

千米尺度模式降水的检验评估进展及展望

陈昊明 李普曦 赵妍

(中国气象科学研究院 气候与气候变化研究所, 北京 100081)

摘要: 数值模式检验评估是理解模式偏差、完善数值模式和提升预报服务产品质量的有效方法,是现代数值模式发展的基础环节之一。传统的降水业务检验主要基于累积降水量的评估,较少考虑降水的发展演变过程。随着模式分辨率和降水精细化预报需求的提升,仅基于降水量的检验评估已不能满足定时、定点、定量的降水预报服务需求。在回顾已有降水检验评估方法的基础上,重点介绍基于降水频次、强度等小时尺度特征量及其日变化峰值时间的新评估方法,并示例给出此类评估结果对于深入认识日内降水预报偏差的应用价值。初步结果表明,相关方法和结论为丰富和深化数值模式评估,改进客观预报方法具有有益的启发作用。随着千米尺度模式在降水精细化预报业务中的广泛应用,未来仍需进一步发展和完善针对不同区域降水过程分类、分型的细致评估方法,并进一步研发面向快速更新循环系统和对流可分辨集合预报模式应用的检验评估指标,为千米尺度数值模式的改进和模式产品的合理应用提供科技支撑。

关键词: 检验评估, 对流可分辨模式, 精细化预报, 小时尺度降水特征

DOI: 10.3969/j.issn.2095-1973.2021.03.018

A Review and Outlook of Verification and Evaluation of Precipitation Forecast at Convection-Permitting Resolution

Chen Haoming¹, Li Puxi¹, Zhao Yan¹

(1 Institute of Climate System, Chinese Academy of Meteorological Sciences, Beijing 100081)

Abstract: Evaluation and verification of numerical model provides an effective method to understand the model deviation, improve the numerical model as well as improve the quality of forecast service products. The traditional operational verification of precipitation is mainly based on the evaluation of accumulated rainfall amount, and less attention has been paid to the evolution of rainfall systems. With the increasing of the model resolution and the requirement of fine precipitation prediction, the verification and evaluation based on amount only will not satisfy the demand of fine-scale quantitative precipitation forecast service. On the basis of reviewing the existing methods of precipitation verification and evaluation, this paper mainly introduces a new evaluation method based on hourly scale characteristics such as frequency and intensity of precipitation and the peak time of diurnal variation. Examples are given to show the application of such evaluation methods for in-depth understanding of the sub-daily precipitation forecast deviation. The preliminary results show that the results can enrich the numerical model evaluation and provide more useful information to improve the objective prediction method. With the widely application of convective permitting models in operational precipitation forecast, it is necessary to further develop and improve the detailed evaluation methods for different regions through classification, and further develop the verification and evaluation metrics for the application of rapid updating system and convective permitting ensemble forecasting model, so as to improve the performance of kilometer scale numerical model and post-possess the model products to provide scientific and technological support.

Keywords: verification, convective permitting model, fine-scale forecast, hourly precipitation characteristics

0 引言

数值模式的发展已是气象乃至地球系统科学的核心内容,如何合理、有效地评估、改进、完善和应用数值模式系统是发展的主要内容。数值模式产品或多或少的长期存在着不确定性,因此必须细致开展对

数值模式结果的评估和订正工作,以获得更高质量的预报服务产品^[1]。模式评估是理解模式偏差、完善数值模式和提升预报服务产品质量的有效方法,是现代数值模式发展的基础环节之一。在以数值模式为核心的气象业务新时期,模式评估是避免或减轻数值模式结果不确定性影响的关键。

近年来,在多源观测系统及天气气候一体化模式研发的支撑和推动下,我国区域高分辨率数值预报业务发展迅速^[2-4]。国家级^[5-6]和华北^[7-8]、华东^[9]、华南^[10]

收稿日期: 2020年9月15日; 修回日期: 2020年1月28日
第一作者: 陈昊明(1983—), Email: chenhm@cma.gov.cn
资助信息: 国家重点研发计划(2018YFC1507605)

等区域气象中心相继建立了水平分辨率为3 km的区域高分辨率数值预报系统，逐步实现了区域对流可分辨尺度模式的业务化运行，模式产品广泛应用在短时临近预报预警以及短期预报业务和服务中。用于开展区域气候模拟和预估的模式水平网格距，也在近十年被国际领先的模式研究机构提升至千米级^[11]。模式在此水平分辨率下可以显式表征深对流过程，从而摆脱对于对流参数化方案的依赖，此类模式称为“对流可分辨模式（Convection-Permitting Model）”。传统数值模式所采用的对流参数化方案中对流触发的判据相对简单，对流在边界层的夹卷过程不受约束，同时缺乏对对流过程由浅到深过渡过程的刻画^[12-14]，长期以来被广泛地认为是降水模拟不确定性的主要来源^[15-17]。相比于传统数值模式，对流可分辨模式具有较高的模式水平分辨率，并且能够显式表征深对流过程，在对流过程相关各要素的总体统计特征以及与对流过程相关的辐射过程、陆表过程和大尺度环流等方面的模拟上均呈现出显著的模拟优势^[11, 18-19]，已逐渐成为开展区域高影响天气预报、季节预测、气候模拟和未来预估的有力工具。

中国地处东亚季风区，受复杂地形、不均匀地表加热和季风环流等多种不同尺度因素的综合影响，降水成因复杂，特征多样。当前数值模式对东亚夏季风降水的模拟/预报尚面临诸多问题，而数值模式对云及对流过程的不准确表征是导致模式模拟/预报偏差的重要原因之一^[20-22]。强降水是我国致灾最为严重的气象灾害之一，常导致流域洪涝和城市内涝，继而引发滑坡、泥石流等地质灾害。受东亚特殊地理环境、多尺度天气系统和海陆气相互作用的影响，我国强降水的时空分布和演变过程复杂，区域特征明显且差异显著^[23-25]。当前数值模式对强降水的预报准确率依旧很低，且不同地区预报偏差各不相同^[26-27]。目前仍缺乏对模式降水预报特征及其产生原因的深入认识和理解，特别是针对区域对流分辨模式产品的检验评估方法严重滞后于当前高分辨率数值预报系统的快速发展。基于新的高分辨观测资料的客观综合评估方法和手段，揭示模式的预报能力及关键偏差，发现模式预报产品的问题，从而促进数值模式的改进和完善，同时通过评估给出预报的可适用性和不确定性，对于提高强降水的预报能力和数值预报产品的应用能力具有重要的科学意义和实用价值。

