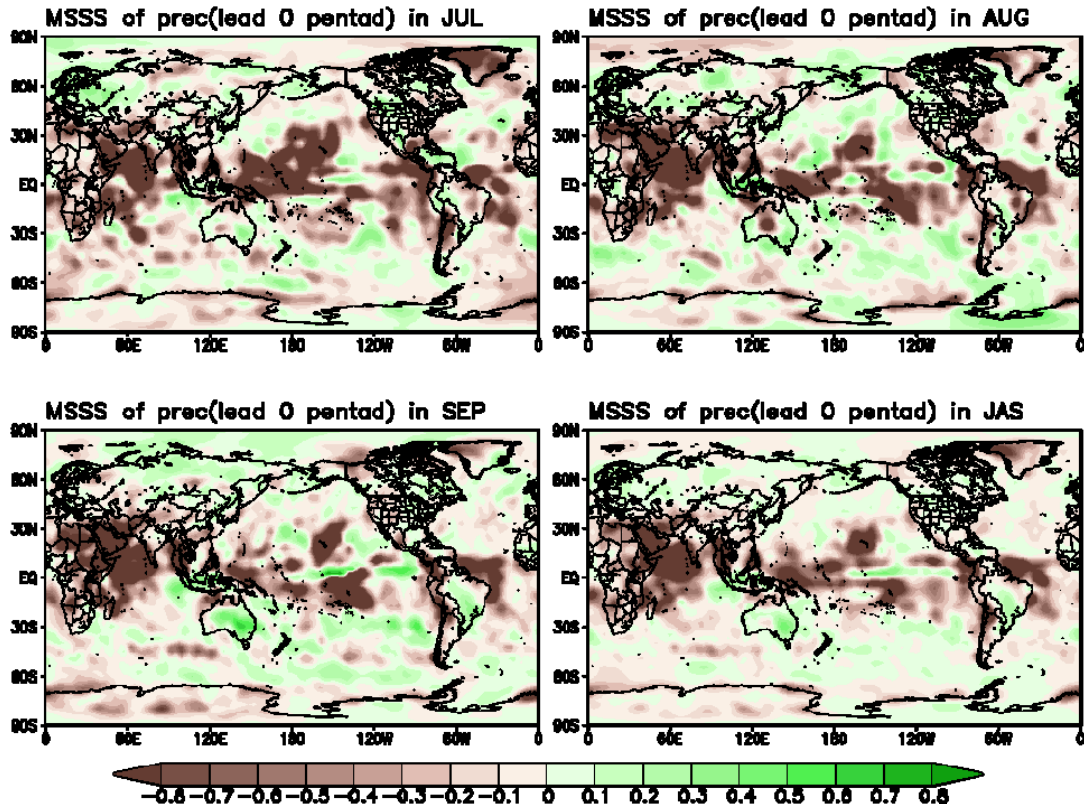


图 23 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的降水 (prec) 的平均方差技巧 (MSSS) 分布

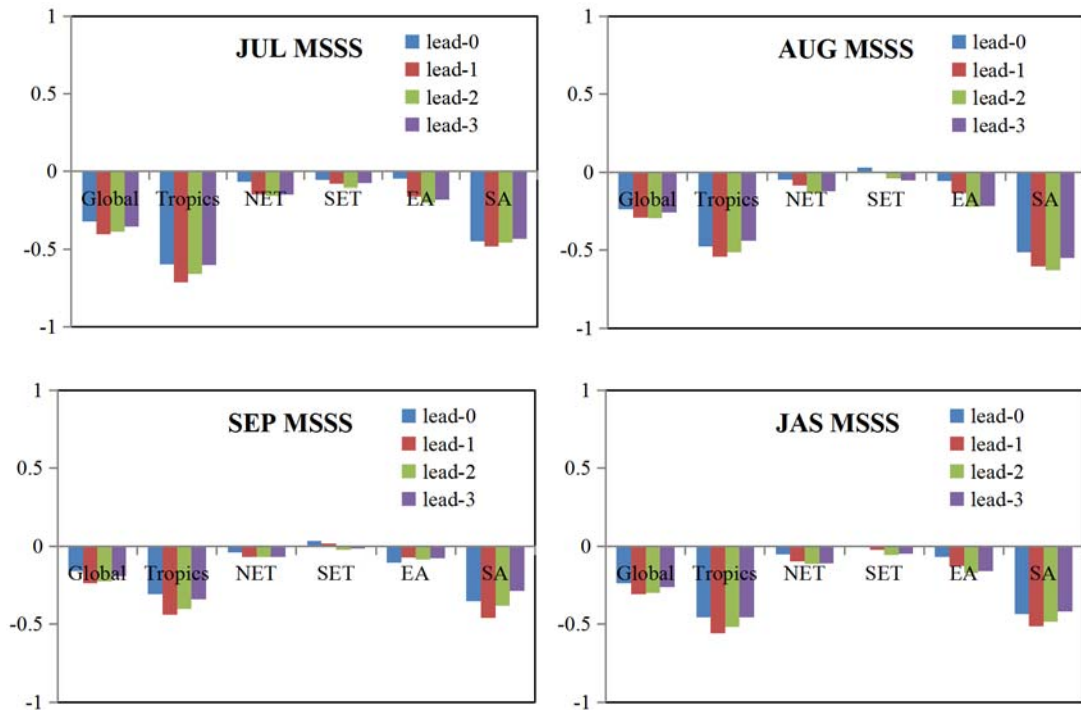


图 24 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的降水 (prec) 的平均方差技巧 (MSSS) 在不同地区的平均

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，热带西太和西印度洋降水明显偏强，东亚中纬度降水偏弱；第二，整体而言，模式对降水有一定预测能力，全球平均 TCC 在 0.2 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；第三，使用 ACC，TCC 等评估方法表明，模式对热带太平洋、特别是 ENSO 地区的降水预测能力较强，对我国而言，对东北和西南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 7 月、8 月和 9 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，在极地和副热带地区略有偏高，这与气温的气候态漂移相互印证（图 1），南半球的环流也有类似的特征，即中高纬偏低而极地偏高，而在西太副高控制地区及其南侧模式与观测偏差不大。

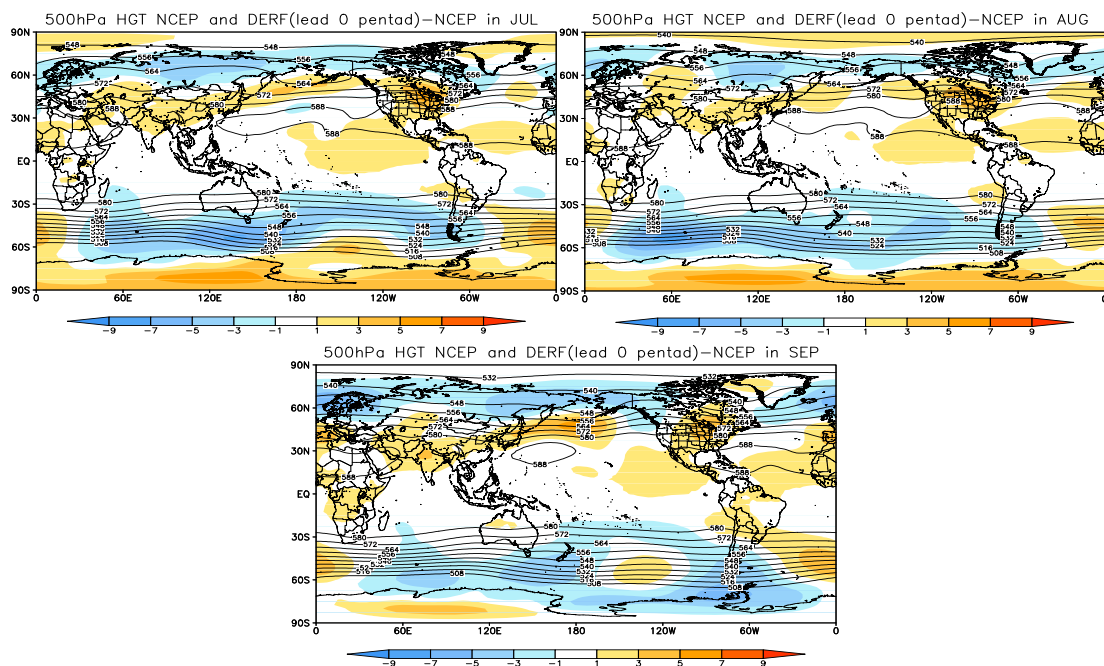


图 25 全球多年 (20a) 平均的 7 月、8 月、9 月 500hPa 位势高度 (单位: 位势什米), 等值线代表 NCEP2 观测, 阴影为 DERF2.0 超前 0 候预报与观测之差

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, 整体高于其他北半球热带外地区, 模式对副高也有较好的把握能力。但需要注意的是, 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。

