

张小雯,唐文苑,郑永光,等,2020. GRAPES\_3 km 数值模式对流风暴预报能力的多方法综合评估[J]. 气象,46(3):367-380.  
Zhang X W, Tang W Y, Zheng Y G, et al, 2020. Comprehensive evaluations of GRAPES\_3 km numerical model in forecasting convective storms using various verification methods[J]. Meteor Mon, 46(3):367-380(in Chinese).

# GRAPES\_3 km 数值模式对流风暴预报能力的 多方法综合评估\*

张小雯 唐文苑 郑永光 盛 杰 朱文剑

国家气象中心,北京 100081

**提 要:** 利用传统点对点 TS 评分、邻域法以及对象检验等多种方法,综合评估了 GRAPES\_3 km 模式的对流风暴预报性能,分析了传统检验方法和新型空间检验方法对高分辨率模式评估的适用性和差异性,并同 GRAPES\_Meso 模式的相关结果进行了对比。结果表明:对强对流典型个例的预报评估发现,综合应用多种评估方法能够更全面地评估对流风暴预报,获取模式在对流风暴初生和发展变化过程中的预报性能。使用点对点评分方法,GRAPES\_3 km 模式对风暴和强风暴的预报都明显优于 GRAPES\_Meso 模式。对于模式不同起报时间的预报,起报时间越新预报效果越好。邻域 TS 方法考虑了时空偏差,GRAPES\_3 km 模式 20 和 35 dBz 采用时间邻域 1 h,空间点对点时预报技巧最高;50 dBz 时空偏差较大,时间邻域尺度为 3 h 技巧最高。分数技巧评分(FSS)显示 GRAPES\_3 km 模式对不同阈值的对流风暴预报均能达到最低技巧尺度,而 GRAPES\_Meso 模式对 35 dBz 以上的对流风暴基本无预报能力。对象检验可以评估对流风暴特征的预报效果,GRAPES\_3 km 模式的对流风暴个数预报与实况较为一致,但面积预报明显低估。该模式对  $\beta$  中尺度的对流风暴形态、位置等预报较好,对  $\gamma$  中尺度的对流风暴预报尺度偏大、形状偏圆、轴角偏小,对  $\alpha$  中尺度的对流风暴预报尺度偏小、形状偏扁、轴角偏大。GRAPES\_Meso 模式的对流风暴面积、个数、尺度预报较实况均偏小,位置预报偏差较大,形状预报较实况偏圆、轴角偏小。传统点对点 TS 评分方法和新型空间检验方法对高分辨率模式对流风暴预报的检验结论一致,依然具有一定的参考价值,但新型空间检验方法能够提供更详细的评估信息。

**关键词:** 对流许可尺度模式,GRAPES,时空尺度分析,邻域法,对象检验,分数技巧评分

中图分类号: P456

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2020.03.008

## Comprehensive Evaluations of GRAPES\_3 km Numerical Model in Forecasting Convective Storms Using Various Verification Methods

ZHANG Xiaowen TANG Wenyuan ZHENG Yongguang SHENG Jie ZHU Wenjian

National Meteorological Centre, Beijing 100081

**Abstract:** GRAPES\_3 km (Global/ Regional Assimilation and Prediction System) is a convection-permitting model, which provides an important objective basis for short-term forecast of severe weather. In this paper, the performance of GRAPES\_3 km model in severe weather forecasting was comprehensively evaluated using traditional pixel-versus-pixel threat score, neighborhood and object-based methods. The applicability and differences of the traditional and new spatial verification methods for high resolution model assessment were analyzed, and the results were compared with those of GRAPES\_Meso model. The results are showed as follows. Through case analysis, it is found that the characteristics of convective storm forecast and the evolution of convection can be comprehensively and objectively evaluated by various verification

\* 国家重点研发计划(2018YFC1507504 和 2017YFC1502003)共同资助

2019 年 4 月 22 日收稿; 2019 年 9 月 13 日收修定稿

第一作者:张小雯,主要从事强对流天气研究. Email: zhangxw@cma.gov.cn

methods. GRAPES\_3 km is superior to GRAPES\_Meso in forecasting convective storms, especially severe storms over 50 dBz. The latest forecast of the initial forecast time is the best. Neighborhood TS method takes space-time deviation into account. The forecasting skill is the highest when the time neighborhood of GRAPES\_3 km model is 1 h for 20 dBz and 35 dBz, and 3 h for 50 dBz. The fractions skill score (FSS) shows that GRAPES\_3 km model can achieve the lowest forecast skill scale for convective storms with different thresholds, while GRAPES\_Meso model usually fails to reach the lowest skill scale for storms above 35 dBz. Method for object-based diagnostic evaluation (MODE) can be used to evaluate the forecast of convective storm attributes. GRAPES\_3 km model is consistent with the actual number of storms at all scales, but the area is obviously underestimated. The model can predict the shape and location of meso- $\beta$  scale convective storms, while for meso- $\gamma$  scale convective storms, the forecast scale is larger, the shape is more circular and the axis angle is smaller, but conversely for meso- $\alpha$  scale convective storms. The traditional pixel-versus-pixel verification method and new spatial verification methods have the same conclusion for convection-permitting model. Comparatively, the new spatial verification methods can provide more detailed information of convective storms.

**Key words:** convection-permitting model (CPM), GRAPES, spatio-temporal scale analysis, neighborhood based method, method for object-based diagnostic evaluation (MODE), fractions skill score (FSS)

## 引 言

强对流天气时空尺度小、发展快、致灾性强,准确预报其落区和强度可最大限度减少生命财产损失。传统的强对流天气预报是根据对流发展的环境条件,使用全球或者区域数值模式预报的对流参数来预报(薛纪善,2006),无法提供强对流天气的精确信息(对流形态、对流移动速度、对流强度等)。随着计算能力的提高,对流许可尺度数值模式 CPMs (convection-permitting models,通常认为水平格距小于 5 km)已成为强对流天气预报中不可或缺的工具,如美国、法国、英国、日本等气象局目前使用的区域中尺度模式空间分辨率达到了 1~4 km,可以直接预报  $\gamma$  中尺度对流风暴。CPMs 在中尺度对流系统的回波形态、对流组织性、发生频率等方面明显优于采用对流参数化方案的低分辨率(格距 10~20 km)数值模式(Weisman et al,2008;陈德辉和沈学顺,2006),能够为预报员提供直接的对流尺度产品(如雷达反射率因子、10 m 瞬时大风等),还可计算闪电、雷暴、冰雹等对流的强度和特征(McCaul et al,2009),为预报员判断强对流天气的类型和强度提供了更为直接的指导产品。

GRAPES\_3 km 模式作为我国自主研发的对流

许可尺度中尺度数值模式,许晨璐等(2017)评估了该模式的降水预报,指出模式夏季日降水频率随降水量的变化与观测相似程度高,大降水频率和区域分布和实况较为吻合,并能够反映降水过程的日变化特征。唐文苑等(2018)利用分数技巧评分(FSS)对比评估了该模式和华东中尺度模式的组合反射率预报在华北地区几次典型强对流天气过程中的预报能力,探讨了该模式在强对流天气中良好的预报性能,但尚未利用多种检验方法对其进行全面评估。因此,本文以 GRAPES\_3 km 模式的对流风暴预报为评估对象,并同 GRAPES\_Meso 模式的预报评估结果对比,分析不同分辨率中尺度模式在强对流天气中的预报能力,从而为基于该类模式预报产品进一步发展强对流天气的智能网格预报技术和产品提供参考依据。

## 1 观测资料和模式资料

评估所用的观测资料为全国雷达组合反射率数据,空间分辨率为  $0.01^\circ$ ,时间分辨率为 6 min,区域范围为  $12.2^\circ\sim 54.2^\circ\text{N}$ 、 $70^\circ\sim 135^\circ\text{E}$ 。模式资料来源于 GRAPES\_3 km 和 GRAPES\_Meso 两个中尺度模式,两者差异主要体现在:(1)预报区域和空间分辨率不同(徐枝芳等,2013;黄丽萍等,2017);(2)起

报时间和预报时效不同;(3)模式初值和背景场不同(毛冬艳等,2014)。具体说明见表 1。

表 1 模式系统对比  
Table 1 NWP model systems

	GRAPES_3 km	GRAPES_Meso
预报区域	17°~50°N,102°~135°E	15°~65°N,70°~145°E
空间分辨率	0.03°	0.1°
	冷启动	热启动
时间分辨率	08/20 时起报 逐小时输出	隔 3 h 预报起报 逐小时输出
预报时效	36 h	72 h
模式背景场	T639 模式同时刻分析场	T639 模式 6 h 预报场
模式初值	降尺度 云分析	三维变分同化(3DVAR)
积云参数化	无	KF_ETA