1 基于累积降水量的检验评估

1.1 基于站点或格点的传统常规检验评估方法

降水预报准确率的业务检验主要关注于累积降

水量的分量级评分^[28-30]，如基于二分类事件的成功指数（Threat Score, TS）、公平成功指数（Equitable Threat Score, ETS）和考虑降水气候概率的概率空间稳定公平误差（Stable Equitable Error in Probability Space, SEEPS）等。此类降水检验评分以累积降水量为检验对象，基于点对点的对比或者利用列联表，根据事件发生与否对预报、观测进行分类并统计观测点上预报、观测事件发生的次数，在此基础上定义一系列评分指数来判别模式的预报技巧。此类评估方法优点是简单直接，计算易于实现且结果容易理解；但往往并不能反映预报的真实水平，特别是对于小概率事件，如暴雨的预报检验存在较大不足^[31]。考虑中国暴雨可预报性的时空分布不均匀问题，为合理对比不同区域暴雨预报能力的差异，陈静等^[32]构建了一种新型的中国暴雨可预报性综合指数（Synthetic Predictability Index of Heavy Rainfall, SPI），以暴雨气候频率、暴雨面积比率和模式暴雨预报TS三个分量来定量描述中国各区域的暴雨可预报性特征。在此基础上，陈法敬等^[33]提出了预报与观测的邻域匹配方法，以给出更稳定的评分结果。

随着模式分辨率的提升，常规检验方法已不能满足于高分辨率模式降水检验评估的需求。即使降水的雨带结构及强度均与实况一致，雨带位置的细微偏差依然可能导致过大的空报率及漏报率，使得最终的预报评分偏低。特别是降水预报同时受到空间和时间微小差异所带来的“双重惩罚”，很难获得足够的检验评估信息^[29, 34]。除“双重惩罚”问题外，传统检验方法的局限性更多地体现在评分与误差种类和大小无关，如TS评分往往倾向于奖励“更平滑”的空间型预报。更重要的是，此类评分结果不能提供足够多的模式偏差信息，比如模式偏差的种类（位置、强度和分布型）以及针对模式偏差的诊断信息，无法将检验评分的高低与模式对真实物理过程表征的正确与否建立内在的物理联系，很难从评分中获取哪些预报更接近于实况的时空分布特征^[35]，因此传统检验评分结果难以针对区域高分辨率模式的研发给出基于物理过程的改进意见或建议。

1.2 空间检验方法

为解决传统预报检验方法在高分辨率模式检验评估中的“双重惩罚”问题，近年来国内外相继发展了一系列新的基于降水空间结构特征的检验方法^[36-37]（如邻域法^[38-39]、尺度分离法^[40-41]、场变换法^[42-43]以及基于对象的评估方法^[39, 44-46]等），以从高分辨率模式预报结果中提取更多的空间信息。

邻域空间检验方法是考虑相邻区域的空间特征的一种检验方法,将较高分辨率的预报与观测信息采用升尺度的方法调整至较大尺度上,通过空间平滑或统计概率分布的方式减小高分辨率资料的偶然性信息,以衡量预报与观测间的相似度特征。目前应用较为广泛的邻域空间检验方法为Roberts和Lean^[47]在2008年提出的分数技巧评分(Fraction Skill Score, FSS),该方法通过设置不同的邻域半径与不同降水阈值计算FSS值,考察数值模式对于不同空间尺度和不同强度的降水空间分布的预报性能。邻域半径概念的引入较为有效地解决了面向高分辨率模式降水传统检验中所遇到的“双重惩罚”问题;同时,由于与传统检验方法采用的评分指标类似,FSS的计算同样简单。欧洲中期天气预报中心、英国气象局、美国环境预报中心近年来都先后将FSS作为高分辨率模式降水业务预报的业务评估方法。但是,由于此类方法侧重于空间上的对比,升尺度的方法可能会漏掉一些中小尺度信息^[48],同时对天气过程的发生、持续和结束等事件特征仍缺少客观、定量的检验评估。

美国国家大气研究中心(National Center for Atmospheric Research, NCAR)研发的基于对象的诊断方法(Method for Object-Based Diagnostic Evaluation, MODE)^[44-45]已经应用于对美国降水确定性预报的检验评估,并在近年来不断发展、完善。当前已有部分前沿性研究工作使用基于对象的时域诊断方法(MODE that Incorporates the Time Dimension, MODE-TD)开展针对美国地区降水系统的甄别与追踪^[49],并用来考察对流分辨模式对于中尺度对流系统降水特性的模拟性能^[50];英国气象局也已围绕降水的时空间尺度开展相关研究,考察降水系统的特性与时空尺度的联系,并以此为观测指标进行针对气候模式的模拟评估^[51-52]。此类基于对象、面向过程的时空连续分析诊断方法,可以较为系统、全面地考察模式对于特定物理过程的模拟性能及模式偏差。同时把数值模式偏差的具体特征和改进建议及时反馈到模式研发团队,使模式检验评估和模式研发二者有机结合,加速推动一体化模式的研发进程^[53]。

2 面向强降水精细化预报的检验评估

随着城市化进程不断推进,社会对精细化气象预报的要求越来越高。定时、定点、定量的降水预报准确率受到越来越多的关注。基于累积降水量的检验评估已不能满足精细化预报服务对于高分辨率模式产品的应用需求^[54]。中国气象局《区域高分辨率数值预报业务发展规划(2018—2020年)》中明确提出“区域模式重点面向可用时效在24小时内的短时临近预报预

警和无缝隙精细化气象预报业务需求”,针对千米尺度区域模式的发展定位,相应的检验评估工作也必须从基于累积降水量的评估转变为针对降水发展演变过程的评估,特别要关注降水在日内时间尺度的特征和演变过程的预报能力。如50 mm/d的强降水过程,既可能来自持续时间为24小时的长持续弱降水,也可能来自于持续3小时的短时强降水;而后者既可能是发生在午后的局地强对流降水,也可能发生在夜间(图1)。传统的基于累积降水量的检验评估未考虑降水日内准确率的预报效果。围绕精细化预报发展需求,需要从降水小时尺度特征,如降水频次、强度、持续时间和日变化等方面对数值预报产品进行更细致的评估,只有模式对于降水的发展演变过程的预报都是正确的,模式才能预报出真实的降水^[55,56]。

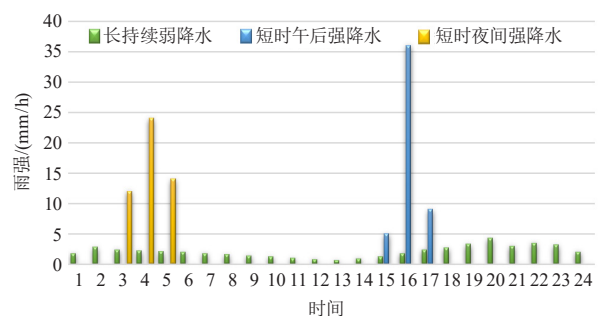


图1 不同的50 mm/d强降水过程示例图

Fig. 1 Examples of different rainfall events with same 50 mm/d amount

小时降水频率和强度是反映降水小时尺度特征的重要物理量^[57]。小时频率的定义为统计时段内有降水时次除以非缺测时次,小时强度的定义为降水量除以有降水时次,如公式(1)和(2)所示。基于频率和强度可以有效区分图1中两类不同持续时间的降水过程,即短时(长持续)降水表现为频率小(大)而强度大(小)的特征。

$$Freq(h) = \frac{N_{rain}}{N_{sample}}, \quad (1)$$

$$Inten(h) = \frac{A_{rain}}{N_{rain}}, \quad (2)$$

其中: h 为预报时间, N_{rain} 为统计时段内 h 时刻的有效降水时数(降水量 ≥ 0.1 mm/h), N_{sample} 为统计时段内 h 时刻的总非缺测样本数, A_{rain} 为统计时段内 h 时刻的累积降水量。