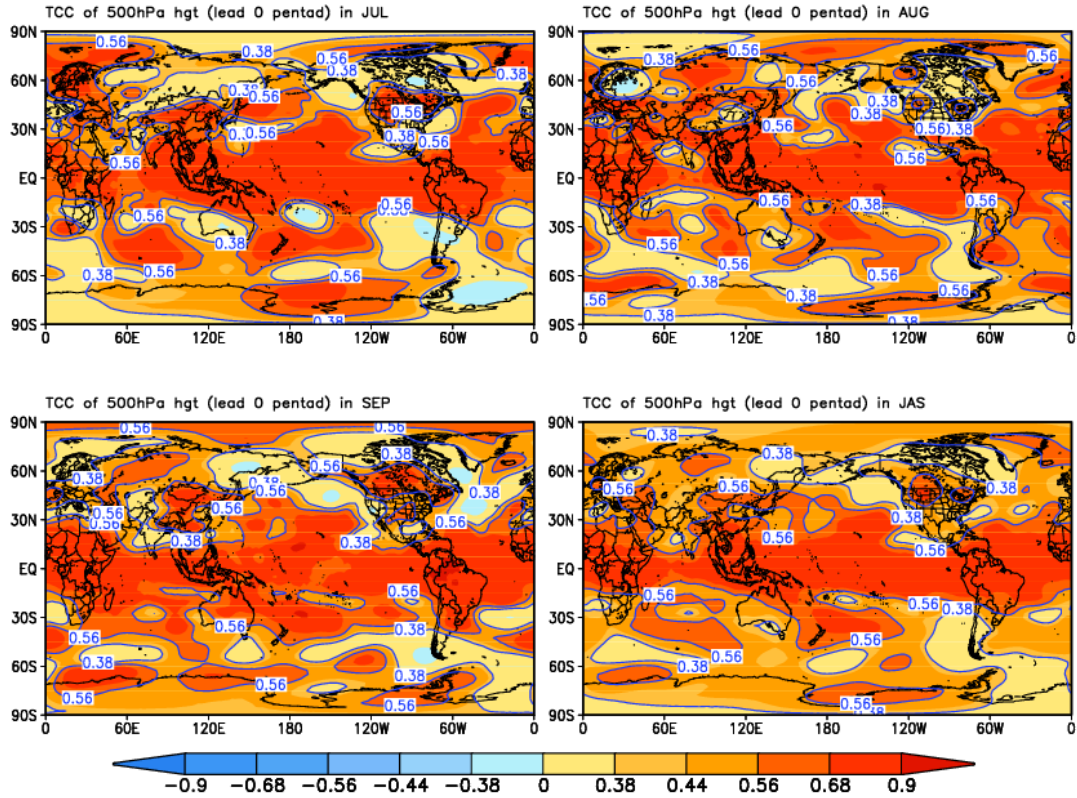


图 26 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的 500hPa 位势高度 (HGT) 的时间距平相关系数 (TCC) 分布, 等值线标定分别代表通过 0.1 和 0.01 显著性检验

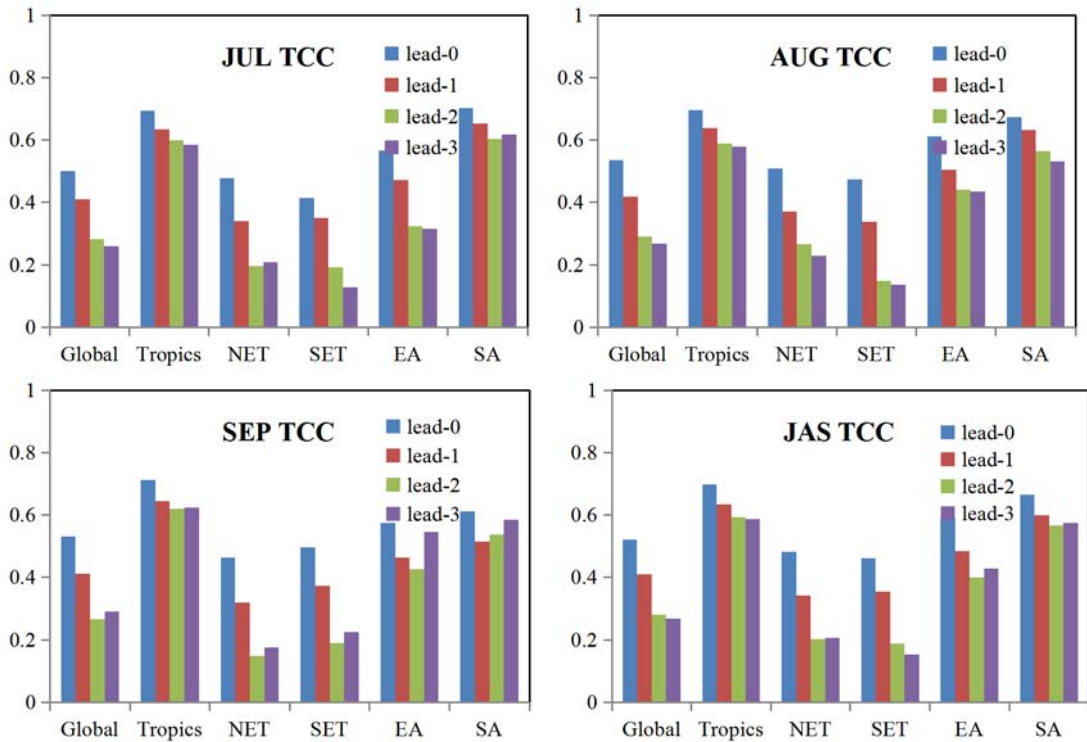


图 27 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的 500hPa 位势高度 (HGT) 的时间距平相关系数 (TCC) 在不同地区的平均

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, 模式对西北太平洋副热带高压、东亚地区的环流有较高的预测技巧, 但热带外的预报技巧均随着 lead time 的延长而迅速减小。

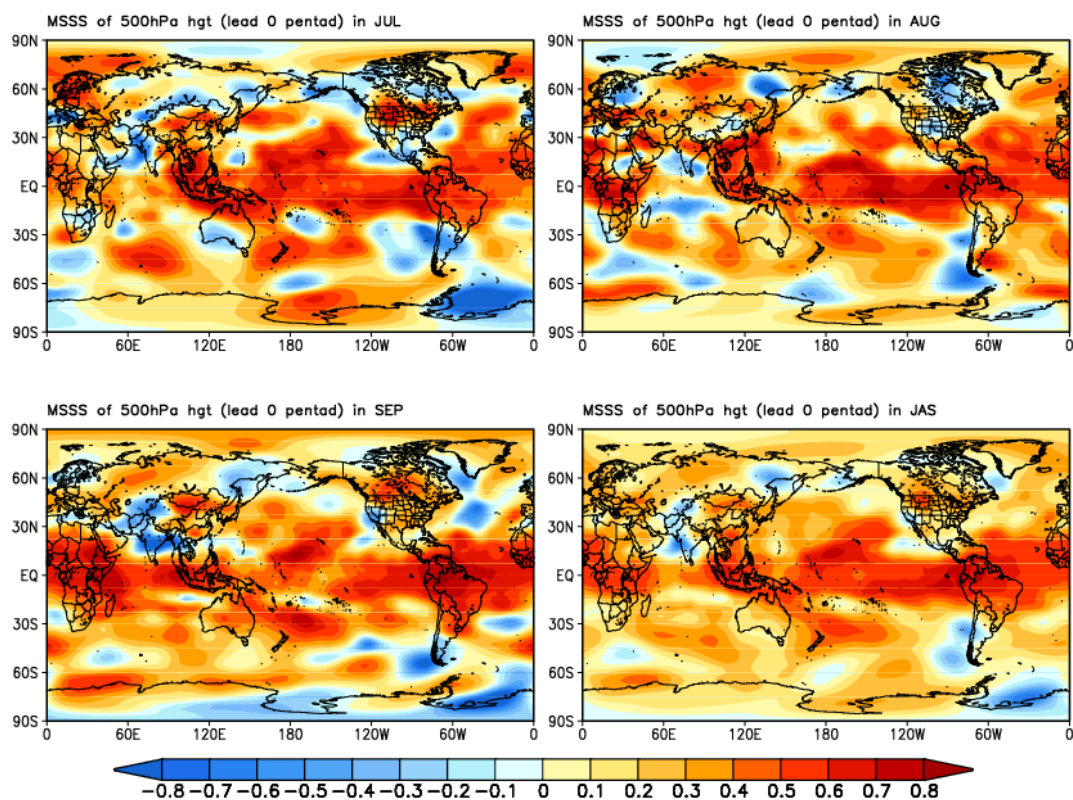
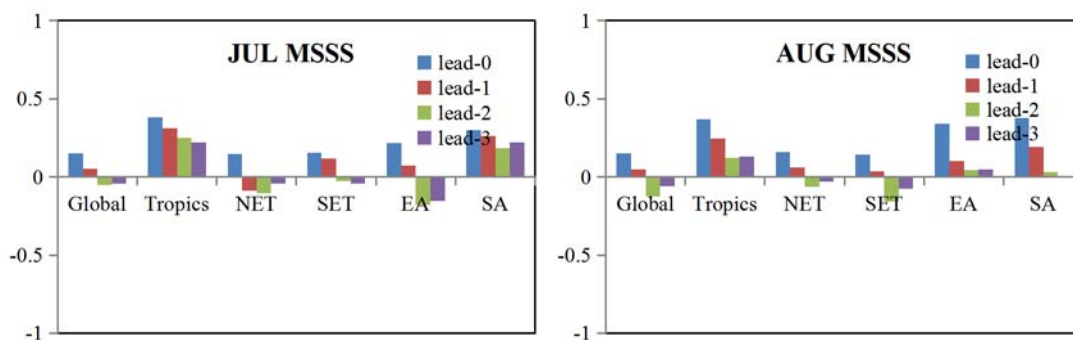


图 28 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的 500hPa 位势高度 (HGT) 的平均方差技巧 (MSSS) 分布



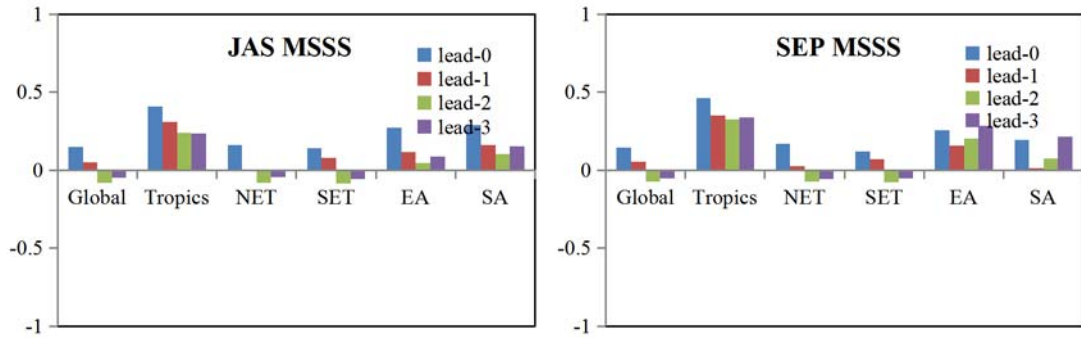


图 29 DERF2.0 在不同 lead time 起报的 7 月、8 月、9 月和第 3 季度的 500hPa 位势高度 (HGT) 的平均方差技巧 (MSSS) 在不同地区的平均