由于模式预报与实况观测数据时空分辨率不同,统一空间分辨率时将观测资料进行了稀疏化,但为保留对流天气的强回波信息,取所稀疏网格周边区域的最大值。

## 2 检验方法和检验对象

### 2.1 检验对象

以对流风暴作为检验对象来评估 GRAPES\_3 km 模式对流天气的预报能力。对流风暴在雷达观测中定义为组合反射率超过一定阈值的具有一定体积的连续区域。本文评估时将其划分为三个等级(Dixon and Wiener,1993):对流风暴外围的层状云带、对流风暴及强对流风暴;其最低阈值分别定义为 20、35 和 50 dBz。

首先使用点对点 TS 评分对两个模式的逐小时预报进行整体评估,检验区域为我国中东部地区(17°~50°N、102°~135°E),检验时段为 2018 年 4—8 月。由于 GRAPES\_3 km 模式为一日两次的预报,GRAPES\_Meso 模式为一日 8 次滚动预报,评估时采用两者相同的起报时间(每日 08 时和 20 时)进行对比评估。

应用新型检验方法进行评估时,为使结果更有代表性,选取了 10 次强对流天气过程(见表 2),预报区域包含了华南、江汉江淮、华北黄淮等强对流较为多发的区域,检验时段为强对流个例发展较为强盛的时段。模式评估对象为:对流发生时,GRAPES\_3 km 模式前一日 20 时和当日 08 时起报的预报;

GRAPES\_Meso 模式最临近(最新)起报和前一时次(次临近)起报的预报。通过两个模式不同起报时间的预报效果对比,评估对流许可尺度模式和快速更新同化模式在对流风暴预报中的性能。

表 2 2018 年强对流天气个例列表

Table 2 List of severe weather events in 2018			
日期 (月/日)	范围	时段/BT	起报时间
5/9	华南(20°~28°N,105°~115°E)	15—20 时	GRAPES_
5/12	华北黄淮(36°~43°N,110°~118°E)	16—21 时	3 km 模式前
5/16	华北黄淮(30°~38°N,115°~122°E)	11—16 时	一日 20 时
5/18	江汉江淮(28°~33°N,109°~122°E)	12—17 时	起报、当日
5/19	江南(25°~30°N,113°~120°E)	11—16 时	08 时起报、
5/21	江汉江淮(27°~33°N,110°~118°E)	21—02 时	GRAPES_
5/6	江南(27°~33°N,110°~118°E)	15—20 时	Meso 对于
3/19	华南(22°~26°N,106°~116°E)	16—21 时	该时段最临
3/4	江南(25°~31°N,109°~117°E)	12—17 时	近时刻、次
5/15	华北黄淮(34°~40°N,110°~118°E)	21—02 时	临近时刻的
			预报

### 2.2 检验方法

#### 2.2.1 传统点对点检验

TS 评分是基于二分类表,针对预报和观测发生和未发生有四种可能的情况计算模式与实况匹配的程度。传统点对点检验是对比预报格点和观测格点重合程度,给出总体评分。TS 评分综合考虑了命中( $a$ )、空报( $b$ )和漏报( $c$ )的情况,计算公式如下:

$$TS = \frac{a}{a+b+c} \quad (1)$$

#### 2.2.2 时空邻域 TS 检验

传统预报检验具有一定的参考意义,但是对于高分辨率模式,随着格点间距的减小,微小的位置和形态的差异在传统的评分会造成很大的影响(命中率和虚警率的双重惩罚)。邻域法解决了“双惩罚”现象(Ebert,2008),能够接受预报值和观测点位置上的细微差异。邻域法通常是指空间邻域法(Ghelli and Lalanrette,2000),是指将预报格点和以其为中心向外拓展的实况邻域进行的点对面评分。强对流天气发生发展迅速、局地性强,对流风暴在短时间内形态、强度均可能有比较大的变化,评估时使用预报整点的瞬时值和对应时刻的整点实况,可能无法客观评价对流在该时段内发展变化,因此需要引进时间邻域的概念,使用累积实况来体现该时间段强对流的影响区域和强度变化。

本文的邻域 TS 法同时引入时间邻域和空间邻域(简称为时空邻域 TS 法),具体算法是:以 0.0、0.4°、0.8°、1.2°、1.6°、2.0°、3.0°、4.0° 方形网格为空

间邻域,分别以雷达回波在  $\Delta t$ 、 $2\Delta t$ 、 $3\Delta t$  时段内超过某一阈值(如 35 dBz)的次数为时间邻域,计算 TS 评分。邻域 TS 法是传统点对点 TS 方法的延伸,当邻域尺度为 0 时,与点对点评分一致。利用邻域尺度可变的特性,时空邻域 TS 法可以得到单点预报时(比如机场预报等)模式预报的最优技巧尺度,这个尺度是向用户显示预报结果的合适尺度。

### 2.2.3 分数技巧评分

分数技巧评分(FSS)是基于邻域法给出空间技巧评估的另一个手段,比较预报邻域与相应观测邻域的差异,是“面对面”的检验,衡量场的相似度(Roberts and Lean,2008)。

$$FSS = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (P_{fcst} - P_{obs})^2}{\frac{1}{n} \sum_{i=1}^n P_{fcst}^2 + \frac{1}{n} \sum_{i=1}^n P_{obs}^2}$$

式中: $P_{fcst}$  和  $P_{obs}$  分别表示邻域内预报格点数和观测格点数占邻域窗口的比例; $n$  是窗口的个数,当 FSS 值趋向于 1 时预报技巧最高。

将“最低技巧尺度”定义为 FSS 值超过  $0.5 + 0.5f_{obs}$  的尺度,其中  $f_{obs}$  为研究区域内满足某一给定阈值的观测格点占整个区域的百分比。FSS 能够评估模式对于不同尺度对流风暴预报的位置偏差,并给出相应的最低预报技巧尺度。

### 2.2.4 对象检验

对象检验(MODE)也是“面对面”的检验(Davis et al,2006),通过识别和匹配不同空间场里的对象,计算对象的相似度并评估两个场的综合相似度(李佰平等,2016;茅懋等,2016;刘凑华和牛若云,2013)。该方法除了能够评估空间场的相似程度外,还可以检验对象的位置、面积、轴向、形状等特征。其有 5 个步骤:(1)卷积滤波;(2)聚类:将离散的小对流风暴合并成大对流风暴(Marzbán and Sandgathe,2006);(3)椭圆识别:对聚类后的对流风暴进行椭圆拟合(Rodrigue and Laio,2014);(4)匹配:根据面积膨胀法,取椭圆交集最大的配对;(5)评分:利用面积重叠比计算最后的评分值。

## 3 典型强对流个例检验

新型检验方法是否能够客观评估模式对流风暴的预报能力?以 2018 年 3 月 19 日的强对流过程为例进行探讨。该过程是一次较为典型的华南前汛期

由地面锋面触发的强对流天气,锋面呈“人”字型分布。午后到傍晚,广西多站观测到冰雹、8 级以上雷暴大风。

首先,检验 MODE 拟合的椭圆能否较为准确地代表对流风暴尺度、位置、轴角等基本特征,评估是否具有可信性?以 19 日 18 时实况为例(图 1),从预报员的主观检验来看,GRAPES\_3 km 模式前一日 20 时起报的预报位置偏东,强度偏弱,当日 08 时起报的预报最好,能够较为准确地预报对流风暴及其外围层状云带,但是对强对流风暴的形态预报(南北向)与实况(东西向)有一定的差异。GRAPES\_Meso 模式预报对流风暴的位置明显偏东、强度偏弱,对强对流风暴的预报基本漏报。

图 2 为不同聚类尺度的 MODE 评分,可以评估模式对不同尺度对流风暴的预报能力(Tustison et al,2001;Roberts and Lean,2008)。GRAPES\_3 km 模式的 MODE 评分明显高于 GRAPES\_Meso 模式,这也与预报员的主观判断一致。以 GRAPES\_3 km 模式 08 时起报的最优评估尺度(分别为  $0.2^\circ$ 、 $0.6^\circ$ 、 $3.0^\circ$ )拟合的椭圆空间分布(图 3)可见,MODE 识别和匹配的椭圆基本能反映对流风暴的尺度、形状等特征,可以用其作为模式评估的依据。

FSS 评估结论与 MODE 大致相同(图 4),GRAPES\_3 km 模式的 FSS 评分显著高于 GRAPES\_Meso 模式。GRAPES\_3 km 模式当日 08 时起报的预报最好,各阈值均达到了最低技巧尺度( $0.3^\circ$ 、 $0.8^\circ$ 、 $3.5^\circ$ ),而前一日 20 时起报的空间偏差较大,尤其是强风暴预报没有达到最低技巧尺度。GRAPES\_Meso 模式对于 35 dBz 以上的对流风暴均没有达到最低技巧尺度。