图2给出国家气象信息中心多源融合降水产品(CMPA,图2a~2c)和ECMWF高分辨率全球模式(模式分辨率9 km,下文简称EC,图2d~2f)与国家气象中心GRAPES-Meso-3km区域模式(下文简称GRAPES,2g~2i)9~33小时预报的2019年暖季

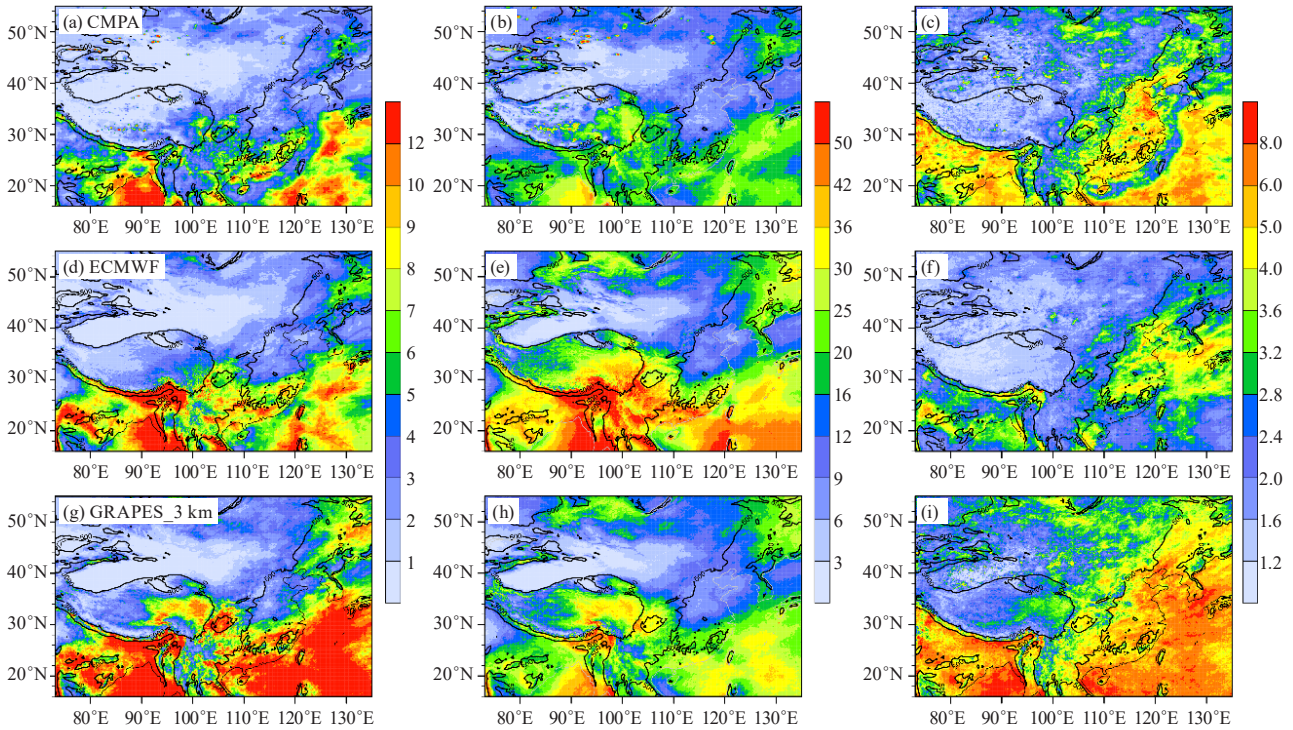


图2 2019年暖季(5—9月)CMPA(a~c)、ECMWF(d~f)和GRAPES-Meso-3 km(g~i)模式9~33小时小时预报的平均降水量(a, d, g, 单位: mm/d)、小时降水频次(b, e, h, 单位: %)和降水强度(c, f, i, 单位: mm/3 h)的空间分布(图中等值线为500 m和3000 m地形高度线)

Fig. 2 Spatial distribution of mean precipitation (a, d, g, unit: mm/d), hourly precipitation frequency (b, e, h, unit: %) and hourly intensity (c, f, i, unit: mm/3h) by CMPA (a-c), ECMWF (d-f) and GRAPES-Meso-3km (g-i) during the warm season (from May to September) in 2019 (The black lines represent 500 m and 3000 m terrain heights)

(5—9月)平均降水量(图2a、2d、2g)、频率(图2b、2e、2h)和强度(图2c、2f、2i)的空间分布。模式较好地再现了降水量的空间分布(图2),但对南方大部分地区的平均降水量有高估,且GRAPES对于地形周边降水的高估更为明显。EC和GRAPES与CMPA的空间相关系数分别为0.83和0.74,均方根误差分别为3.02 mm/d和4.92 mm/d,可见仅从日累积降水量看,ECMWF全球模式对于降水中心和量级的预报仍略优于区域模式的结果。

由小时降水频率的分布可见,我国大陆暖季降水高频区主要位于华南沿海、青藏高原东南部至高原东坡以及四川盆地以东的二阶阶梯地形周边。模式均可以再现主要降水高频区的空间分布,EC和GRAPES与CMPA的相关系数分别为0.75和0.76,均方根误差分别为7.55%和14.93%,即模式对小时降水频率均有高估,但GRAPES更接近融合实况产品的结果。EC模式显著高估了青藏高原和我国南方地区暖季降水频次的量值,且降水频次的偏多在35°N以南、二阶阶梯地形以西区域更明显。从降水强度分布来看,EC和GRAPES与CMPA的相关系数分别为0.61和0.76,均

方根误差分别为1.33 mm/3 h和1.67 mm/3 h。EC模式除华南沿海地区外,能够较为合理地再现二阶阶梯地形以东降水强度大、以西强度相对较小的空间分布特征。与降水频率的高估不同,EC模式对我国绝大部分地区降水强度均有较明显低估。GRAPES对于小时降水强度的预报明显优于全球模式的结果。综合频率和强度的预报结果可见,尽管EC模式对于累积降水量的预报较好,但其对我国大陆大部分区域小时降水特征的预报存在高估降水频次与低估降水强度的问题,表明其合理的日降水量并不是来自于准确的降水过程预报。而千米尺度区域模式GRAPES,尽管对于降水量的预报较EC无明显优势,但其对于降水频率和强度的预报更为真实,因此在日内时间尺度上具有更高的参考价值。

为进一步定量比较数值模式预报的不同强度降水量的区域差异,采用降水量随强度分布的 e 指数拟合方法来量化比较模式预报不同强度降水的结构偏差^[58]。 e 指数拟合方法考虑到降水量随小时降水强度增加呈 e 指数衰减趋势(如图3c曲线所示),可用式(3)表示,其中 I 表示小时降水强度, $A(I)$ 表示降水