3.3.4 500hPa 位势高度评估小结

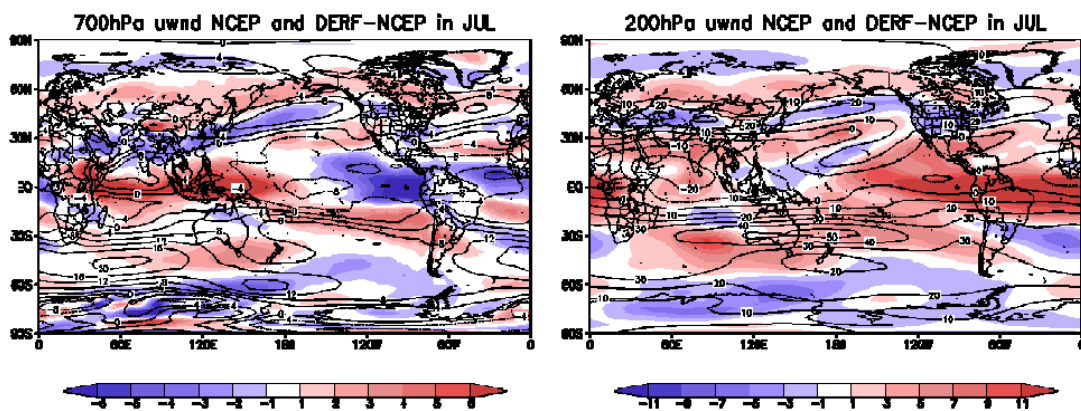
模式对 500hPa 位势高度的预测能力，代表着模式把握整体环流形势的性能，可以为预报员进行主观判断提供重要参考，对其预测能力的综合评估表明：首先，DERF2.0 对环流的把握能力较好，整体明显高于气温和降水；第二，模式主要对热带地区位势高度预报能力较强，具体到影响东亚地区主要系统来说，模式对西北太平洋副热带高压和中高纬度乌拉尔山和鄂霍次克海高压所在地区的环流有较高的预测技巧，有利于预报员通过对环流形势把握从主观上对降水和温度预报做出判断；最后，需要注意的是，模式在热带外地区，对初值十分敏感，随着 lead time 的延长，模式预报性能显著下降，从春到夏也存在着下降趋势。

3.4 200hPa 和 700hPa 纬向风场

与第 2 季度的评估报告相比，本次报告增加了对 200hPa 和 700hPa 纬向风场的评估，进一步检验 DERF2.0 对高空急流和低空主要环流系统的预报能力。

3.4.1 气候态对比

同样首先对模式预报的 7 月、8 月和 9 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西风偏弱，输送水汽偏弱，而在 200hPa 高层，北半球西风急流的强度也偏弱，这也与东亚中纬度地区降水偏弱相互印证。此外，在低层赤道东太平洋的东风信风偏强，高层的东风则偏弱。



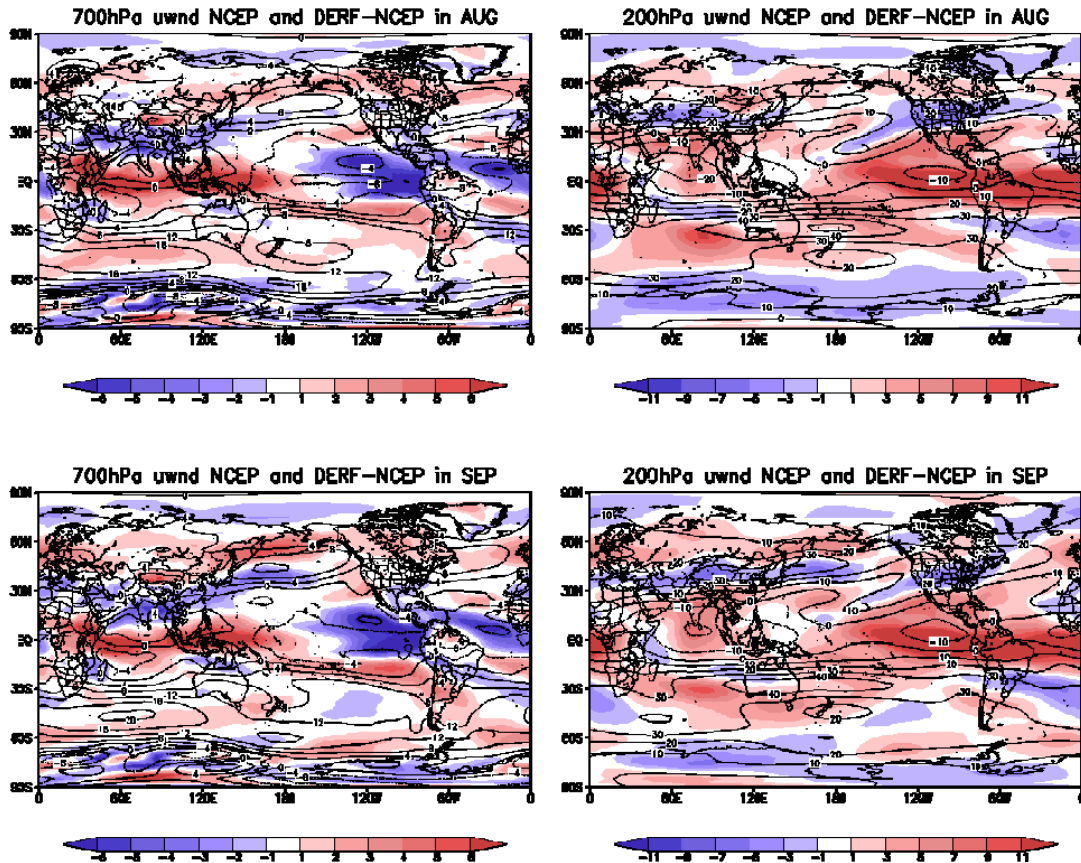


图 30 全球多年 (20a) 平均的 7 月、8 月、9 月 700hPa (左列) 和 200hPa (右列) 纬向风 (单位: m/s), 等值线代表 NCEP2 观测, 阴影为 DERF2.0 超前 0 候预报与观测之差

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 0.4, 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, 特别的在东亚和热带西太平洋地区预报技巧略高于其他地区, 这对把握菲律宾附近对流有很好的帮助。

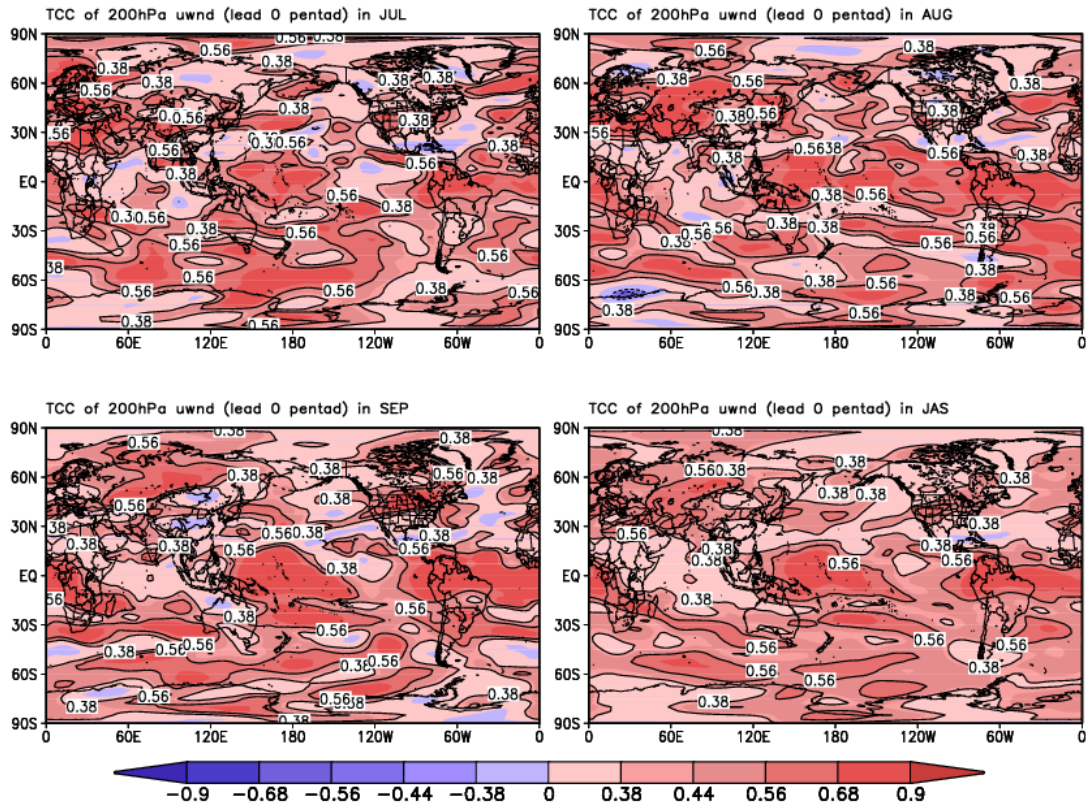


图 31 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的 200hPa 纬向风的时间距平相关系数(TCC) 分布，等值线标定分别代表通过 0.1 和 0.01 显著性检验