采用 MODE 拟合的风暴面积结合邻域 TS 评分和 FSS 的最低技巧尺度来分析对流风暴发展变化的预报能力。图 5a~5c 中黑色实线为基于 MODE 最优技巧尺度拟合的对流风暴面积实况。19 日午后对流风暴的面积快速增长,在 15—16 时和 21—22 时两个时段,面积达到峰值,呈现出双峰结构。GRAPES\_3 km 模式的对流风暴面积预报较为准确,但是对初生时间预报滞后,其中当日 08 时起报滞后 2 h 左右,对于午后到傍晚时段预报较好;前一日 20 时起报滞后 4 h,对于傍晚到前半夜的预报与实况更接近。GRAPES\_Meso 模式是快速更新同化模式,对流初生和发展的时段预报和实况较为一致,能够看到明显的双峰结构,但由于空间分辨

率较粗,对流风暴的面积预报明显偏小。

邻域 TS 评分和 FSS 可以评估模式对不同阶段的对流风暴发展变化的能力。对于 08—21 时时段, GRAPES\_3 km 模式当日 08 时起报的对流风暴 (35 dBz) 邻域 TS 评分高于其他预报, 其中 17 时 TS 评分最高, FSS 也显示该时刻最低技巧尺度最小 (0.3°), 位置预报偏差最小。对于 50 dBz 的强对流风暴, 评分具有明显的日变化特征, 模式在热力条件更好的午后到前半夜的刻画能力明显好于其他时段, 傍晚到前半夜评分最高, 其他时次评分基本为 0。但由于模式对强对流风暴的面积预报仅为实况面积的一半, 所以邻域 TS 评分最高仅为 0.07, FSS 未达到最低技巧尺度。对于 21—02 时时段, FSS 和邻域 TS 评分均显示 GRAPES\_3 km 模式前一日 20 时起报的对流风暴预报最好, 其中在 00 时时 TS

评分最高, 35 dBz 的邻域 TS 评分达到 0.25, 最低技巧尺度(空间偏差)为 0.2°左右。该起报时间的强对流风暴面积预报为实况的 1/5, 漏报较多, 因此邻域 TS 评分最大值仅为 0.06, FSS 也显示无预报技巧。GRAPES\_Meso 模式对流风暴预报的 TS 评分仅为 GRAPES\_3 km 模式的一半, 对强对流风暴无预报能力为 0(图 5f), FSS 评估该模式预报未达到最低技巧尺度(图略), 这可能与对流风暴面积预报偏小有关。

从本次强对流过程来看, GRAPES\_3 km 模式预报对流的初生时间滞后, 对流风暴的位置和面积预报较好。GRAPES\_Meso 模式能够较为准确地描述对流初生和发展的时间, 但是对对流风暴的强度、面积预报能力较差。预报员制作强对流天气预报时可以综合考虑两模式的特点对主观预报进行订正。

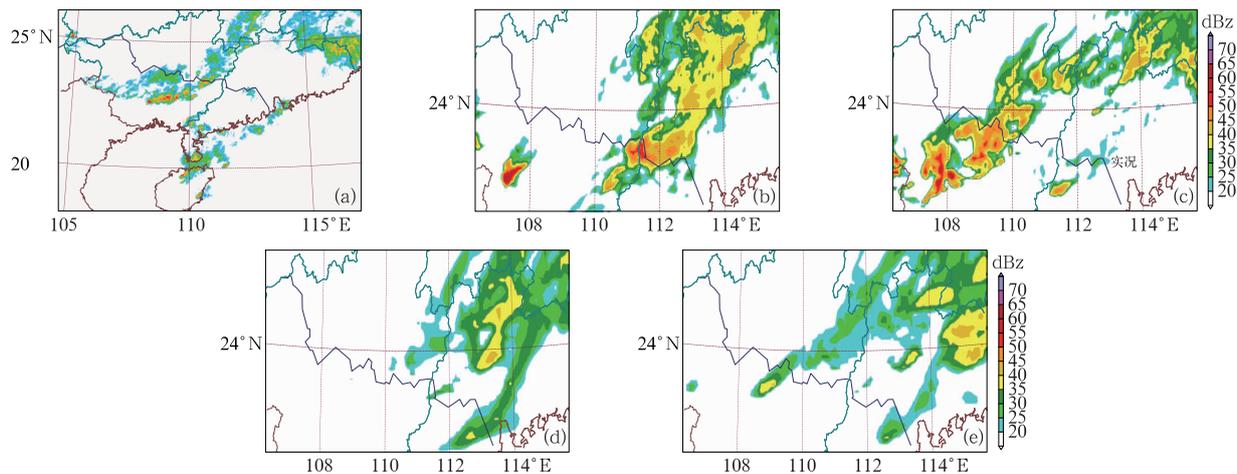


图 1 2018 年 3 月 19 日 18 时雷达组合反射率观测(a), GRAPES\_3 km 模式 18 日 20 时(b)和 19 日 08 时(c)起报的雷达组合反射率预报(预报对象为 3 月 19 日 18 时, 下同),

GRAPES\_Meso 模式 19 日 11 时(d)、14 时(e)起报的雷达组合反射率预报

Fig. 1 Observation of composite reflectivity at 18:00 BT 19 March 2018 (a), the 22 h forecast of composite reflectivity starting at 20:00 BT 18 (b) and the 10 h forecast starting at 08:00 BT 19 (c) of GRAPES\_3 km, the 7 h forecast of composite reflectivity starting at 11:00 BT 19 (d) and the 4 h forecast starting at 14:00 BT 19 (e) by GRAPES\_Meso

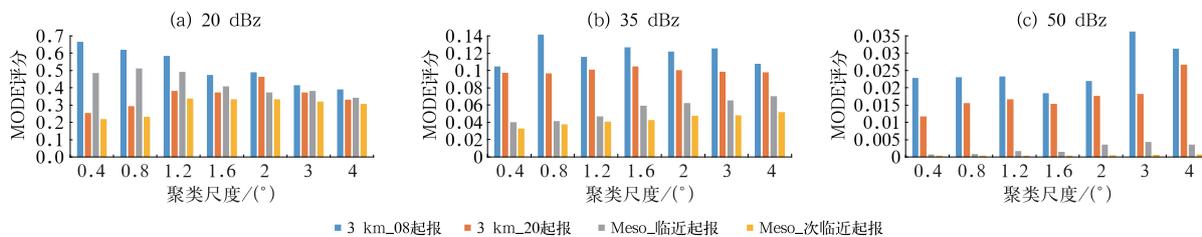


图 2 不同聚类尺度拟合对流风暴的 MODE 评分

Fig. 2 The MODE of the fitting convective storms with different clustering scales

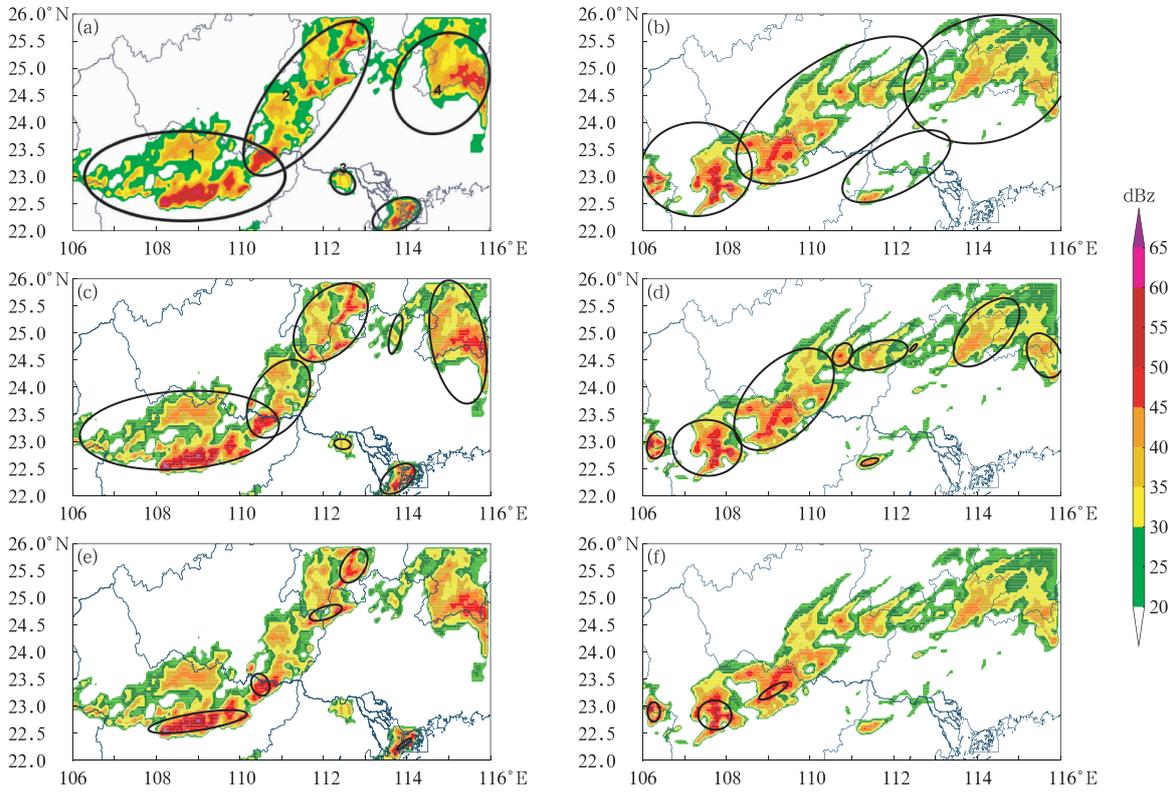


图 3 使用 MODE 最优评估尺度对实况(a,c,e)和模式拟合(b,d,f)椭圆的分布特征  
(a,b)20 dBz,(c,d)35 dBz,(e,f)50 dBz