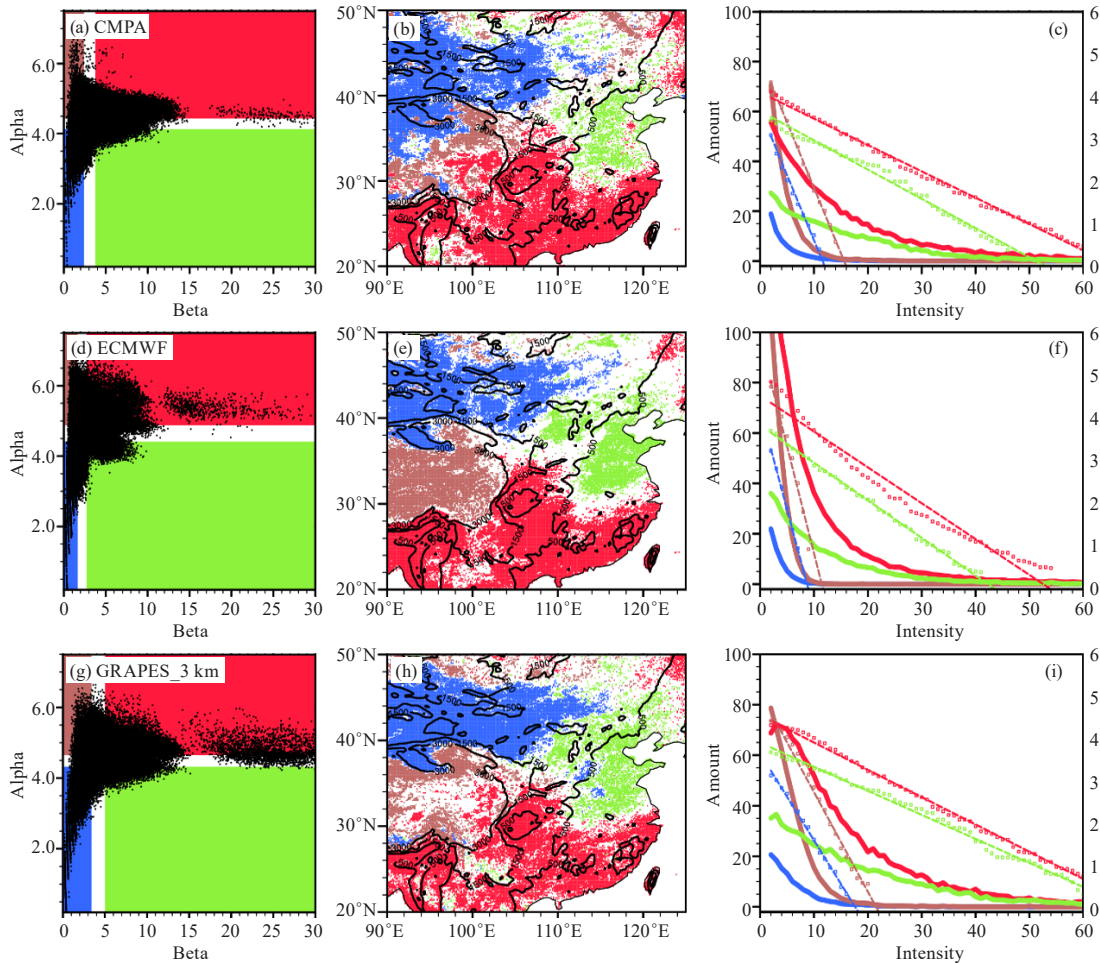


图3 CMA (a~c)、ECMWF (d~f) 和GRAPES-Meso-3 km (g~i) 模式中青藏高原以东地区(范围如图3b所示)每个格点的降水量对数随强度的回归参数值(α , β)在 α - β 平面上的分布(a, d, g)、四类分组的空间分布(b, e, h)和四类格点年平均降水量随强度的分布(c, f, i)

(图3a, 3d, 3g中黑色点表示回归参数值的分布;图3c, 3f, 3i中实线表示四类格点年平均降水量(单位: mm/3 h)的分布, 空心框和虚线分别表示年平均降水量取对数和基于最小二乘法线性回归后的拟合线;图3a~3i中四色背景表示四类分组, 其中蓝色(红色)表示弱降水和强降水的占比均较低(高), 绿色(棕色)表示弱降水的频次低(高)而强降水的占比高(低))

Fig. 3 (a-c) CMA, (d-f) ECMWF, (g-i) GRAPES-Meso-3 km model, (a, d, g) the distribution of the regression parameters (α , β) of the logarithm amount with the hourly intensity of each grid point over the east of Tibet Plateau (regions as shown in b), (b, e, h) the spatial distribution of four groups, (c, f, i) the distribution of mean annual precipitation with intensity of four types of grid points

(The black dots in Fig.3a, 3d, 3g represent the distribution of regression parameter, the solid line in Fig. 3c, 3f, 3i is the distribution of mean annual precipitation (mm/a) with intensity (unit: mm/3 h), the hollow box and dotted line in Fig. 3c, 3f, 3i are the fitting lines of logarithm of annual precipitation and linear regression based on least square method respectively, and the four color backgrounds in (a-i) indicate four groups, in which blue (red) indicates that the contribution of weak precipitation and strong precipitation is low (high), while green (brown) indicates that the contribution of weak precipitation is low (high) and that of strong precipitation is high (low))

强度($I-1$)~ I 的累积降水量。对式(3)取对数可得到对数降水量随降水强度的线性变化, 当降水强度不超过95%时, 可通过最小二乘法拟合得到回归系数 α 和 β (式(4)), $\alpha(\beta)$ 可定量表征弱(强)降水的对总降水量的相对贡献, $\alpha(\beta)$ 越大表示弱(强)降水的相对占比越大。

$$A(I) = \exp\left(\alpha - \frac{1}{\beta} I\right), \quad (3)$$

$$\ln[A(I)] = \alpha - \frac{1}{\beta} I. \quad (4)$$

图3a~3c示例给出了2019年5—9月CMA产品在青藏高原以东地区平均的不同强度降水量按照公式(3)拟合后的结果, 以说明拟合系数 α 和 β 表征的意

义。为比较不同 α 和 β 代表的强度结构差异的区域分布特征,将 α 和 β 按其量值进行组合分类,按照 α 、 β 数值的35%和70%进行了分组(图3a),用不同颜色代表不同强度结构特征,分别将 α 和 β 大于(小于)70%(35%)的点定义为弱和强降水占主导(相对较小)的格点,而在35%~70%的点未进行统计。图3b给出了四类分组的分布。图3c给出四类分组的年平均降水量随不同降水强度的分布曲线和按照式(4)取对数拟合后的直线,可清楚地看出通过 α 和 β 定量区分出了不同的降水强度分布特征。对于 α 和 β 均较小(蓝线)和较大(红线)的格点,其总降水量分别小于和大于其他类的格点。对于 α 较大(小)而 β 较小(大)的格点,弱(强)降水的占比更大。基于这一定量指标,可以评估数值模式对不同强度降水的预报能力。

图3d~3f和3g~3i分别给出EC和GRAPES模式对不同降水强度拟合后的结果。可见,模式预报能合理呈现出华南和西南强、弱降水在30°N以南、1500 m高度以下地区占比都相对更多(红色),500 m地形等值线以东的江淮至黄淮流域强降水相对较多(绿色),青藏高原在3000 m高度以上以弱降水主导(棕色)等区域特征。棕色与红色区域对应于图3a、3d、3g中 α 值较大的区域,而绿色区域主要对应于 β 值较大的区域。模式均高估了35°N以北、二阶阶梯地形以西