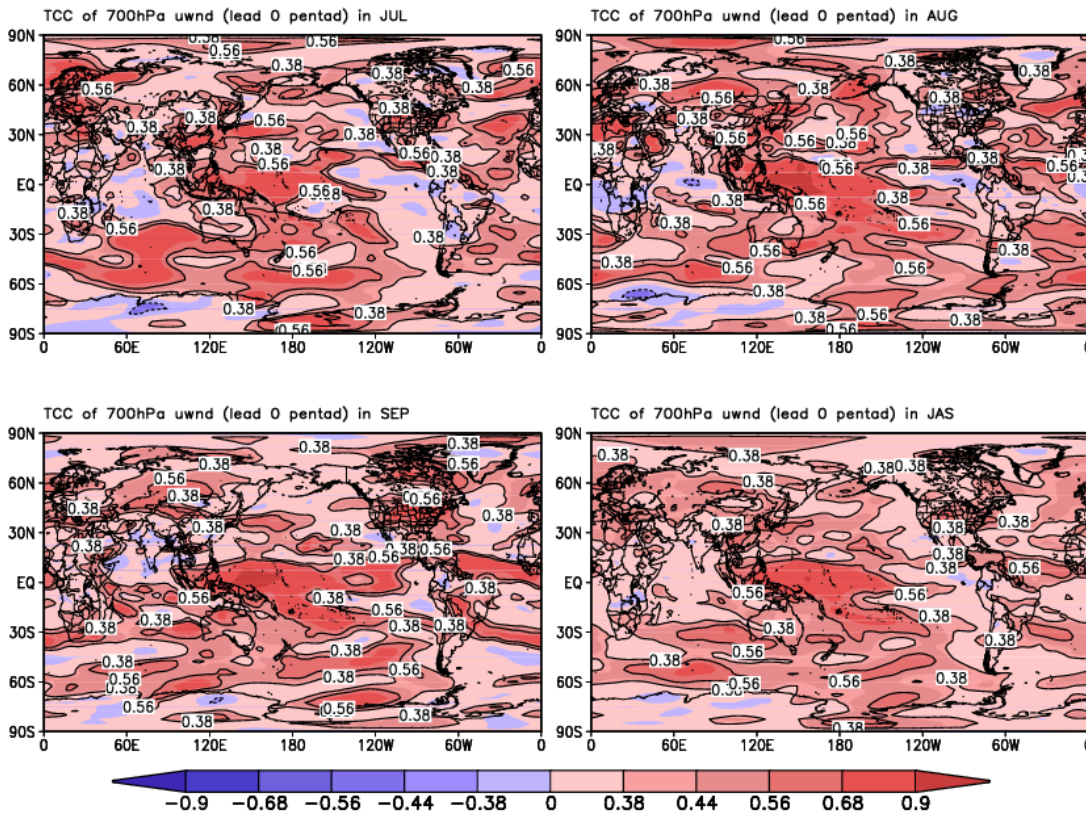


图 32 同图 31，但为 700hPa 纬向风的时间距平相关系数 (TCC) 分布

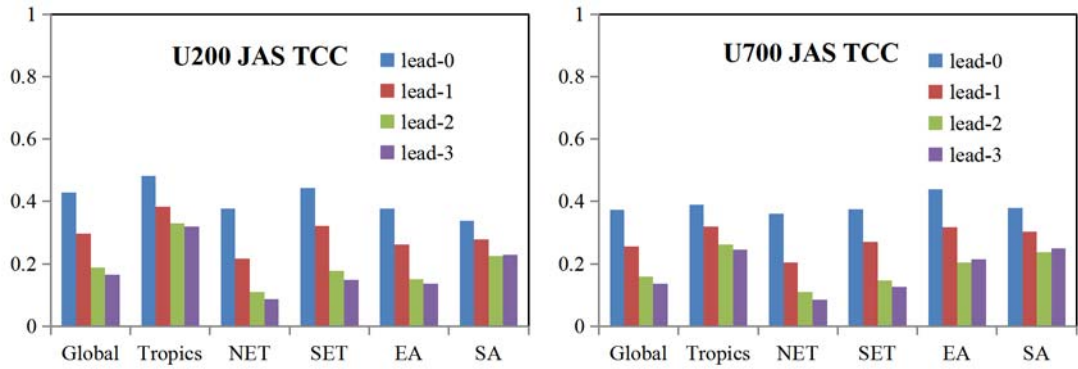


图 33 DERF2.0 在不同 lead time 起报的第 3 季度的 200hPa 和 700hPa 纬向风的时间距平相关系数 (TCC) 在不同地区的平均

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。与 TCC 不同的是，高层 200hPa 热带地区的 MSSS 评分并未显著高于其他地区，在南半球中高纬度和北半球欧亚大陆中高纬地区，MSSS 评分较高。在高低层，模式均对北印度洋地区的预报技巧很低，这一点值得注意。