Fig. 3 Characteristics of observation (a, c, e), and modeled ellipses (b, d, f) of the storms by using optimal MODE evaluation scale  
(a, b) 20 dBz, (c, d) 35 dBz, (e, f) 50 dBz

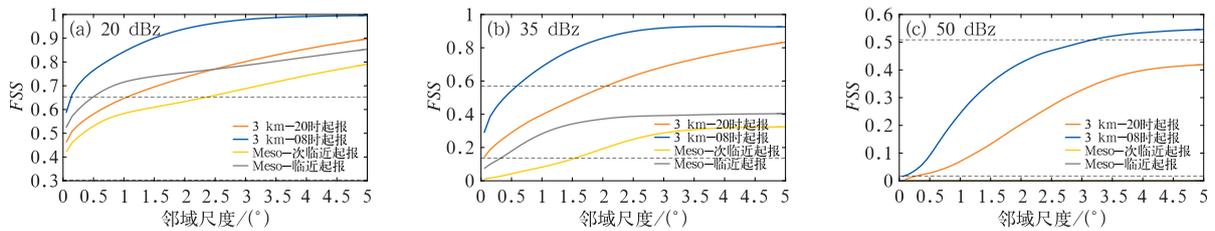


图 4 不同阈值对流风暴随窗区变化的 FSS 评分

Fig. 4 FSS against neighborhood windows using different thresholds of the storms

## 4 整体检验评估

从前文的典型个例来看,新型检验方法能够较为客观评估模式对流风暴的预报能力,因此下文将针对 2018 年 4—8 月两个模式预报效果进行整体检验,以得到具有统计意义的评估结论。

### 4.1 点对点 TS 评分整体评估

图 6 是基于传统点对点 TS 方法对 GRAPES\_

3 km 模式对流风暴预报进行的整体评估。为分析不同起报时间的预报效果,将不同起报时间对同一时刻实况的预报评分放在一起对比。例如预报员在制作强对流天气预报时,在某一时刻(如 09 时),对应有三个不同起报时间的预报,即图中标注的当日 08 时、昨日 20 时和昨日 08 时起报的预报,哪个起报时间的预报效果最好?

从逐小时评分变化可以看到:(1)模式起报的最初 3 h 预报评分最高,基本不存在“spin-up”的问题,3 h 后评分明显下降,这与初始场加入云分析

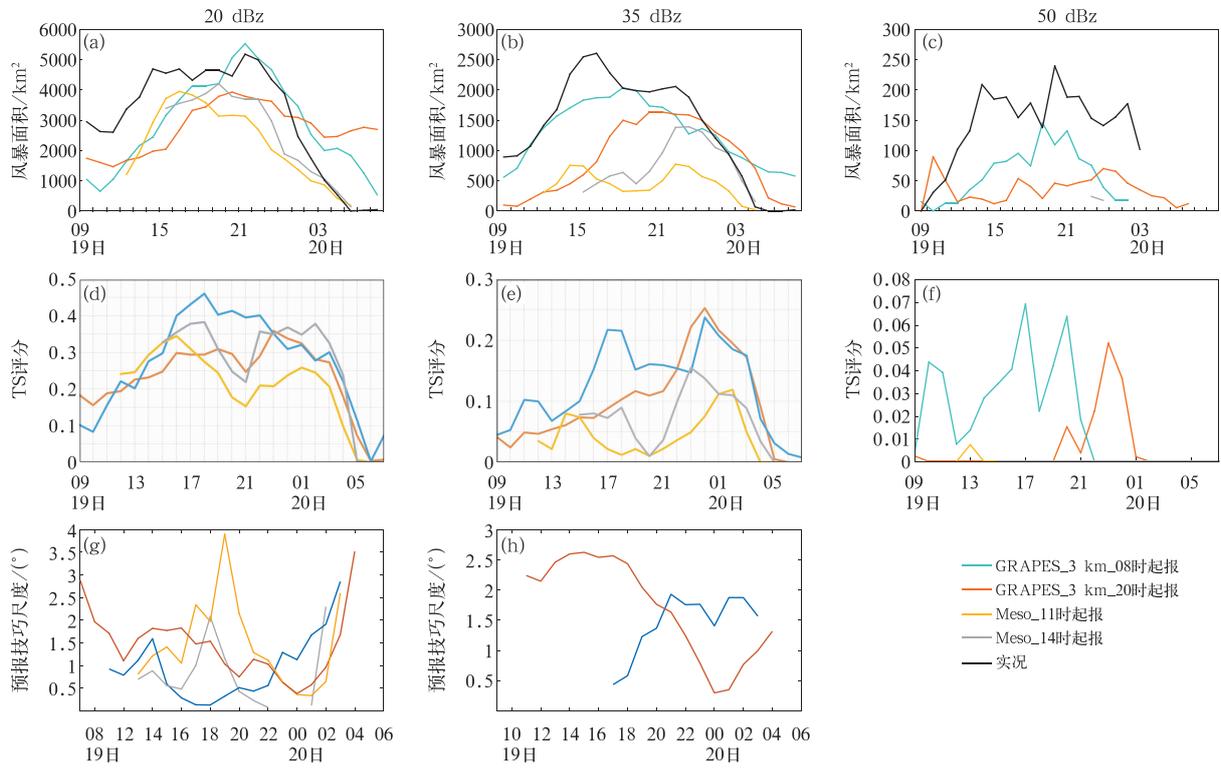


图 5 2018 年 3 月 19 日华南区域强对流过程的对流风暴面积(a,b,c)和邻域 TS 评分(d,e,f) 以及 3 月 19—20 日 FSS 最低技巧尺度(g,h)的时间演变

Fig. 5 Evolution of storm area (a, b, c), TS (d, e, f) on 19 March and FSS lowest forecast skill scale (g, h) of severe convective weather in South China during 19—20 March 2018

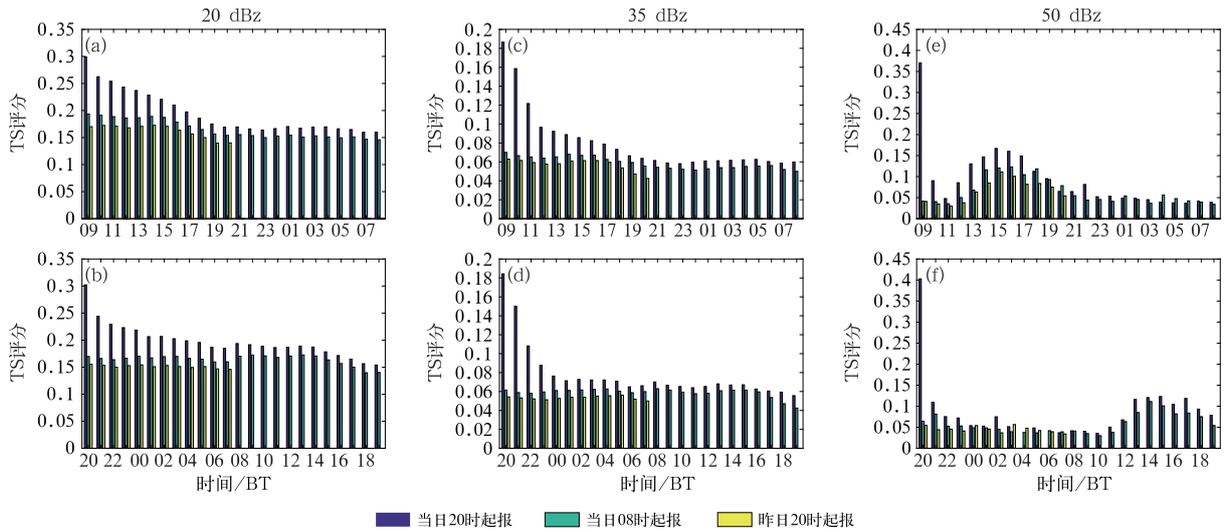


图 6 GRAPES\_3 km 模式起报的逐小时预报检验 (a, c, e)08—08 时时段,(b,d,f)20—20 时时段

Fig. 6 Hourly forecast verification of GRAPES\_3 km model predicting at different initial forecast time for the periods of 08:00 BT—08:00 BT (a, c, e) and 20:00 BT—20:00 BT (b, d, f)