至河套地区的相对强降水,CMPA中500 m等值线以西相对弱降水占比更高(棕色),而模式中在该区域存在较多强降水占比相对更高的格点(绿色)。从四类分组平均的降水量随不同强度分布的曲线及取对数拟合后的直线可见,EC模式中弱降水的降水量均明显偏大,这也与图3d中模式的 α 值整体偏大对应。同时模式中四组分类的降水量随强度递减的速率均较CMPA更快,这与模式对 β 整体的低估相对应。对比可见,GRAPES模式对于不同强度降水的分布均更接近于CMPA的结果。

降水日变化是地球系统中大气热力和动力过程对水循环过程综合影响的结果,已成为评估数值模式不确定性问题的重要环节或指标^[59-62]。图4为CMPA和模式预报的2019年暖季平均的降水量主峰值发生时间(北京时)。EC模式再现了东南地区和500 m等高线西侧高地地形区下午至傍晚的峰值,但峰值出现时间较观测均有提前,而GRAPES模式对于下午峰值时间的预报更接近CMPA。模式均能合理地再现四川盆地和华北500 m等高线东侧低地形区午夜至清晨的峰值特征。EC模式预报的华北夜间至清晨降水峰值区的范围偏大,且未合理预报出CMAP和GRAPES模式中沿地形梯度降水峰值时间自高地形向下游滞后的现象。

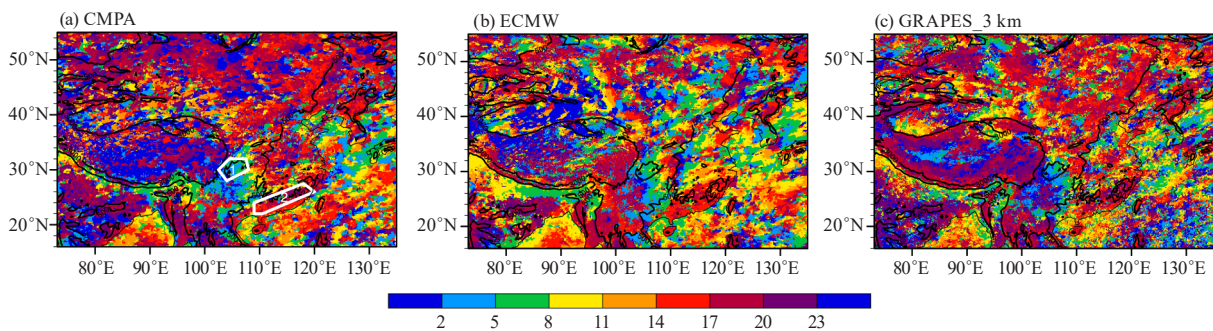


图4 2019年暖季(5—9月)CMPA(a)、ECMWF(b)和GRAPES-Meso-3 km(c)模式预报的降水量日变化主峰值发生时间(北京时)

(白色框标注CMPA和模式降水量日变化主峰值时间均出现在夜间至清晨(22:00—08:00)的西南地区(区域1)和下午至傍晚(14:00—20:00)的东南地区(区域2);黑色等值线为500 m和3000 m地形高度线)

Fig. 4 The time of the main diurnal peak for daily precipitation by CMPA (a), ECMWF (b) and GRAPES-Meso-3 km model (c) in the warm season (May-September) of 2019

(The white boxes indicate the southwest region (region 1) with the diurnal peak appears in the night to morning (22:00—08:00 BT) in both CMPA and model, and the southeast region (region 2) with the diurnal peak appears from afternoon to evening (14:00—20:00 BT), the black contour lines are 500 m and 3000 m terrain heights)

图5以CMPA和模式降水日峰值时间预报接近的西南(图4a中白色实线标注的区域1)和华南(区域2)这两个降水日变化振幅较大的区域(图略)为例,给出CMPA和模式预报的区域平均降水日变化曲线。在以夜间峰值为主的西南地区,不同模式均合理地再现

了凌晨的降水峰值,但EC模式高估了下午时段的降水频次,即EC模式中虚假的高频弱降水在下午时段更为突出。GRAPES模式对西南地区各时间段的降水均有高估,但在夜间时段降水强度高估的特征更明显。在以下午峰值为主的东南区域,尽管EC模式各时段

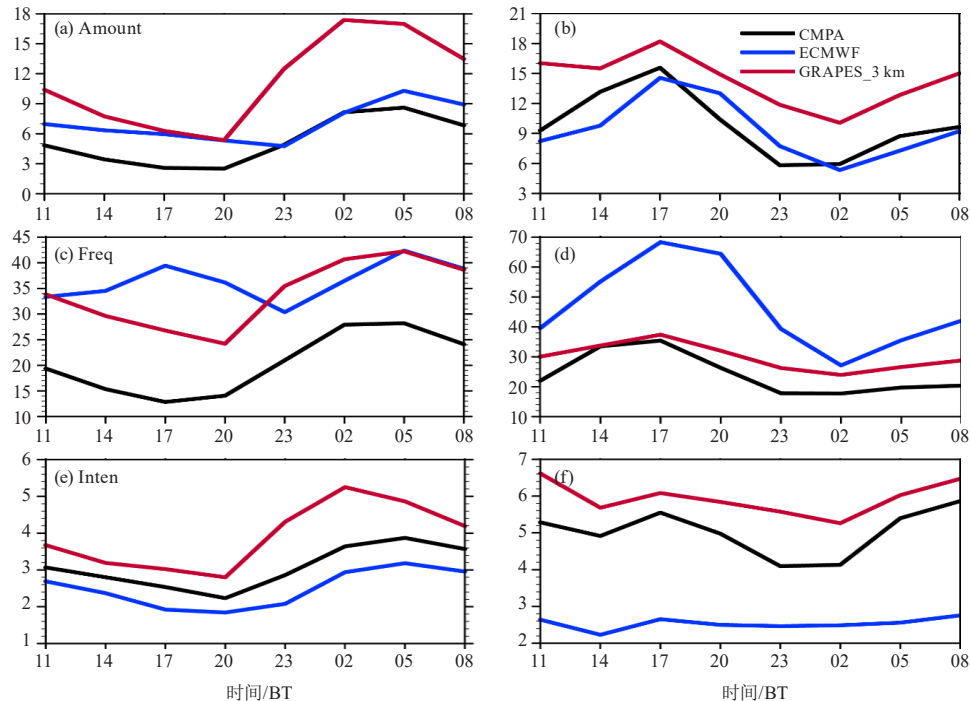


图5 图4a中白框标注的西南地区(区域1, a, c, e)和东南地区(区域2, b, d, f)平均的降水量(a, b, 单位: mm/d)、频率(c, d, 单位: %)、强度(e, f, 单位: mm/3h)的日变化曲线

Fig. 5 The diurnal curves of precipitation (a, b, unit: mm/d), frequency (c, d, unit: %), intensity (e, f, unit: mm/3 h) averaged over southwest and southeast (as marked by white boxes in Fig. 4a) regions