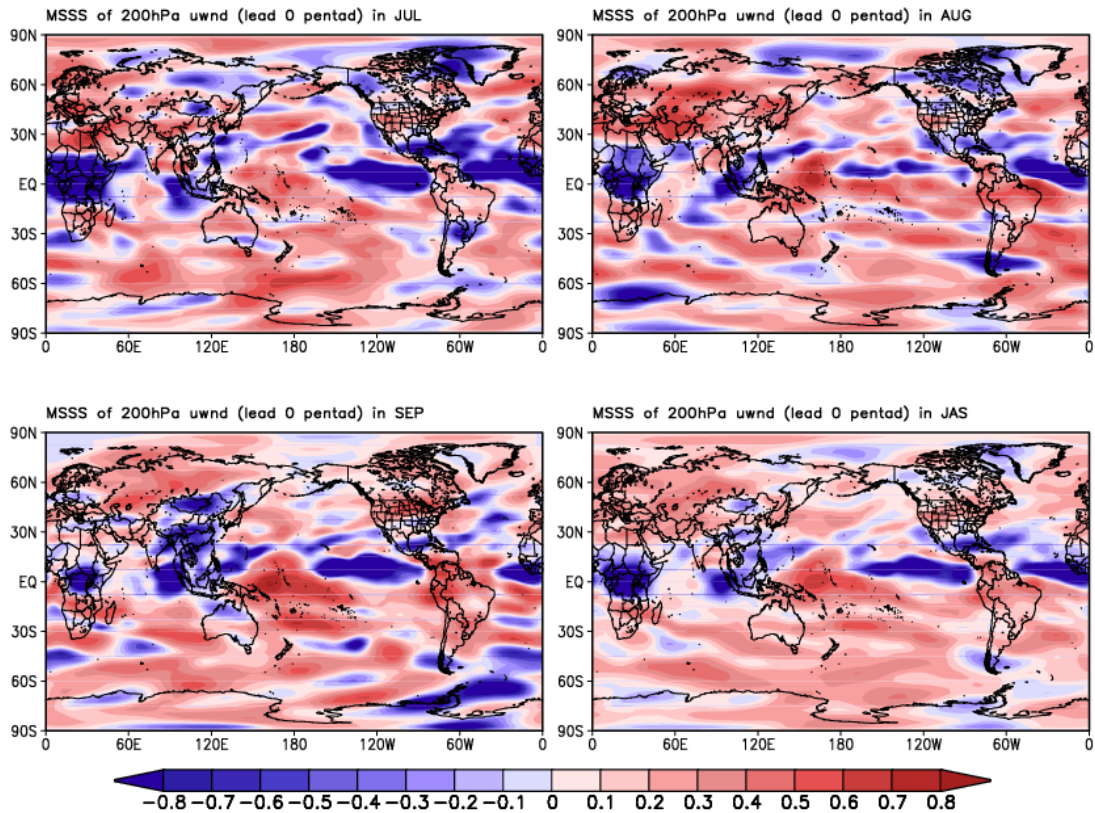


图 34 DERF2.0 超前 0 候起报的全球 7 月、8 月、9 月和第 3 季度的 200hPa 纬向风的平均方差技巧 (MSSS) 分布

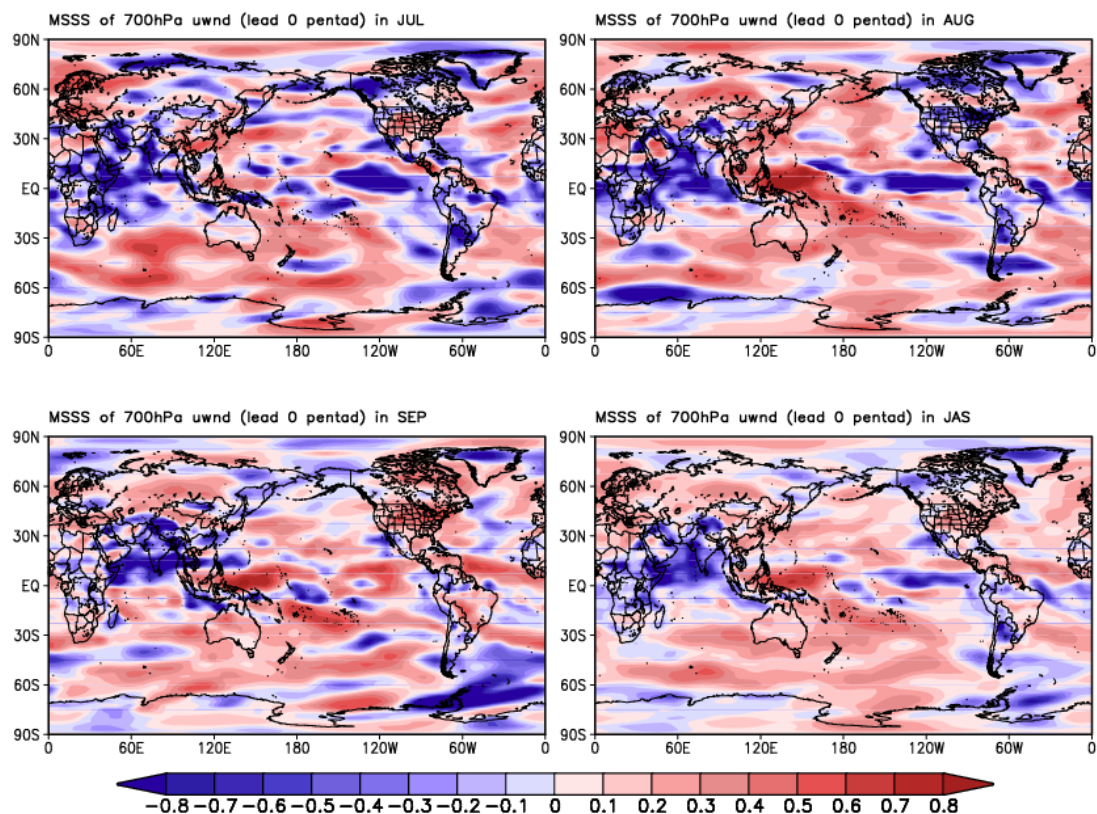


图 35 同图 34，但为 700hPa 纬向风的平均方差技巧 (MSSS) 分布

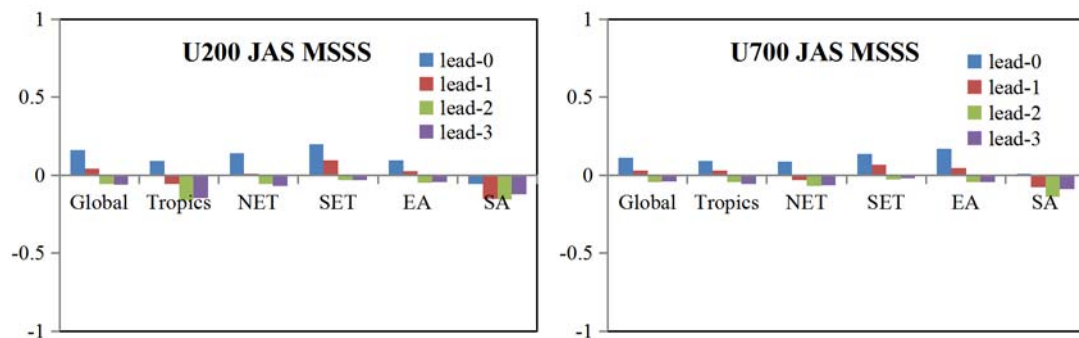


图 36 DERF2.0 在不同 lead time 起报的第 3 季度的 200hPa 和 700hPa 纬向风的平均方差技巧 (MSSS) 在不同地区的平均

3.4.4 纬向风场评估小结

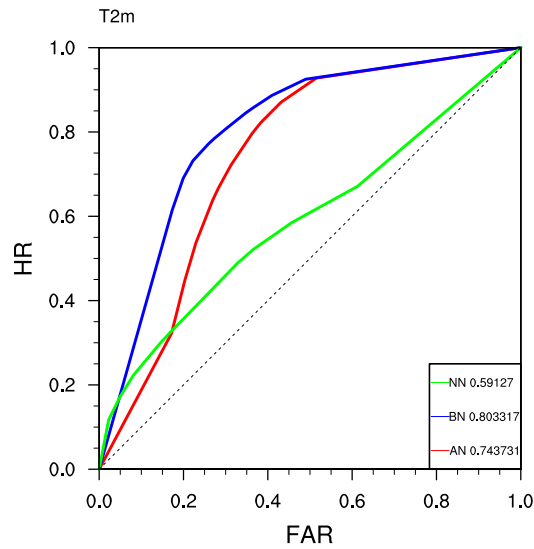
模式对高低层纬向风场的预测，如对高空急流，低层西南季风的预测能为预报员主观判断降水异常的大小和区域提供重要参考，对其预测能力的综合评估表明：首先，DERF2.0 的气候态，在第 3 季度存在着一定的系统偏差，表现为高层西风急流较弱，低层北印度洋上的西风也较弱；第二，DERF2.0 对纬向风场的把握能力好于气温和降水，但要低于 500hPa 位势高度的预报，在高层模式主要对热带地区纬向风场预报能力较强；最后，需要注意的是，模式在高层和低层，对南亚季风区纬向风场异常的预报能力均较弱。

4 概率预报评估 (ROC)

在本次评估报告中，我们采用 ROC 方法，对概率预报结果进行了评估。

4.1 2m 气温 (t2m)

对 2m 气温场 (T2m) 的 ROC 评估表明，模式在热带外地区的预报技巧最高，对热地技巧较低，即有相对较多的样本能够准确判断出气温高/低/常值事件出现与否，特别是对低温异常出现与否把握能力较强。



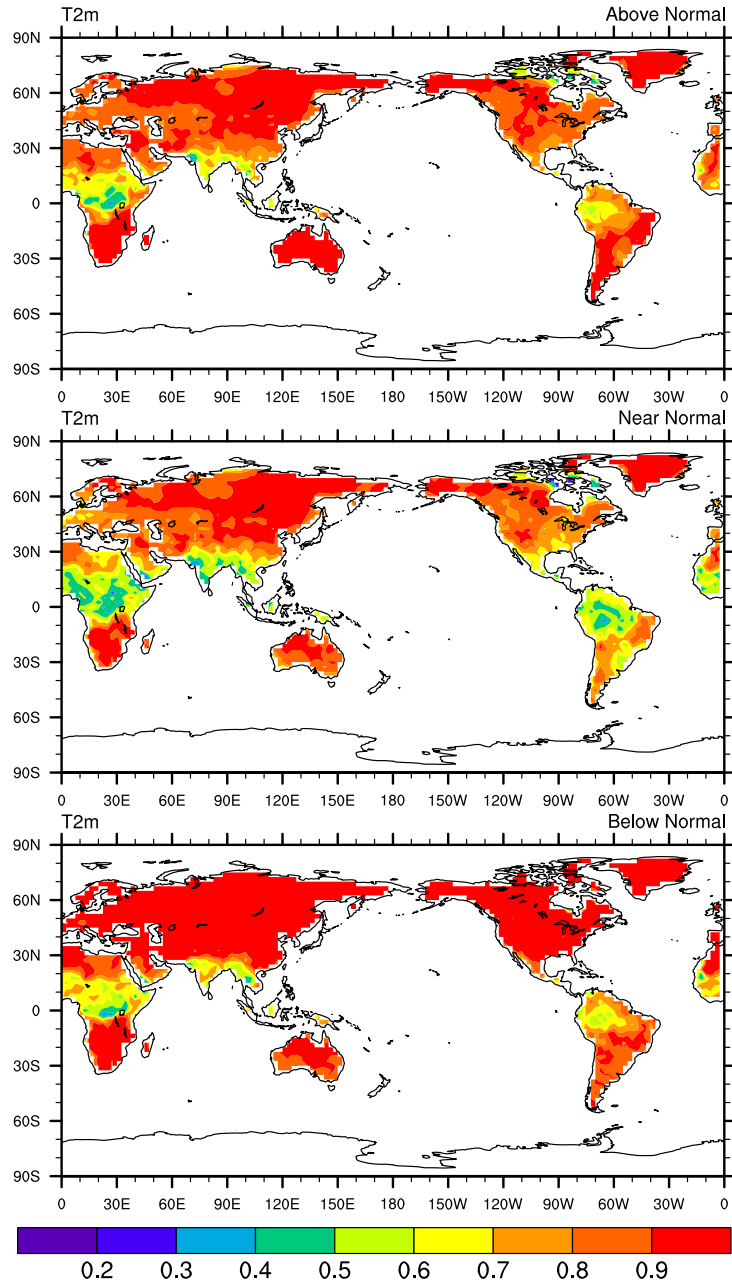


图 37 1991-2010 年 7-9 月 2m 温度 ROC 曲线和 ROCA 面积指数

4.2 降水 (prec)

模式对降水场预报能力与温度相比较差，仅略高于无技巧预报，对不同地区而言，降水场的高/低值 ROCA 指数在热带中太平洋、大西洋西部、非洲大陆、海洋大陆和中国地区均可达到 0.7 左右，具有较高的预报技巧（图 38b）

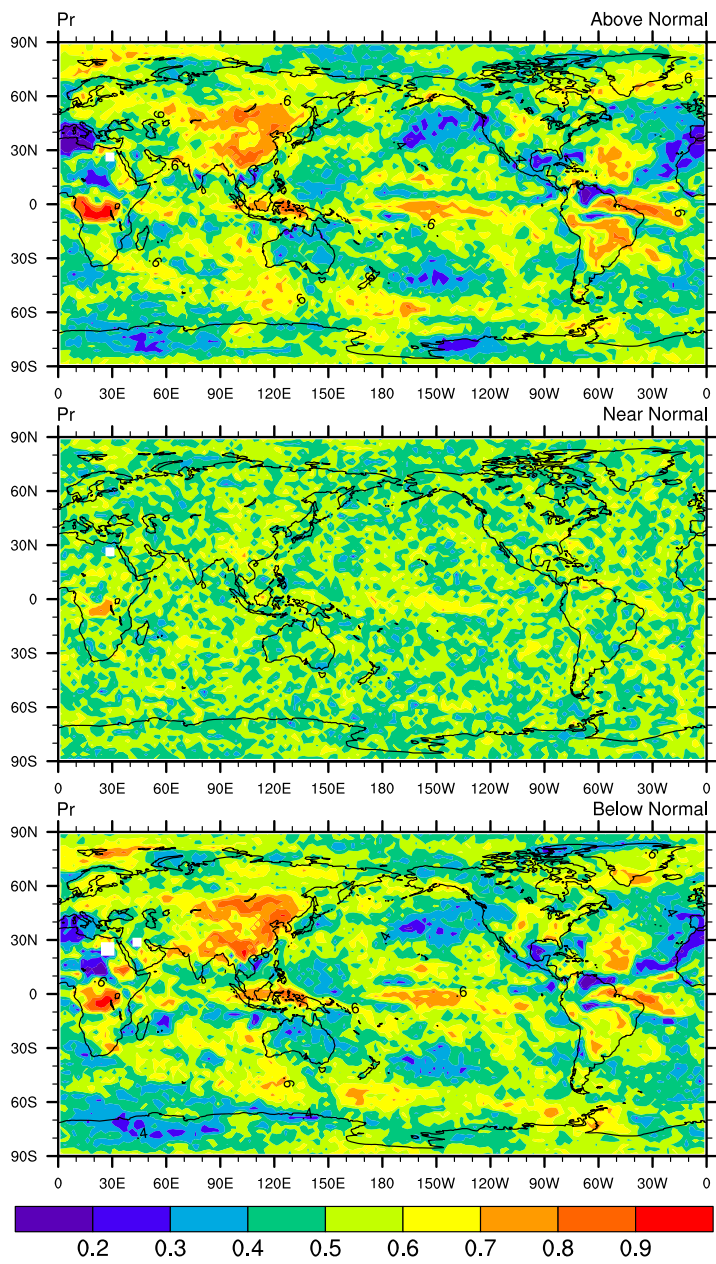
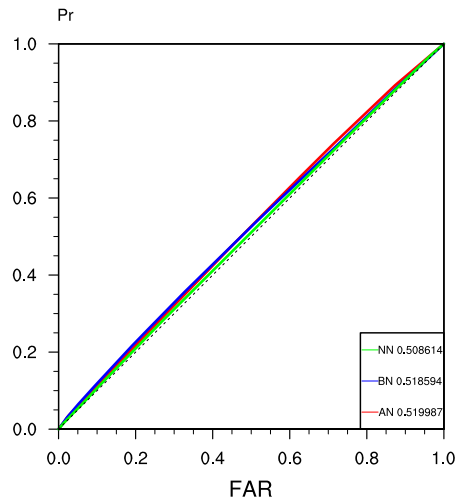
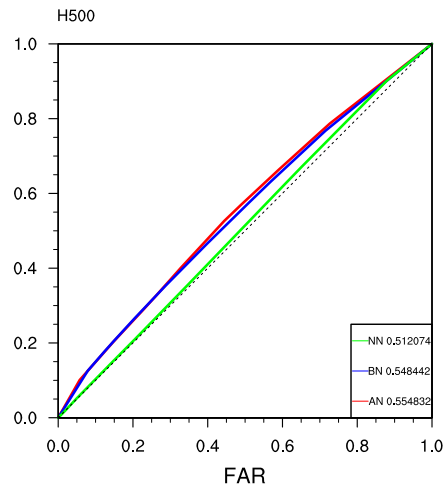