有关。云分析是指加入观测和模式的初始分析构造初始云水物理场,缓解模式需要运行几个小时才能成云致雨的过程。云分析的有效时间大概在 3 h 左右,与之前的研究结论 (Xiao and Sun, 2007) 相吻合。(2) 50 dBz 以上的强对流风暴预报评分具有明显的日变化特征,午后到傍晚评分最高,凌晨评分最低。20 和 35 dBz 日变化不显著。

从不同起报时间对于相同时间的预报对比可以看到:(1)起报时间越新预报效果越好,预报员可以直接采用最新预报产品;(2)评分发展变化的趋势并不一致,这间接反映了不同起报时间的预报对于对流风暴发展过程的刻画是有区别的;(3)50 dBz 评分峰值与 20、35 dBz 峰值无明显对应关系。通过个例检验发现,50 dBz 评分高的个例,对应的 20、35 dBz 的范围通常偏小。

从 GRAPES\_3 km 和 GRAPES\_Meso 两个模式不同预报时效的对比来看(图 7),无论是 08 时还是 20 时起报,GRAPES\_3 km 模式的预报效果均优于 GRAPES\_Meso 模式。20 和 35 dBz 的预报,GRAPES\_3 km 模式 TS 评分是 GRAPES\_Meso 模式的两倍。大于 50 dBz 的预报评分差异更加明显,GRAPES\_3 km 模式在严格点对点检验中评分仍能达到 0.01 以上,说明该模式对于强对流风暴具有一定的预报能力。GRAPES\_Meso 模式的预报评分在

多个时次几乎接近于零,说明该模式对于强对流风暴的预报能力非常有限,这与位置和强度预报偏差都有关系。

## 4.2 时空邻域 TS 方法评估

Lynn et al(2012)对空间分辨率为 4 km 的高分辨率模式预报进行检验评估时,由点对点检验改为空间邻域为 12、36 km 时,预报评分大大提高。但邻域的扩大降低了预报的空间分辨率,会使预报信息被平滑(Mass et al, 2002)。因此使用邻域法检验需要关注的问题是:模式对于哪个邻域尺度可预报性最强、预报技巧最高? 模式最优预报技巧尺度 (Roberts, 2005) 是多少?

图 8 是时空邻域 TS 评分,横坐标为空间邻域尺度(点对面直径)。从传统的点对点评分来看(第一行第一列当空间邻域为 0 时),对于 20 和 35 dBz, 预报评分高低依次是:GRAPES\_3 km 模式当日 08 时起报、GRAPES\_Meso 模式最临近时刻起报、GRAPES\_3 km 模式前一日 20 起报、GRAPES\_Meso 模式前一个时刻起报的预报。考虑了空间邻域时(图 8a~8c),GRAPES\_3 km 模式当日 08 时起报的预报仍为评分最高的预报,与点对点方法判定的结论一致。

空间邻域使 TS 评分有所提高,GRAPES\_3 km

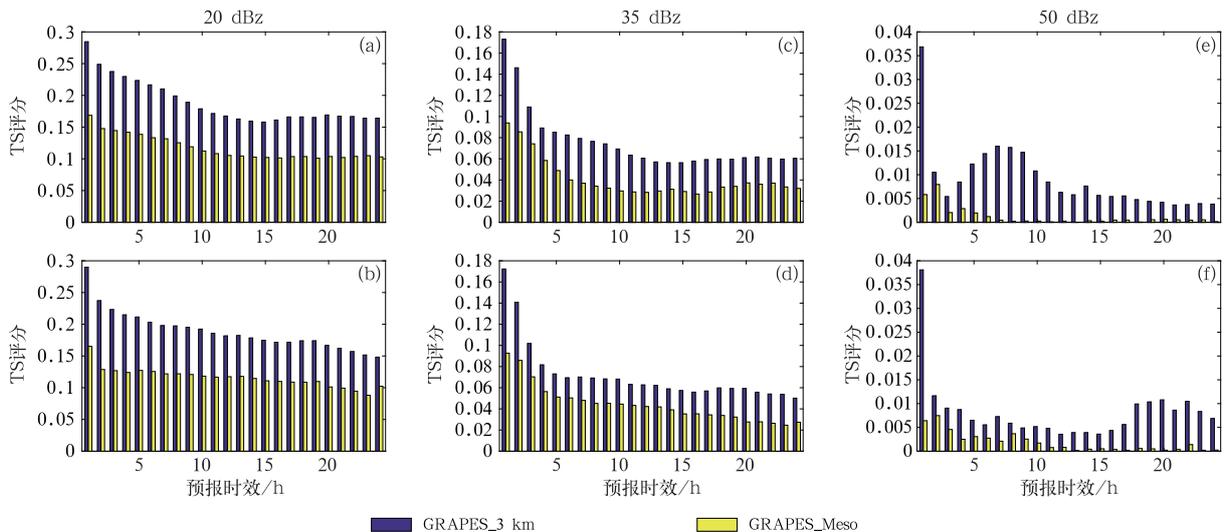


图 7 GRAPES\_3 km 模式与 GRAPES\_Meso 模式不同预报时效的 TS 评分对比

(a, c, e) 08 时起报, (b, d, f) 20 时起报

Fig. 7 Comparison of TS of GRAPES\_3 km model and GRAPES\_Meso model

(a, c, e) the initial forecast time 08:00 BT, (b, d, f) the initial forecast time 20:00 BT

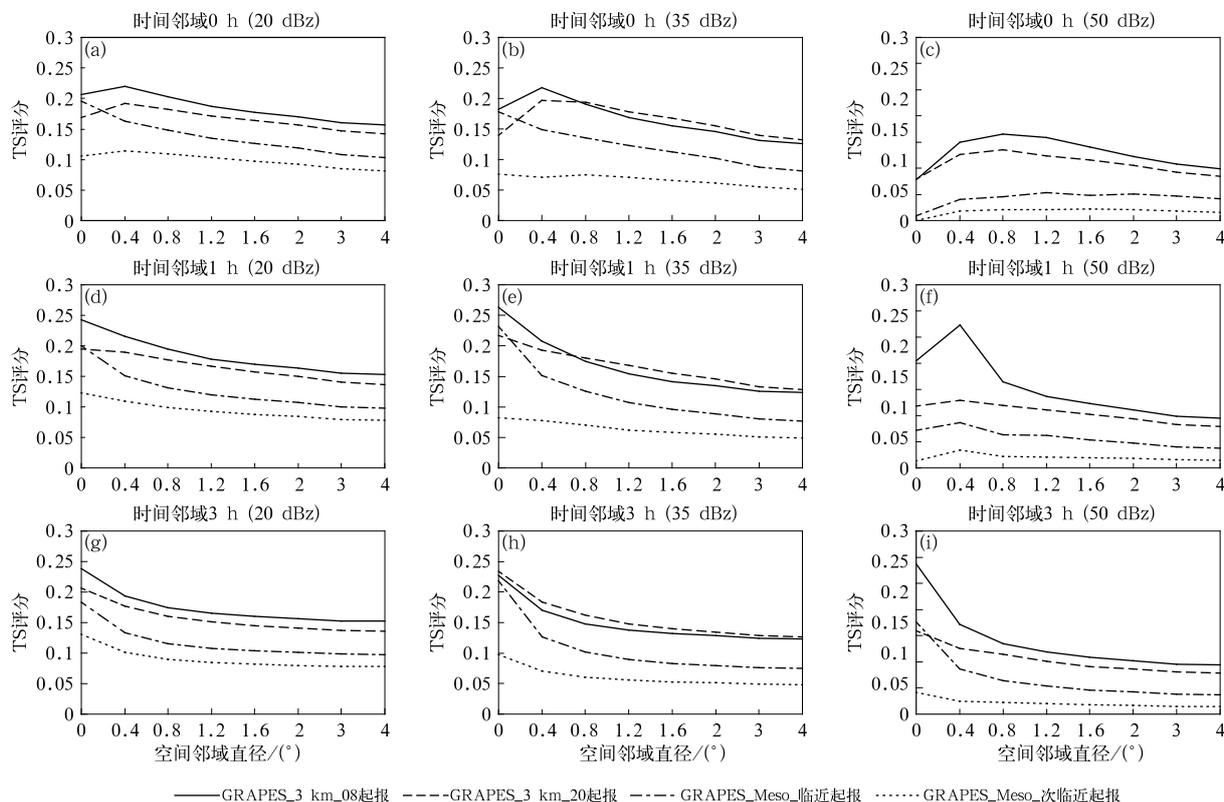


图 8 对流风暴基于不同时空邻域尺度的 TS 评分

Fig. 8 TS based on different temporal and spatial neighborhood using different thresholds of the convective storms