降水量与CMPA更为接近, 但该模式高估频率、低估强度的特征在各时段均较为突出。GRAPES模式对下午时段降水量的高估主要源于对强度的高估, 而夜间时段对降水频率和强度的高估在总降水的高估中均有贡献。

上述示例评估结果表明, 基于降水频次、降水强度等小时尺度特征量及其日变化峰值时间的分布有助于深入认识日内降水精细化预报偏差。但是如何合理评估数值预报模式的精细化预报能力, 目前还缺乏系统、完善的评估标准。基于本节所用的评估指标(小时频次/强度、日变化振幅与峰值时间), 中国气象局首次建立了应用于数值预报模式的业务评估指标体系, 将在未来的业务试运行中不断改进完善。

3 集合预报的检验评估

对流可分辨尺度的集合预报是区域高分辨率模式重要应用^[63]。研究对流可分辨尺度的数值模式预报偏差的起源、演变及影响, 合理、定量表征对流分辨尺度的模式不确定性, 提高对千米尺度模式偏差和模式多成员间差异的诊断和理解对于对流可分辨尺度的集合预报技术尤为重要。多成员集合概率预报相较于以往的统计概率预报, 是一种更具有动力学意义的概率预报。针对集合预报整体预报性能的检验评估, 当

前大多采用Brier Score等两分类评分来评估数值模式的整体预报性能^[64], 在此基础上, 通过相对工作特征(Relative Operating Characteristic, ROC)检验曲线考察集合预报命中率与空报率的预报性能。同时, 针对集合预报中多成员的可靠性与发散程度的检验评估, 依据集合内不同预报成员“准确率等同性”原则^[65-66], 可采用Talagrand直方图衡量各集合预报成员与观测的概率分布差异, 通过考察Talagrand分布直方图(呈现“U”型或“拱形”), 从而检验集合成员发散程度(不足或偏大)。

当前围绕强降水集合预报的相关科学研究, 大多依据各成员对于某次强降水过程累积降水量的空间分布和强度预报性能的优劣, 将不同成员分类并进行合成分析, 通过比较两组成员对于强降水相关的边界层过程、环流场、温湿场等的异同, 探究影响强降水的关键物理过程并分析预报偏差的产生原因。随着集合预报的兴起和广泛应用, 针对集合预报的检验评估也在不断改进与完善: 在常规技巧评分(如传统的两分类评分BS、TS、ETS等)的基础上, 一系列空间检验方法(如FSS), 基于对象、面向物理过程等新型诊断评估方法(如MODE、MODE-TD)得以发展并应用其中。与此同时, 概率预报检验技术得到了新的发

展,诸如可靠性、锐度和解析度等检验指标被用于评价集合预报性能。

4 面向天气-气候一体化模式无缝隙预报性能的检验评估

部分国际前沿研究已经实现利用大气模式进行全球对流可分辨的月尺度稳定积分^[67-69]。针对大气模式的研发而言,将天气和气候模式彼此衔接,实现天气和气候模式动力框架的统一,进而发展具备从天气到气候无缝隙预报能力的数值预报系统是当前重要的国际态势。而对一体化模式无缝隙预报系统开展客观、系统的检验评估,是一体化模式发展与产品应用的重要基础。

天气-气候一体化模式的目标是准确描述大气的多尺度结构特征。因此,围绕一体化模式的检验评估方法,有必要从当前的只关注最终统计结果,深入到分析影响基础变量(如温度、降水)动力、热力过程的面向物理过程的评估。在评估对象方面,从时间、空间平均的传统评估过渡到客观分类、分型的基于特定对象(Object-Based)、面向具体物理过程(Process-Oriented)的评估^[53]。

具体而言,一是从基于特定对象的角度出发,提出可连续描述覆盖天气-气候典型预报对象的无缝隙客观评价方法,筛选覆盖从天气到气候时间尺度的一组典型预报对象,通过考察一体化模式对于该组内涵盖天气-气候不同时空尺度典型预报对象的预报性能,检验其对天气-气候的无缝隙预报性能。二是从物理过程的角度入手,通过选取特定的判据集,从观测资料和预报结果的每个时间步长针对降水和气温中的典型预报对象进行甄别与追踪;首先在天气尺度从时空连续的角度描述降水和气温中典型预报对象的强度、范围、持续时间、移动速度等在时空尺度的精细化特性及演变特征,进而在气候尺度考察该典型预报对象的气候统计特征,揭示在我国不同关键区域,降水和气温预报技巧/偏差在天气过程到气候统计特征之间的关系。三是可从针对无缝隙预报流程衔接段的连续性和准确性进行检验评估;由于当前无缝隙预报流程的设计需要权衡预报性能与计算效率之间的关系,会不可避免地采用变分辨率策略;因此,检验与评估无缝隙预报流程中会涉及到诸如变分辨率衔接等技术方案,以及对一体化模式整体性能平稳性和对预报对象的积分连续性的定量影响。

总体而言,面向天气-气候一体化模式的检验评估,不仅要考察一体化模式在气候尺度对于时空平均后气候态的模拟技巧,同时也要面向物理过程考察一

体化模式在天气尺度对于典型预报对象精细化特性及其演变特征的预报性能,检验评估一体化模式“良好的预报结果”是否源自“正确的物理过程”。

5 讨论与展望

目前对于千米尺度对流分辨模式的业务检验仍以传统的二分类等评分方法为主,此类方法在高分辨率模式检验中面临着“双重惩罚”的问题,且难以给出模式偏差的有效诊断信息。近年来,尽管面向高分辨率模式的空间检验方法得以迅速发展,模糊检验方法(如邻域法)、目标对象法等新型检验方法虽然能够部分解决传统检验方法的局限性问题,但是也带来了新的问题,如信息量过多,针对性不够强,不能突出重点等。

本文在回顾数值预报传统检验和近年来新发展的空间检验基础上,给出了新的基于降水小时尺度的检验评估指标。通过示例分析表明,针对小时降水频次、降水强度的对比可以揭示模式对于日内降水特性的预报偏差,而针对日变化特征的评估揭示了模式对气候平均的峰值降水时间的预报能力,在一定程度上反映了模式动力、热力因子偏差的综合影响。此类检验评估结果对丰富和深化数值模式评估和对改进客观预报方法具有重要的参考价值。值得注意的是,此类方法虽然较已有的业务模式评估分析有很大的拓展和细化,但仍然属于通用性的基础评估分析。如何合理评估高时空分辨率数值预报模式的精细化预报能力,目前还缺乏系统完善的评估标准,需要推动新评估方法的研发和推广。

区域对流可分辨尺度模式在天气雷达等高分辨率、高频次观测资料的应用等方面具有不可替代的作用,基于千米尺度模式的快速更新循环系统能够更高效地同化利用实时观测资料,在强对流等高影响天气的短时临近预报预警等业务方面起着重要作用。针对快速更新同化系统的检验评估,如何客观、定量地检验评估不同观测资料快速协同同化的效果及不同起报时间预报的差异,仍面临较大的困难。