图 38 1991-2010 年 7-9 月降水 ROC 曲线和 ROCA 面积指数

4.3 500hPa 位势高度

如图39所示，在超前0-4候的模式预报结果中，全球平均的1991-2010年7-9月500hPa位势高度ROC曲线表明模式对500hPa位势高度的高值和低值预报技巧相对较高，即有相对较多的样本能够准确判断出高/低值事件出现与否；但对常值的预报技巧偏低，ROCA指数接近0.5。图39为1991-2010年7-9月500hPa位势高度ROCA面积指数，在超前0-4候的模式预报结果中，热带地区的ROCA指数较大，即在热带地区模式具有较高的预报技巧。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。在热带东太平洋、南美洲、大西洋、非洲、东印度洋和中国西部地区，ROCA指数均可达到0.7以上。



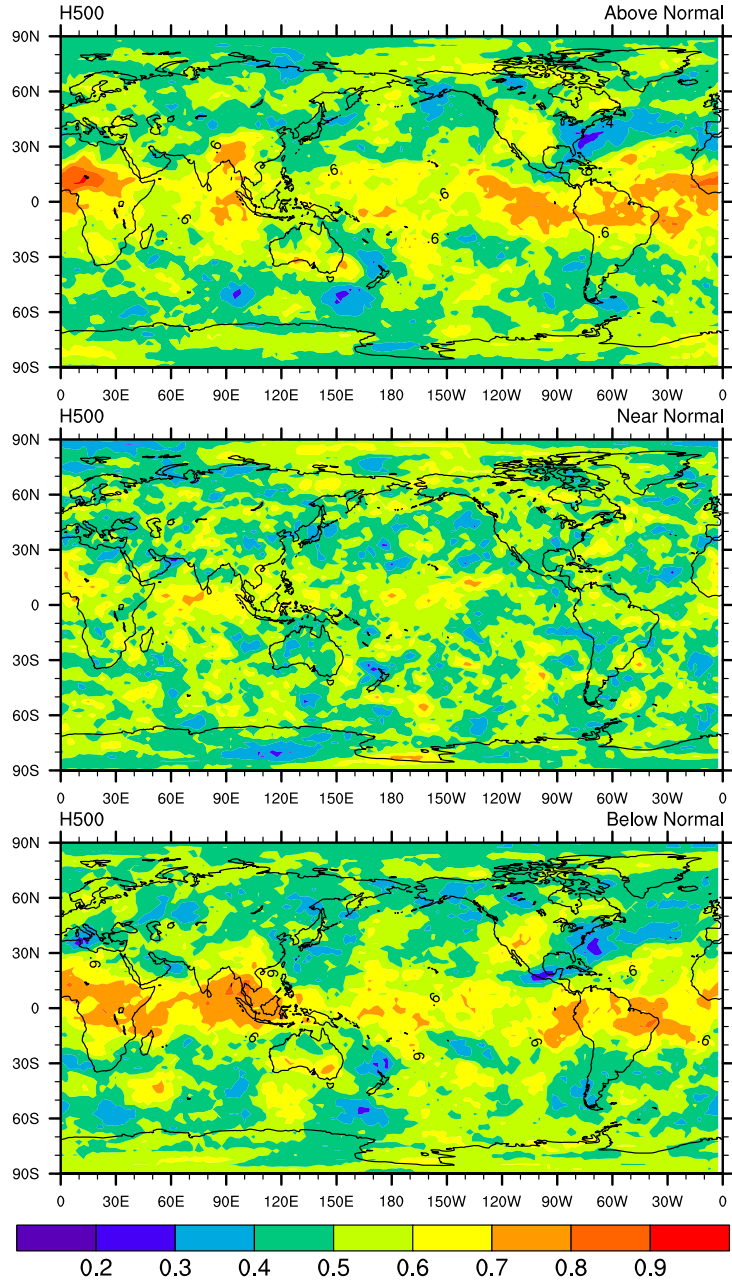


图 39 1991-2010 年 7-9 月 500hPa 位势高度 ROC 曲线和 ROCA 面积指数

4.4 纬向风场

在 200hPa 纬向风和 700hPa 温度场的全球平均结果表明，其常值同样无预报技巧，即相当于随机预报；高值和低值的预报技巧略高（图 40）。在 ROCA 指数的水平分布图中，热带地区的 200hPa 和 700hPa 纬向风的预报技巧均相对其它纬度较高。尤其在热带太平洋地区的 ROCA 最大，分别约为 0.6 和 0.7。

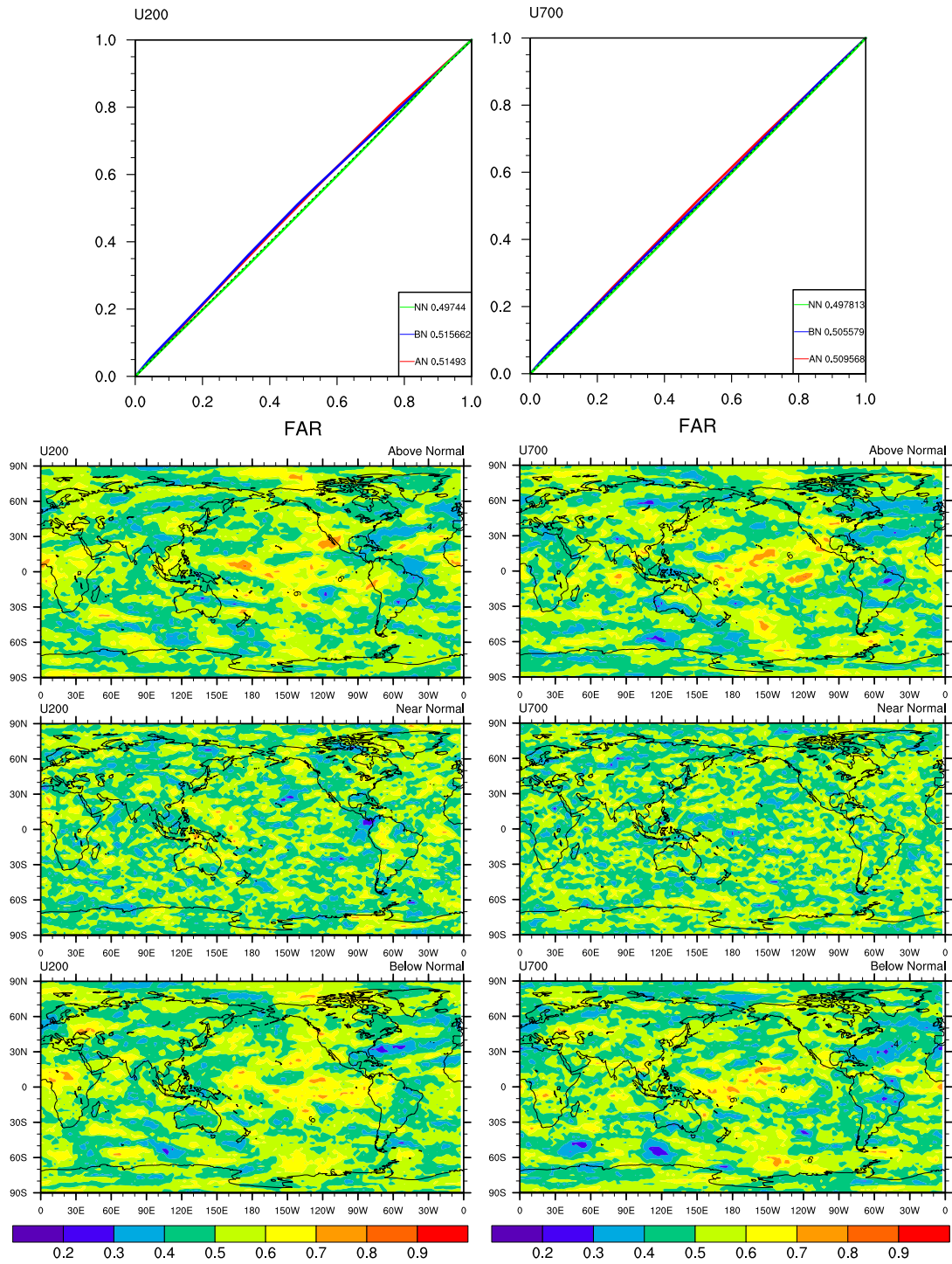


图 40 1991-2010 年 7-9 月 200hPa 和 700hPa 纬向风场 ROC 曲线和 ROCA 面积指数

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，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，除了在热带地区相对较高外，在我国大部分地区也具有较高的预报技巧。气温场（T2m）在热带外地区的预报技巧最高，即有相对较多的样本能够准确判断出气温高/低/常值事件出现与否，可为模式预报结果的校正提供参考。