模式在邻域尺度  $0.4^\circ$  时 TS 评分达到最大, GRAPES\_Meso 模式随邻域尺度增大 TS 评分持续下降, 说明该模式由于实时同化最新资料的缘故, 虽然对位置预报较好, 但当忽略一定的空间偏差(邻域)时, 模式对范围和形态的预报能力逊于 GRAPES\_3 km 模式。再同时考虑时间邻域时(图 8d~8f、图 8g~8i)结果又有所不同。时间邻域尺度的增加扩大了实况样本的数量, 当拓展到 1 h(图 8d~8f)可以使模式预报的空间偏差完全忽略, 且评分高于空间邻域的瞬时实况评分。当拓展到 3 h(图 8g~8i)时由于实况样本过多导致漏报率增加, 因此评分低于时间邻域为 1 h 的评分, 这也说明中尺度系统对于局地影响的时间尺度在 1 h 左右。

对于尺度较小的 50 dBz 的强对流风暴预报, 传统点对点 TS 评分(图 8c 中空间邻域为 0 时)已经展示出 GRAPES\_3km 模式的优越性, TS 评分是 GRAPES\_Meso 模式的 4 倍, 这与 4.1 节结论一致。仅考虑空间邻域时, GRAPES\_3 km 模式邻域尺度在  $0.8^\circ$  时, GRAPES\_Meso 模式在  $1.2^\circ$  左右预报评

分最高, 说明 GRAPES\_3 km 模式对强对流风暴预报的空间偏差更小。同时考虑时间邻域时, 两个模式在时间邻域拓展到 3 h 评分最高, GRAPES\_3 km 模式采用空间点对点 TS 评分最高, GRAPES\_Meso 模式仍存在一定的空间偏差, 邻域尺度为  $0.2^\circ$ 。

综合以上可以发现:(1)传统点对点方法和时空邻域 TS 方法检验结论是具有一致性, 点对点方法对于高分辨率模式对流风暴预报的评估也具有参考价值。(2)时空邻域 TS 高于传统点对点 TS 评分。对于不同阈值的对流风暴, 模式评估的时空邻域尺度不同: GRAPES\_3 km 模式对于 20 和 35 dBz 采用时间邻域为 1 h, 空间点对点时预报技巧最高; 50 dBz 采用时间邻域 3 h, 空间点对点时预报技巧最高。这与不同尺度对流风暴的可预报性一致, 强对流风暴尺度小可预报性差, 需要的时空邻域尺度也就更大。

### 4.3 FSS 评估

FSS 通过邻域(窗区)大小的变化, 评估预报和

实况在邻域中的发生频率差异来分析模式的预报能力,当模式预报存在可容忍的空间偏差时仍能给出有价值的评分结果 (Millermaier and Roberts, 2010)。图 9 可见 FSS 评估与点对面的邻域法 TS 评估结论一致,GRAPES\_3 km 模式在各个邻域的 FSS 评分均高于 GRAPES\_Meso 模式,其中当日 08 时起报的预报技巧最高。

表 3 为 FSS 评估的模式最低技巧尺度,该尺度是根据图 9 中第二个横虚线对应的横坐标确定,尺度越小说明模式预报的空间偏差越小。GRAPES\_3 km 模式对于不同阈值的对流风暴均能达到最低技巧尺度,而 GRAPES\_Meso 模式对于 35 dBz 以上的预报无法达到最低预报技巧尺度,对流风暴的预报能力较差。

#### 4.4 MODE 评估

使用 MODE 评估时,聚类尺度等参数很大程度依赖于研究者的主观判断 (Davis et al, 2006)。本节通过对比两个模式不同聚类尺度下的对流风暴预报与实况的差异,评估 MODE 方法的最优评估聚类尺度。

图 10a~10c 为不同聚类尺度的 MODE 评分。GRAPES\_3 km 模式不同的聚类尺度 MODE 评分均高于 GRAPES\_Meso 模式。20 和 35 dBz 的评分呈中间高两边低的“拱形”, GRAPES\_3 km 模式最优评估聚类尺度为 0.8°, GRAPES\_Meso 为 1.2°,说明 GRAPES\_3 km 模式能够刻画更小尺度对流风

暴的特征。对于 50 dBz 的强对流风暴,MODE 评分随着聚类尺度增大而提高,模式最优预报尺度为 4°,该尺度说明两个模式强对流风暴特征的预报能力均比较有限。对比来看,GRAPES\_3 km 模式对于强对流风暴预报更好,当日 08 时起报的预报 MODE 评分是 GRAPES\_Meso 模式的 3 倍。

表 4 为模式基于 MODE 方法的最优预报技巧尺度。MODE 的聚类尺度与 FSS 的邻域窗口相似,都是分析不同邻域中预报场和观测场的差异,但是 MODE 同时考虑了对流风暴的个数、面积、形状、轴向等信息, FSS 仅考虑邻域发生概率,所以虽然两方法对模式预报优劣的评估结果一致,但最优空间邻域尺度的范围有所差异。50 dBz 的强对流风暴预报 MODE 的最优尺度更大,说明模式相对于对流风暴的面积和强度,对于形状、轴向等特征的预报较差。

图 10d~10f 为不同聚类尺度的椭圆个数偏差比(预报与实况的比值,等于 1 时最接近实况)。对于 20 和 35 dBz, GRAPES\_3 km 模式 08 时预报偏差比接近于 1,与实况个数较为一致,前一日 20 时起报的预报椭圆个数较实况略偏少; GRAPES\_Meso 模式预报较实况偏少明显,仅为实况个数的一半,这也是其 TS 评分较低的原因之一。对于 50 dBz 以上的强对流风暴,两个模式预报个数均偏少, GRAPES\_3 km 模式为实况的一半, GRAPES\_Meso 模式仅为实况的 1/5。

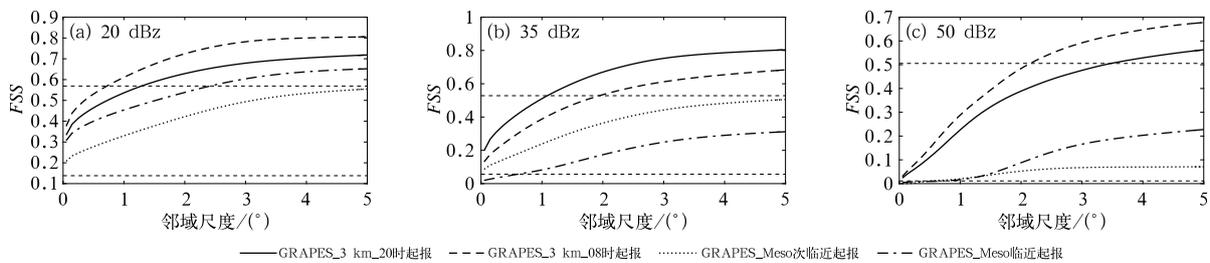


图 9 不同阈值对流风暴随窗区变化 FSS 评分

Fig. 9 FSS against neighborhood windows using different thresholds of the convective storms

表 3 基于 FSS 评估不同模式的预报技巧尺度

Table 3 Forecast skill scale evaluation of different models based on FSS

不同阈值/dBz	GRAPES_3 km 模式		GRAPES_Meso 模式	
	08 时起报	20 时起报	最临近时刻起报	前一个时刻起报
20	0.8°	1.4°	2.4°	无
35	1.0°	1.8°	无	无
50	2.2°	3.6°	无	无

表 4 基于 MODE 评估不同模式的预报技巧尺度

Table 4 Forecast skill scale evaluation of different models based on MODE

不同阈值/dBz	GRAPES_3 km 模式		GRAPES_Meso 模式	
	08 时起报	20 时起报	最临近时刻起报	前一个时刻起报
20	0.8°	0.8°	1.2°	1.2°
35	0.8°	1.6°	1.6°	3.0°
50	4.0°	4.0°	4.0°	4.0°