受东亚季风和多尺度地形与海陆分布的复杂影响,我国不同区域大气受到的区域强迫差异很大,以降水日变化为表征的降水精细化演变过程和机理在不同区域存在明显不同,数值模式对降水过程的模拟和预报偏差在不同区域也差异显著,模式对于不同降水过程的预报能力差异明显,针对关注区域开展不同类型降水过程的细致评估是提升区域模式应用水平的关键。所以,未来针对降水预报的细致检验评估必须考虑到不同区域强降水过程的差异。基于对关注区域

细化分类的精细过程的全面把握,细致对比分析所用的数值模式或客观预报系统,形成基于特征环流结构或要素关联特征的细化分类的降水和相关要素的评估指标体系,才能系统评估、客观预报系统在具体类别的偏差特征分布。同时,数值预报的评估也将从时间、空间平均的传统评估,过渡到基于特定对象、面向物理过程的客观分类、分型评估以及面向高影响天气过程(如强降水、冰雹、大风)的空间位置、强度、过程形态差异和日内尺度的演变特征等综合定量化检验评估,为对流可分辨尺度模式预报能力的改进及其预报产品的订正与应用提供更有意义的参考。

随着气象科学、高性能计算和观测系统的发展,对流可分辨尺度数值预报模式已在精细化预报业务中得以广泛应用,在智能网格预报业务中发挥着不可替代的作用。尽管对流可分辨尺度模式可以显式表征深对流过程,从而减小了对流参数化方案对模式不确定性的影响,但受限于边界层过程和云物理过程等方案的不足,模式同样呈现出长期存在的不完整性和不精准性,而强化对千米尺度数值模式结果等的不确定性问题的深入评估分析,是完善模式性能、改进模式结果和提高模式结果应用水平的切入点和突破口。数值预报的偏差源于诸多因素的综合非线性作用结果,提升千米尺度模式在降水精细化预报中的应用能力,需要通过强化对数值模式结果等的不确定性问题的深入评估分析,从而减小其结果的不确定性,实现对现有方法的不断改进和完善,得到更高质量的业务产品。Yu等^[1]围绕减小数值模式结果的不确定性影响,给出了通过深化评估践行“自上而下”的发展途径,指出充分完备的观测和模式数据是必须的基础支撑,同时提出对现有客观预报方法所存在的问题首先要通过基本评估得到基本认知的理念;而更重要的是针对现有客观方法所存在的问题,在充分熟悉已有科学成果的基础上,从数据中去获取更加深入和更加细化的科学认知,以此建立针对性的、细致深入的评估体系和创新评估方法,得到改进和订正现有“方法”的有效途径,实现现有“方法”的不断改进和完善。

参考文献

- [1] Yu R C, Jian L, Jia P Q. Development of operational weather forecasting shaped by the “Triple-In” properties of numerical models. *WMO Bulletin*, 2019, 68(2): 56-62.
- [2] 张小玲, 杨波, 盛杰, 等. 中国强对流天气天气预报业务发展. *气象科技进展*, 2018, 8(3): 8-18.
- [3] 金荣花, 代刊, 赵瑞霞, 等. 我国无缝隙精细化网格天气预报技术进步与挑战. *气象*, 2019, 45(4): 445-457.
- [4] 王毅, 周庆亮, 代刊, 等. 全球数据处理和预报系统发展及展望. *气象科技进展*, 2019, 9(2): 9-10.
- [5] 黄丽萍, 陈德辉, 邓莲堂, 等. GRAPES Meso V4.0主要技术改进和预报效果检验. *应用气象学报*, 2017, 28(1): 25-37.
- [6] 许晨曦, 王建捷, 黄丽萍. 千米尺度分辨率下GRAPES-Meso4.0模式定量降水预报性能评估. *气象学报*, 2017(6): 851-876.
- [7] 范水勇, 卢冰, 仲跻芹, 等. 华北区域多尺度分析和短临数值预报系统RMAPS-ST介绍. 中国气象学会. 第33届中国气象学会年会S10城市、降水与雾霾——第五届城市气象论坛. 中国气象学会: 中国气象学会, 2016: 120-121.
- [8] 卢冰, 孙继松, 仲跻芹, 等. 区域数值预报系统在北京地区的降水日变化预报偏差特征及成因分析. *气象学报*, 2017, (2): 58-69.
- [9] 徐同, 李佳, 杨玉华, 等. SMS-WARMS V2.0模式预报效果检验. *气象*, 2016, 42(10): 1176-1183.
- [10] 徐道生, 张艳霞, 张诚忠, 等. 华南区域高分辨率模式中不同雷达回波反演技术方案的比较试验. *热带气象学报*, 2016, 32(1): 9-18.
- [11] Prein A F, Langhans W, Fossier G, et al. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 2015, 53(2): 323-361.
- [12] Emanuel K A. A scheme for representing cumulus convection in large-scale models. *Journal of the Atmospheric Sciences*, 1991, 48(21): 2313-2329.
- [13] Guichard F, Petch J C, Redelsperger J L, et al. Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Quarterly Journal of the Royal Meteorological Society*, 2004, 130(604): 3139-3172.
- [14] Rooy W C, Bechtold P, Fröhlich K, et al. Entrainment and detrainment in cumulus convection: An overview. *Quarterly Journal of the Royal Meteorological Society*, 2013, 139(670): 1-19.
- [15] Liang X Z, Li L, Dai A, et al. Regional climate model simulation of summer precipitation diurnal cycle over the United States. *Geophysical Research Letters*, 2004, 31(24): L24208.
- [16] Dai A. Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, 2006, 19(18): 4605-4630.
- [17] Zhang Y, Chen H. Comparing CAM5 and superparameterized CAM5 simulations of summer precipitation characteristics over continental East Asia: Mean state, frequency-intensity relationship, diurnal cycle, and influencing factors. *Journal of Climate*, 2016, 29(3): 1067-1089.
- [18] Ban N, Schmidli J, Schär C. Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, 2014, 119(13): 7889-7907.
- [19] Clark P, Roberts N, Lean H, et al. Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 2016, 23(2): 165-181.
- [20] Jin Q, Yang X Q, Sun X G, et al. East Asian summer monsoon circulation structure controlled by feedback of condensational heating. *Climate Dynamics*, 2013, 41(7-8): 1885-1897.
- [21] Guo Z, Zhou T, Wang M, et al. Impact of cloud radiative heating on East Asian summer monsoon circulation. *Environmental Research Letters*, 2015, 10(7): 074014.
- [22] 周天军, 吴波, 郭准, 等. 东亚夏季季风变化机理的模拟和未来变化的预估: 成绩和问题, 机遇和挑战. *大气科学*, 2018, 42(4): 902-934.
- [23] 陶诗言. 中国之暴雨, 北京: 科学出版社, 1980: 225.
- [24] 丁一汇. 中国暴雨理论的发展历程与重要进展. *暴雨灾害*, 2019, 38(5): 395-406.
- [25] 罗亚丽, 孙继松, 李英, 等. 中国暴雨的科学及预报: 改革开放40年研究成果. *气象学报*, 2020, 78(3): 419-450.
- [26] 宗志平, 陈涛, 徐珺, 等. 2012年初秋四川盆地两次西南涡暴雨过程的对比分析与预报检验. *气象*, 2013, 39(5): 567-576.
- [27] 官宇, 代刊, 徐珺, 等. GRAPES-GFS模式暴雨预报天气学检验特征. *气象*, 2018, 44(9): 34-45.
- [28] Schaefer, Joseph T. The critical success index as an indicator of warning skill. *Weather and Forecasting*, 1990, 5(4): 570-575.
- [29] Casati D B, Wilson L J, Stephenson D B, et al. Forecast verification: current status and future directions. *Meteorological Applications*, 2008, 15(1): 3-18.
- [30] Rodwell M J, Richardson D S, Hewson T D, et al. A new equitable score suitable for verifying precipitation in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 2010, 136(650): 1344-1363.