5 重要季节内振荡现象评估

在本次报告中，我们以 MJO 和北半球夏季季节内振荡 (BSISO) 为例，评估了 DERF2.0 对重要的季节内过程的预测能力。

5.1 MJO

5.1.1 MJO 的主要模态

Wheeler and Hendon (2004) 对 850hPa、200hPa 纬向风和 OLR 构建了一对实时监测 MJO 的 RMM 指数，并被广泛的应用于对 MJO 的业务和科研工作中。但是在 DERF2.0 资料中并未提供 OLR 场，因此我们使用降水场替代 OLR 场，选择气候态为 1998-2010 年共 13 年 (GPCP 逐日降水资料始于 1997 年)，对逐日资料去掉气候上的季节循环 (前 3 个谐波)、前 120 天平均，并对 15S-15N 的资料进行经向平均，再对每个场分别进行标准化，最终进行多变量 EOF 分析，得到 MJO 的两个模态 (图 41)，二者的时间系数具有较高的超前滞后相关，共同表示 MJO 在赤道地区的动传，与 W&H (2004) 使用 OLR 场得到的结果基本一致。因此将 DERF2.0 的风场和降水场投影到该模态上，得到两个 RMM 指数，并对其技巧进行评估。

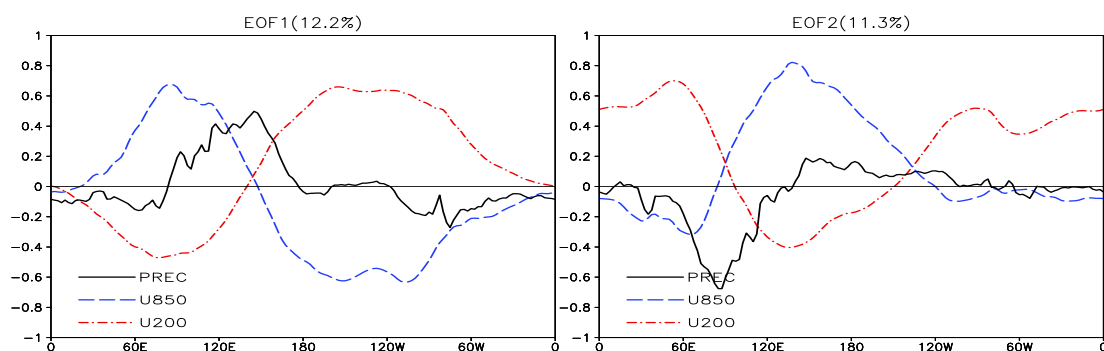


图 41 赤道地区经向平均的多变量 (降水、200hPa 和 850hPa 纬向风异常) EOF 分解的前两个模态的空间分布

5.1.2 RMM 指数的预报技巧

分别对 RMM 指数的相关技巧 (COR)、均方根差 (RMSE)、均方技巧评分 (MSSS) 进行评估，技巧的阈值分别取 0.5、 $\sqrt{2}$ 和 0，对全年和春夏秋冬各个季节的技巧进行计算，如图 42 所示，三种评分标准得到的结论基本一致，整体而言，DERF2.0 对 MJO 的预报技巧约为 14 天，对各个季节来说，秋季最好，夏季次之，春季和冬季的技巧较低，这一点与目前国际上其他模式得到的结论，即冬季较好而夏季较差，有明显的不同。

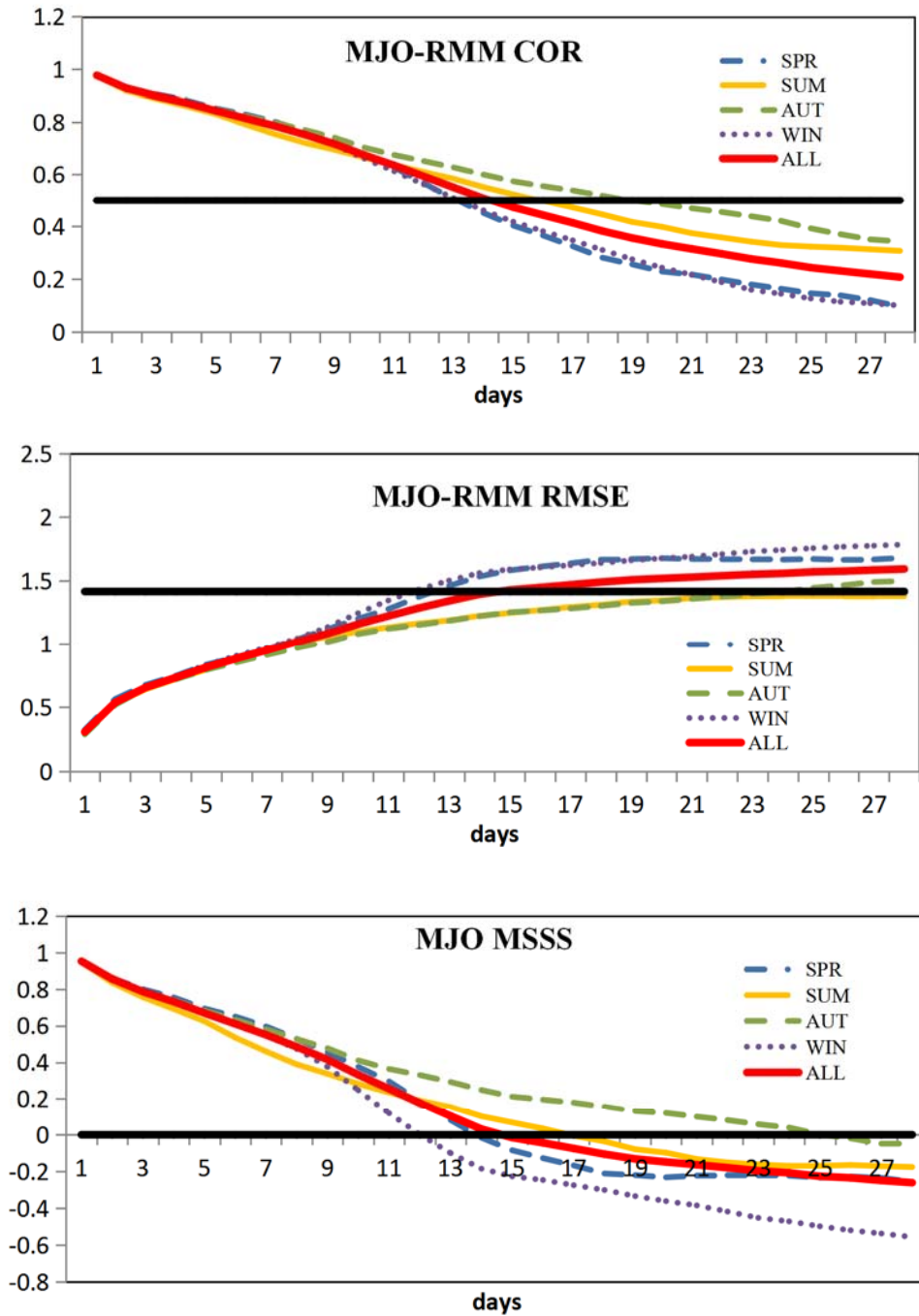


图 42 DERF2.0 对 RMM 指数预报的相关技巧 (COR)、均方根误差 (RMSE) 和均方技巧评分 (MSSS)

图 43 给出在 MJO 位于不同的初始位相和不同的月份时, RMM 指数的相关技巧评分。由图可见, 当 MJO 对流位于第 2 位相和第 3 位相时, 即对流刚刚发展起来强度较强时, 模式对 MJO 的预报技巧较高, 达到 17 天; 当对流位于第 7 和第 8 位相, 即移出热带西太平洋暖池区强度减弱是, 预报技巧较低, 约为 12 天。这一点与国际其他模式的结论相类似。对不同月份而言, 模式在 6 月和 10 月, 预报技巧最高, 可以接近 20 天, 但 4 月技巧最低, 只有 10 天左右。

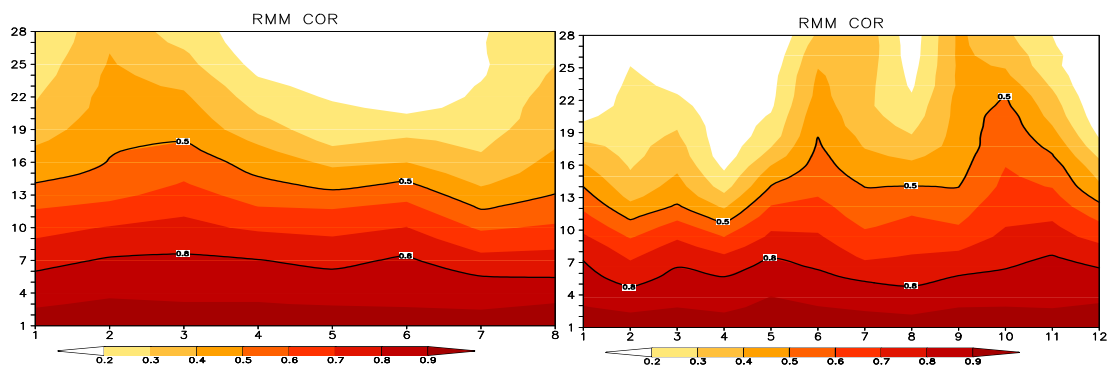


图 43 DERF2.0 对 RMM 指数在不同初始位相 (左图) 和不同月份 (右图) 预报的相关技巧 (COR)

5.2 北半球夏季季节内振荡 (BSISO)