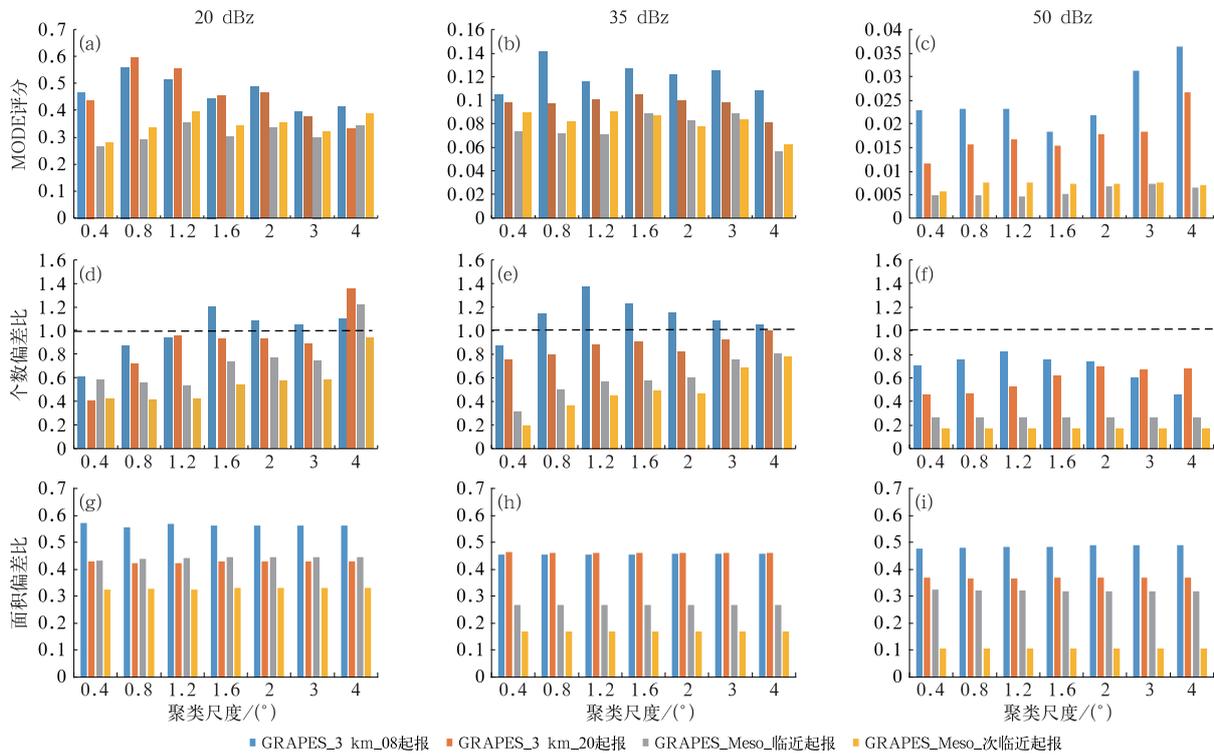


图 10 不同聚类尺度拟合的对流风暴特征

Fig. 10 Fitting elliptical features of convective storms with different clustering scales

图 10g~10i 为不同聚类尺度的椭圆面积偏差比,两个模式对于各个尺度的对流风暴面积均低估,特别是强对流风暴预报面积明显偏小,这与 Cai and Dumais(2015)在对流许可尺度模式的检验结果类似。对比来看,GRAPES\_3 km 模式面积预报明显好于 GRAPES\_Meso 模式,其中当日 08 时起报的预报效果最好。

图 11 以箱线图的形式统计不同聚类尺度椭圆的距离偏差、长轴、长短轴比和轴角,分别体现了对流风暴的尺度,形状和角度。

距离偏差(图 11a~11c)是指 MODE 拟合的对流风暴中心的距离偏差,距离越小说明模式预报位置偏差越小。从距离偏差的中值和离散度来看,GRAPES\_3 km 模式预报对流风暴的距离偏差在 1°~3°左右且离散度较小, GRAPES\_Meso 模式的距

离偏差在 3°~4°且离散度较大,说明 GRAPES\_3 km 模式对位置的预报确定性更高、更稳定,同时距离偏差较小。

椭圆长轴(图 11d~11f)基本反映了对流风暴的尺度,总体来看,两个模式对小尺度的对流风暴(聚类尺度小于 1.2°)预报尺度偏大、对于较大尺度的对流风暴(聚类尺度大于 1.6°)预报尺度偏小。从不同模式不同起报时间的预报来看,GRAPES\_3 km 模式 08 时起报与实况较一致,对中小尺度(聚类尺度为 0.8°~2.0°)对流风暴的尺度预报较好,前一日 20 时起报预报尺度偏小且离散度较大。GRAPES\_Meso 模式最临近时刻起报的对流风暴预报尺度明显偏小,仅是实况的一半,50 dBz 的强对流风暴尺度预报为实况的 1/6;而前一时刻起报的预报基本无预报能力。

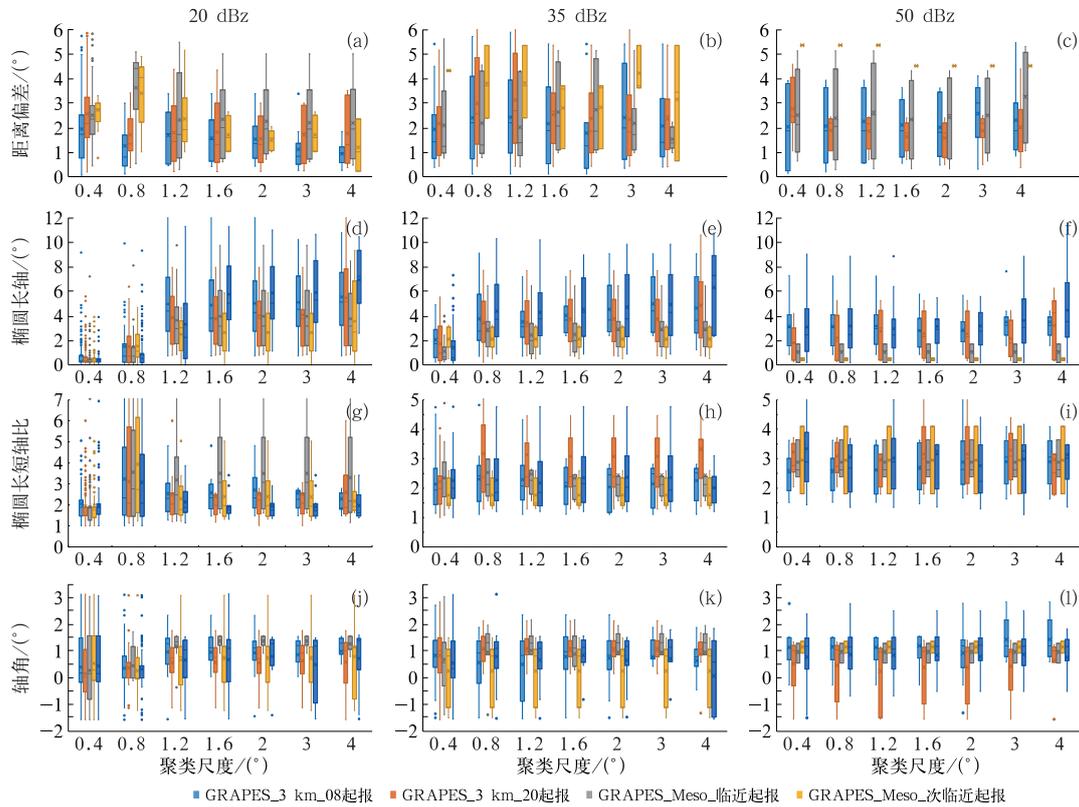


图 11 对流风暴在不同聚类尺度下拟合的椭圆特征分析

Fig. 11 Characteristics of storm fitting ellipses at different clustering scales

椭圆长短轴比和轴角(图 11g~11i)反映了对流风暴的形态。长短轴比接近 1 是圆形,大于 1 时是椭圆形。轴角是椭圆与正东分量的夹角,夹角范围为,逆时针旋转为正,夹角为正时椭圆为西南—东北向,夹角为负则椭圆为东南—西北向,轴角越大说明越倾斜(南北方向分量越大)。两个模式对尺度较小的对流风暴形状预报更圆且轴角更小,尺度较大的对流风暴预报形状更扁轴角更大。GRAPES\_3 km 模式当日 08 时起报的形态预报与实况较为一致,前一日 20 时起报较实况更扁,轴角更偏向东南—西北向。GRAPES\_Meso 模式形态预报较实况更圆,轴角更小。GRAPES\_Meso 模式对流风暴形态预报的偏差较 GRAPES\_3 km 模式更大,无法准确刻画出对流风暴的细节特征,这可能与分辨率较低有关。

综合以上,基于 MODE 方法的对流风暴特征预报,GRAPES\_3 km 模式位置预报的偏差更小且确定性更高,能够较好地反映出  $\beta$  中尺度对流风暴的尺度和形态,但是对  $\gamma$  尺度对流风暴尺度预报偏大、形状偏圆、轴角偏小,对  $\alpha$  尺度对流风暴尺度预报尺

度偏小、形状偏扁、轴角偏大。GRAPES\_Meso 模式预报的位置偏差较大,各个尺度预报均偏小,形状预报较实况偏圆、轴角偏小,对 50 dBz 以上的强对流风暴预报能力较差。

## 5 结论与讨论

本文通过传统点对点 TS 评分,时空邻域 TS、FSS 和 MODE 对比分析了 GRAPES\_3 km 和 GRAPES\_Meso 两个模式对流风暴的预报能力,结论如下:

(1)传统点对点 TS 评分显示,GRAPES\_3 km 模式对于不同阈值的对流风暴预报均优于 GRAPES\_Meso 模式,对于不同起报时间的预报,起报时间越新预报效果越好。GRAPES\_3 km 模式最初 3 h 评分最高,基本不存在“spin-up”,3 h 后评分明显下降。50 dBz 以上的强对流风暴预报具有明显的日变化特征,20 和 35 dBz 日变化不显著。