- [31] Baldwin M E, Kain J S. Sensitivity of several performance measures to displacement error, bias, and event frequency. *Weather and Forecasting*, 2006, 21(4): 636-648.
- [32] 陈静, 刘凑华, 陈法敬, 等. 一种基于可预报性的暴雨预报评分新方法 I: 中国暴雨可预报性综合指数. *气象学报*, 2019, 77(1): 15-27.
- [33] 陈法敬, 陈静, 韦青, 等. 一种基于可预报性的暴雨预报评分新方法 II: 暴雨检验评分模型及评估试验. *气象学报*, 2019, 77(1): 28-42.
- [34] Mass C F, Owens D, Westrick K, et al. Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bulletin of the American Meteorological Society*, 2002, 83(3): 407-430.
- [35] Ebert E, Wilson L, Weigel A, et al. Progress and challenges in forecast verification. *Meteorological Applications*, 2013, 20(2): 130-139.
- [36] Gilleland E, Ahijevych D A, Brown B G, et al. Verifying forecasts spatially. *Bulletin of the American Meteorological Society*, 2010, 91(10): 1365-1376.
- [37] Dorninger M, Gilleland E, Casati B, et al. The setup of the MesoVICT project. *Bulletin of the American Meteorological Society*, 2018, 99(9): 1887-1906.
- [38] Ebert E E. Neighborhood verification: A strategy for rewarding close forecasts. *Weather and Forecasting*, 2009, 24(6): 1498-1510.
- [39] Mittermaier M, Roberts N. Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Weather and Forecasting*, 2010, 25(1): 343-354.
- [40] Harris D, Foufoula-Georgiou E, Droegemeier K K, et al. Multiscale statistical properties of a high-resolution precipitation forecast. *Journal of Hydrometeorology*, 2001, 2(4): 406-418.
- [41] Casati B. New developments of the intensity-scale technique within the Spatial Verification Methods Intercomparison Project. *Weather and Forecasting*, 2010, 25(1): 113-143.
- [42] Keil C, Craig G C. A displacement-based error measure applied in a regional ensemble forecasting system. *Monthly Weather Review*, 2007, 135(9): 3248-3259.
- [43] Marzban C, Sandgathe S, Lyons H, et al. Three spatial verification techniques: Cluster analysis, variogram, and optical flow. *Weather and Forecasting*, 2009, 24(6): 1457-1471.
- [44] Davis C, Brown B, Bullock R. Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, 2006, 134(7): 1772-1784.
- [45] Davis C, Brown B, Bullock R. Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Monthly Weather Review*, 2006, 134(7): 1785-1795.
- [46] Ebert E E, Gallus Jr W A. Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. *Weather and Forecasting*, 2009, 24(5): 1401-1415.
- [47] Roberts N M, Lean H W. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 2008, 136(1): 78-97.
- [48] Zepeda-Arce J, Foufoula-Georgiou E, Droegemeier K K. Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *Journal of Geophysical Research: Atmospheres*, 2000, 105(D8): 10129-10146.
- [49] Clark A J, Bullock R G, Jensen T L, et al. Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. *Weather and Forecasting*, 2014, 29(3): 517-542.
- [50] Prein A F, Liu C, Ikeda K, et al. Simulating North American mesoscale convective systems with a convection-permitting climate model. *Climate Dynamics*, 2020, 55: 95-110.
- [51] Klingaman N P, Martin G M, Moise A. ASoP (v1.0): A set of methods for analyzing scales of precipitation in general circulation models. *Geoscientific Model Development*, 2017, 10(1): 57-83.
- [52] Martin G M, Klingaman N P, Moise A F. Connecting spatial and temporal scales of tropical precipitation in observations and the MetUM-GA6. *Geoscientific Model Development*, 2017, 10(1): 105-126.
- [53] Maloney E D, Gettelman A, Ming Y, et al. Process-oriented evaluation of climate and weather forecasting models. *Bulletin of the American Meteorological Society*, 2019, 100(9): 1665-1686.
- [54] Yu R, Zhang Y, Wang J, et al. Recent progress in numerical atmospheric modeling in China. *Advances in Atmospheric Sciences*, 2019, 36(9): 938-960.
- [55] Lin X, Randall D A, Fowler L D. Diurnal variability of the hydrologic cycle and radiative fluxes: Comparisons between observations and a GCM. *Journal of Climate*, 2000, 13(23): 4159-4179.
- [56] Trenberth K E, Dai A, Rasmussen R M, et al. The changing character of precipitation. *Bulletin of the American Meteorological Society*, 2003, 84(9): 1205-1218.
- [57] Zhou T, Yu R, Chen H, et al. Summer precipitation frequency, intensity, and diurnal cycle over China: A comparison of satellite data with rain gauge observations. *Journal of Climate*, 2008, 21(16): 3997-4010.
- [58] Yu R, Li J. Hourly rainfall changes in response to surface air temperature over Eastern Contiguous China. *Journal of climate*, 2012, 25(19): 6851-6861.
- [59] Dai A, Trenberth K E. The diurnal cycle and its depiction in the Community Climate System Model. *Journal of Climate*, 2004, 17(5): 930-951.
- [60] Li J, Yu R, Yuan W, et al. Precipitation over East Asia simulated by NCAR CAM5 at different horizontal resolutions. *Journal of Advances in Modeling Earth Systems*, 2015, 7(2): 774-790.
- [61] Yu R, Li J, Chen H, et al. Progress in studies of the precipitation diurnal variation over contiguous China. *Journal of Meteorological Research*, 2014, 28(5): 877-902.
- [62] Yuan W, Yu R, Zhang M, et al. Diurnal cycle of summer precipitation over subtropical East Asia in CAM5. *Journal of Climate*, 2013, 26(10): 3159-3172.
- [63] Roberts B, Jirak I L, Clark A J, et al. Postprocessing and visualization techniques for convection-allowing ensembles. *Bulletin of the American Meteorological Society*, 2019, 100(7): 1245-1258.
- [64] Hersbach H. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 2000, 15(5): 559-570.
- [65] Hamill T M, Colucci S J. Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, 1997, 125(6): 1312-1327.
- [66] Hamill T M. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 2001, 129(3): 550-560.
- [67] Satoh M, Stevens B, Judt F, et al. Global cloud-resolving models. *Current Climate Change Reports*, 2019, 5(3): 172-184.
- [68] Stevens B, Satoh M, Auger L, et al. DYAMOND: The dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, 2019, 6(1): 61.
- [69] Stevens B, Acquistapace C, Hansen A, et al. The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan*, 2020, 98(2): 395-435.