5.2.1 BSISO 的主要模态

Wang 等 (1997) 很早就提出, 在北半球夏季, 除了赤道地区存在东传的 MJO 外, 在亚洲季风区也存在着较强的季节内振荡, 主要包括印度洋北传的 ISO 的西北太平洋西北传的 ISO, 而 RMM 指数对这两个 ISO 并不能很好的表征。Lee 等 (2013) 提出了两对共四个北半球夏季季节内振荡指数 (BSISO index), 发现其能较好的表示亚洲夏季季风区北传和西北传的 ISO。他们具体做法为: 取夏季 (MJAS) 850hPa 纬向风和 OLR 场, 去掉气候上的季节循环 (前 3 个谐波)、前 120 天平均, 再对每个场分别进行标准化, 最终进行多变量 EOF 分析, 取前四个模态, 定义两对 BSISO 指数。这里我们同样将 OLR 场替换为降水场, 图 44 给出前 4 个 PC 回归的 850hPa 风场和降水场, 与 Lee 等 (2013) 的结果非常接近, 其中前两个模态构成 BSISO 指数 1, 代表 30-60 天振荡, 第 3、4 个模态构成 BSISO 指数 2, 代表 10-20 天振荡。该指数的定义虽然使用的是夏季资料, 但通过投影可以得到全年的 BSISO 指数, 此外由于没有进行时间滤波, 该指数也可以进行实时监测。

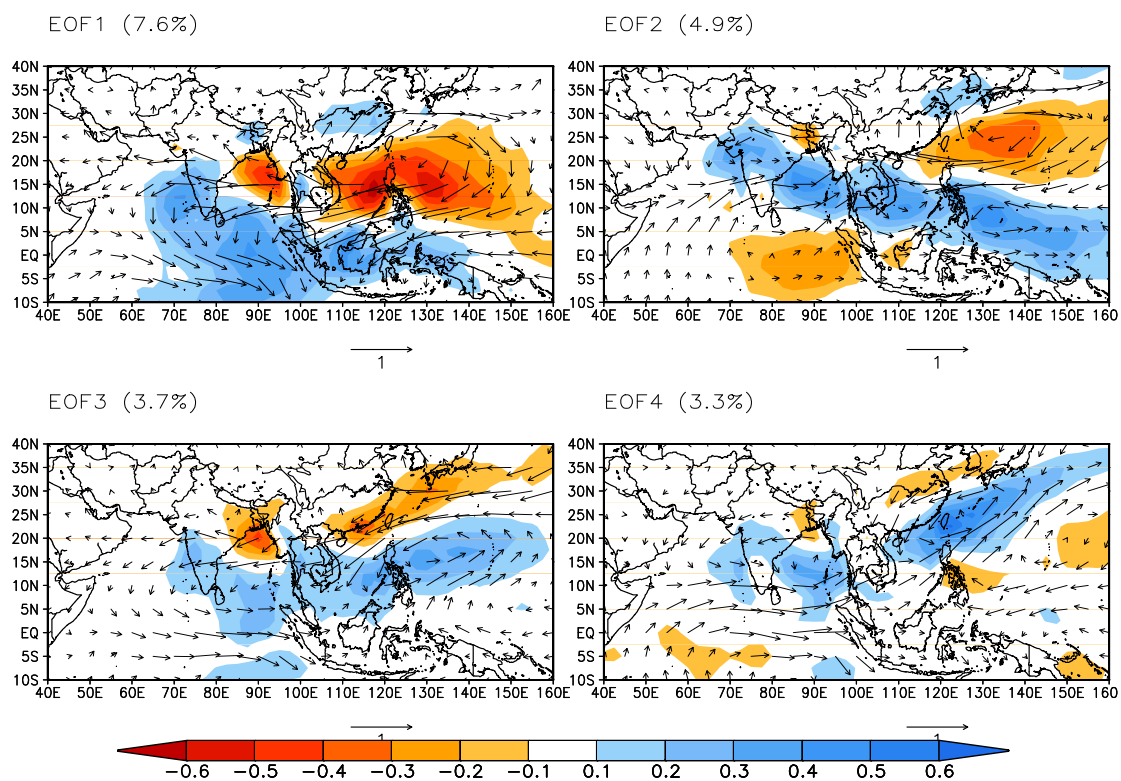


图 44 夏季 BSISO 指数（850hPa 纬向风和降水前 4 个 EOF 模态）回归的（MJJAS）850hPa 风场（矢量，单位：m/s）和降水场（阴影：mm/day）的空间分布

5.2.2 BSISO 指数预报技巧

与对 RMM 指数评估类似，这里也对 BSISO 指数的相关技巧进行评估，如图 45 所示，BSISO1 指数的预报技巧约为 10 天，而 BSISO2 指数的预报技巧约为 8 天，这与两个 BSISO 指数的周期也有一定关系，二者的技巧也均低于 RMM 指数。当然，这里也涉及到二者的可预报性问题。对两个指数在不同月份的技巧评估表明（图 46），BSISO1 在夏季的技巧较高，而 BSISO2 指数的技巧则在冬季较高，这需要进一步的探讨。此外，这两个指数的所代表的模态的方差贡献，总计低于 20%，是否对影响我国的 ISO 有较好的表征也需要进一步研究。

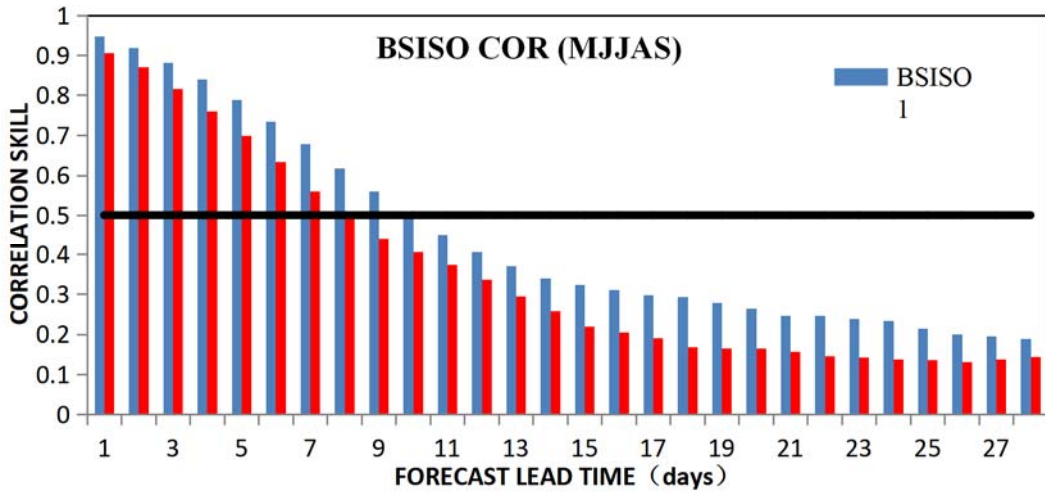


图 45 DERF2.0 对两个 BSISO 指数预报的相关技巧（COR）

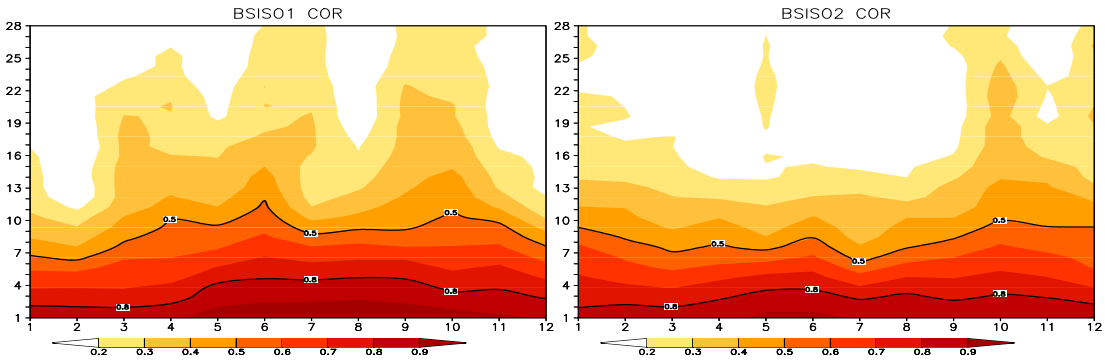


图 46 DERF2.0 对两个 BSISO 指数在不同月份预报的相关技巧（COR）

5 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 3 季度评估结果，与上次报告相比，主要增加了概率预报 ROC

评估和对重要低频过程 MJO/BSISO 的评估，主要得到以下结论：

1) DERF2.0 第 3 季度多年的气候态在某些地区存在系统性的漂移，如在欧亚大陆高纬度地区气温偏低，中纬度气温偏高，对应中纬度地区位势高度场偏高，高空副热带西风急流偏弱，而低层南亚西风偏弱，对应北印度洋和东亚副热带到中纬度地区降水偏弱，热带西太平洋降水偏强。

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；第 3 季度的预报技巧整体低于第 2 季度；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由于其他北半球热带外地区，特别对西太副高所在区域的环流预报技巧较高，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，ENSO 年较高，对我国东北和西南地区的温度和降水预报技巧略高。

4) 概率预报检验 ROC 的结果表明，模式对高/低异常出现对预报的技巧高于正常时的预报技巧，对气温场（T2m）在热带外地区的预报技巧较高，对我国东部地区降水异常的预报技巧也高于其他地区，即有相对较多的样本能够准确判断出气温高/低/常值事件出现与否，可为模式预报结果的校正提供参考。

5) 对 MJO 预报技巧检验表明，模式对 RMM 指数的预报技巧约为 14 天，与其他模式不同的是，DERF2.0 对秋夏 MJO 预报技巧较高而对春冬预报技巧较低；对两个 BSISO 指数预报技巧分别为 10 天和 8 天。

本次报告除了对第 3 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 DERF2.0 对季节内重要过程 MJO/BSISO 的预报技巧，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，增加对重要的大气环流系统，如东亚夏季风、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

- Lee J Y, Wang B, Wheeler M C, et al. 2013. Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Clin Dyn*.
- Lin H, Brunet G, Derome J. 2008. Forecast skill of the Madden-Julian oscillation in two Canadian atmospheric models. *Mon Wea Rev.*136: 4130-4149
- Wheeler M C, Hendon H. 2004. An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon Wea Rev.*132: 1917-1932
- 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