(2)邻域 TS 方法能够得到点对面预报时的最优评估尺度。两个模式对于 20 和 35 dBz 最优评估

尺度为时间邻域尺度为 1 h,空间采用点对点;对于 50 dBz 的强对流风暴时间邻域尺度需要拓展到 3 h,此时 GRAPES\_Meso 模式仍有 0.4° 的空间偏差(空间邻域直径),说明 GRAPES\_Meso 模式预报时间和空间差异均较大。当忽略一定的空间偏差(邻域)时,GRAPES\_3 km 模式预报 TS 评分明显高于 GRAPES\_Meso 模式,其中 GRAPES\_3 km 模式当日 08 时起报的预报效果最好。

(3)GRAPES\_3 km 模式在各个邻域的 FSS 均高于 GRAPES\_Meso 模式,对于不同阈值的对流风暴 GRAPES\_3 km 模式均能达到最低技巧尺度,GRAPES\_Meso 模式对于 35 dBz 以上的对流风暴预报无法达到最低预报技巧分数。

(4)MODE 可以客观识别对流风暴的位置、形状、方向等信息,GRAPES\_3 km 模式对流风暴个数的预报较好,面积预报偏小,对  $\beta$  中尺度对流风暴的预报效果最好。GRAPES\_Meso 模式的对流风暴面积、个数、尺度预报均低估,无法准确刻画对流风暴细节特征。MODE 评估的强对流风暴的最优尺度大于 FSS,原因可能与模式对强对流风暴的形状、轴向等刻画能力较差有关。

(5)通过强对流典型个例评估表明,多种评估方法可以客观全面地评估对流风暴特征的预报性能,分析模式对流风暴初生和发展变化的预报能力,从而给模式开发人员重要的反馈信息,同时为预报员在模式使用时做出修正和正确的决策方案提供依据。

本文通过几种不同检验方法对 GRAPES\_3 km 模式和 GRAPES\_Meso 模式进行了整体评估,每种评估方法虽然给出的信息不同,但结论基本一致,都可以作为高分辨率模式评估的依据。这几种检验方法的区别为:时空邻域 TS 是“点对面”的检验,这对于机场等固定点的单点预报非常有意义。FSS 是将实况邻域和模式邻域的两个场进行比较,衡量不同尺度(邻域)发生概率的相似性,同时 FSS 可以给出模式的最低预报技巧尺度。MODE 也是基于两个场的检验,但对检验对象是否配对的条件更加严格。MODE 需要人为地规定卷积和聚类尺度,更适合针对小区域短时临近预报中的人机交互检验。综合以上,用户可以根据不同目的选择不同的检验方法,单一的检验方式不可能包含模式预报性能的全部信息。本文所得结果可为智能网格预报产品和业务短临预报提供参考。另外,发展检验技术的实时评估

结果也可多模式集成预报技术提供模式评估结果,今后将进一步研究这些评估方法在多模式集成预报技术中的应用方法。

## 参考文献

- 陈德辉,沈学顺,2006.新一代数值预报系统 GRAPES 研究进展[J].应用气象学报,17(6):773-777. Chen D H, Shen X S, 2006. Recent progress on GRAPES research and application[J]. J Appl Meteor Sci, 17(6):773-777(in Chinese).
- 黄丽萍,陈德辉,邓莲堂,等,2017. GRAPES\_Meso V4.0 主要技术改进和预报效果检验[J].应用气象学报,28(1):25-37. Huang L P, Chen D H, Deng L T, et al, 2017. Main technical improvements of GRAPES\_Meso V4.0 and verification[J]. J Appl Meteor Sci, 28(1):25-37(in Chinese).
- 李佰平,戴建华,张欣,等,2016.三类强对流天气临近预报的模糊检验试验与对比[J].气象,42(2):129-143. Li B P, Dai J H, Zhang X, et al, 2016. Fuzzy verification test and comparison of three types of severe convective weather nowcasting[J]. Meteor Mon, 42(2):129-143(in Chinese).
- 刘凑华,牛若芸,2013.基于目标的降水检验方法及应用[J].气象,39(6):681-690. Liu C H, Niu R Y, 2013. Object-based precipitation verification method and its application[J]. Meteor Mon, 39(6):681-690(in Chinese).
- 毛冬艳,朱文剑,樊利强,等,2014. GRAPES\_MESO(V3.3)模式强天气预报性能的初步检验[J].气象,40(12):1429-1438. Mao D Y, Zhu W J, Fan L Q, et al, 2014. Preliminary synoptic verification on the performance of GRAPES\_MESO V3.3 in severe weather forecasting[J]. Meteor Mon, 40(12):1429-1438(in Chinese).
- 茅懋,戴建华,李佰平,等,2016.不同类型强对流预报产品的目标对象检验与分析评价[J].气象,42(4):389-397. Mao M, Dai J H, Li B P, et al, 2016. Object based verification and evaluation for different types of severe convection forecasting products[J]. Meteor Mon, 42(4):389-397(in Chinese).
- 唐文苑,郑永光,张小雯,2018.基于 FSS 的高分辨率模式华北对流预报能力评估[J].应用气象学报,29(5):513-523. Tang W Y, Zheng Y G, Zhang X W, 2018. FSS-based evaluation on convective weather forecasts in North China from high resolution models [J]. J Appl Meteor Sci, 29(5):513-523(in Chinese).
- 许晨璐,王建捷,黄丽萍,2017.千米尺度分辨率下 GRAPES-Meso4.0 模式定量降水预报性能评估[J].气象学报,75(6):851-876. Xu C L, Wang J J, Huang L P, 2017. Evaluation on QPF of GRAPES-Meso 4.0 model at convection-permitting resolution [J]. Acta Meteor Sin, 75(6):851-876(in Chinese).
- 徐枝芳,郝民,朱立娟,等,2013. GRAPES\_RAFS 系统研发[J].气象,39(4):466-477. Xu Z F, Hao M, Zhu L J, et al, 2013. On the research and development of GRAPES\_RAFS[J]. Meteor Mon, 39(4):466-477(in Chinese).
- 薛纪善,2006.新世纪初我国数值天气预报的科技创新研究[J].应用气象学报,17(5):602-610. Xue J S, 2006. Progress of Chinese

- numerical prediction in the early new century[J]. *J Appl Meteor Sci*, 17(5):602-610(in Chinese).
- Cai H Q, Dumais Jr R E, 2015. Object-based evaluation of a numerical weather prediction model's performance through forecast storm characteristic analysis[J]. *Wea Forecasting*, 30(6):1451-1468.
- Davis C, Brown B, Bullock R, 2006. Object-based verification of precipitation forecasts. Part I: methodology and application to mesoscale rain areas[J]. *Mon Wea Rev*, 134(7):1772-1784.
- Dixon M, Wiener G, 1993. TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—a radar-based methodology [J]. *J Atmos Oceanic Technol*, 10(6):785-797.
- Ebert E E, 2008. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework[J]. *Meteor Appl*, 15(1):51-64.
- Ghelli A, Lalaurette A, 2000. Verifying precipitation forecasts using up-scaled observations[J]. *ECMWF Newsl*, 87:9-17.
- Lynn B H, Yair Y, Price C, et al, 2012. Predicting cloud-to-ground and intracloud lightning in weather forecast models[J]. *Wea Forecasting*, 27(6):1470-1488.
- Marzban C, Sandgathe S, 2006. Cluster analysis for verification of precipitation fields[J]. *Wea Forecasting*, 21(5):824-838.
- Mass C F, Ovens D, Westrick K, et al, 2002. Does increasing horizontal resolution produce more skillful forecasts? [J]. *Bull Amer Meteor Soc*, 83(3):407-430.
- McCaul Jr E W, Goodman S J, LaCasse K M, et al, 2009. Forecasting lightning threat using cloud-resolving model simulations [J]. *Wea Forecasting*, 24(3):709-729.
- Mittermaier M, Roberts N, 2010. Intercomparison of spatial forecast verification methods: identifying skillful spatial scales using the fractions skill score[J]. *Wea Forecasting*, 25(1):343-354.
- Roberts N, 2005. An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall[R]. Forecasting Research Technical Report NO. 455, Met Office:80.
- Roberts N M, Lean H W, 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events[J]. *Mon Wea Rev*, 136(1):78-97.
- Rodriguez A, Laio A, 2014. Machine learning clustering by fast search and find of density peaks[J]. *Science*, 344(6192):1492-1496.
- Tustison B, Harris D, Foufoula-Georgiou E, 2001. Scale issues in verification of precipitation forecasts[J]. *J Geophys Res Atmos*, 106(D11):11775-11784.
- Xiao Q and Sun J, 2007. Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP 2002[J]. *Mon Wea Rev*, 135:3381-3404.
- Weisman M L, Davis C, Wang W, et al, 2008. Experiences with 0-36-h explicit convective forecasts with the WRF-ARW Model[J]. *Wea Forecasting*, 23(3):407-437.