0254-6124/2026/46(5)-0001-15 *Chin. J. Space Sci.* 空间科学学报

LIU Dandan, ZHAO Haijuan, LI Mu, CHEN Anqin, TANG Wei, GUO Jianguang. Analysis of Probabilistic Forecasts for NCSW Space Weather Events (in Chinese). *Chinese Journal of Space Science*, 2026, **46**(5): 1–15. CSTR:32142.14.cjss.2025–0188. DOI:10.11728/cjss2026.05.2025–0188

基于 NCSW 空间天气事件概率预报水平分析 *

刘丹丹　赵海娟　李　沐
陈安芹　唐　伟　郭建广

(国家卫星气象中心 (国家空间天气监测预警中心)　北京　100081)

(许健民气象卫星创新中心　北京　100081)

(中国气象局空间天气重点实验室　北京　100081)

摘　要　基于国家空间天气监测预警中心 (NCSW)2005－2022 年空间天气事件概率预报数据, 系统检验 M 级耀斑、X 级耀斑、质子事件、小地磁暴及中等及以上地磁暴预报性能, 并在可靠性–锐度–分辨率框架下与美国 SWPC 对比. 结果表明, 各类事件预报技巧均随时效延长衰减. M 级耀斑预报在样本充足年份性能稳定, 太阳活动低年波动明显; X 级耀斑样本稀缺, 空报现象突出, 业务上呈现“宁可虚报”的策略; 质子事件采用零虚报的保守预报策略; 小地磁暴预报在后期同等样本条件下技巧较早期显著提升; 中等及以上地磁暴发生频次低, 整体预报技巧偏弱. 多指标联合分析显示, TSS 适用于逐年业务能力跟踪, 统计特征分析服务于长期系统校准, PC 指标受事件基础率影响显著, 需结合 TSS 综合评估. 对比可见, NCSW 在预报可靠性与虚报控制方面具备优势, SWPC 锐度表现更强; 短时效 NCSW 综合判别能力更佳, 二者性能差异随预报时效延长逐步缩小. 本文完成 NCSW 18 年概率预报连续检验, 揭示稀少空间天气事件预报特征, 可为中国空间天气业务评估与预报能力建设提供支撑.

关键词　空间天气事件, 太阳活动水平, 太阳耀斑, 太阳质子事件, 地磁暴, TS 评分, TSS 评分

中图分类号　P353

Analysis of Probabilistic Forecasts for NCSW Space Weather Events

LIU Dandan　ZHAO Haijuan　LI Mu
CHEN Anqin　TANG Wei　GUO Jianguang

(*National Satellite Meteorological Center* (*National Center for Space Weather*), *China Meteorological Administration*, *Beijing* 100081)

(*Innovation Center for FengYun Meteorological Satellite* (*FYSIC*),

* 中国气象局创新发展专项项目 2024(202401YW043), 国家重点研发计划项目 (2022YFB3901802) 和中国气象局“电离层天气预报预警服务”青年创新团队项目 (CMA2024QN09) 共同资助

2025-11-10 收到原稿, 2026-07-23 收到修定稿

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License (https://creativecommons.org/licenses/by/4.0/)

China Meteorological Administration, *Beijing* 100081)

(*Key Laboratory of Space Weather*, *China Meteorological Administration*, *Beijing* 100081)

Abstract Based on the probabilistic forecast data of space weather events from the National Center for Space Weather, China (NCSW) during 2005–2022, this study systematically examines the forecast performance of M-class flares, X-class flares, proton events, minor geomagnetic storms, and moderate-and-above geomagnetic storms, and compares the results with the U.S. Space Weather Prediction Center (SWPC) within the reliability-sharpness-resolution framework. The results show that the forecast skill of all event types degrades with increasing lead time. For M-class flares, the forecast performance is stable in years with sufficient samples but fluctuates obviously during solar minimum. X-class flares have scarce samples and prominent false alarms, with an operational strategy of "preferring false alarms to misses". Proton event forecasts adopt a conservative zero-false-alarm strategy. Under comparable sample conditions, the forecast skill of minor geomagnetic storms improves significantly in later periods compared with earlier stages. Moderate-and-above geomagnetic storms have low occurrence frequency and generally weak forecast skill. Multi-metric joint analysis indicates that TSS is suitable for annual operational capability tracking, and statistical characteristic analysis serves long-term systematic calibration. The PC metric is strongly affected by the event base rate and requires comprehensive assessment combined with TSS. The comparison shows that NCSW has advantages in forecast reliability and false alarm control, while SWPC achieves stronger sharpness. NCSW presents better overall discrimination skill at short lead times, and the performance difference between the two agencies gradually narrows with extended lead time. This study conducts continuous verification of NCSW's 18-year probabilistic forecasts and reveals the forecast characteristics of rare space weather events, which provides support for China's space weather operational assessment and forecasting capability development.

Key words Space weather events, Solar activity level, Solar flares, Solar proton events, Geomagnetic storms, TS score, TSS score

0 引言

太阳耀斑和日冕物质抛射等剧烈太阳活动会显著扰动地球磁层、电离层及中高层大气, 对卫星运行、通信、导航、电力输送等关键基础设施构成严重威胁[1–5]. 灾害性空间天气事件不仅可能造成重大经济损失, 其释放的高能辐射还会对宇航员及高空飞行人员产生辐射危害[6]. 因此, 提升空间天气概率预报的精准能力已成为空间环境安全保障的必然需求.

目前, 全球主要空间天气业务中心均已建立常态化预报业务. 美国国家海洋和大气管理局空间天气预报中心 (SWPC), 其前身为 1995 年成立的空间环境中心 (SEC), 2007 年 SEC(Space Environment Center) 更名为 SWPC(Space Weather Prediction Center); 欧洲空间局 (ESA) 于 1999 年启动空间天气专项研究, 英国、比利时等国陆续建立业务中心[7–8]. 中国空间天气业务起步于 20 世纪 90 年代末期, 中国气象局国家空间天气监测预警中心 (NCSW) 于 2002 年成立, 2004 年开始业务化运行[9]; 中国科学院国家空间科学中心空间环境预报中心 (SEPC) 于 1992 年随载人航天工程启动开展筹建工作, 并于 1999 年 12 月正式投入业务运行[10]. 目前, 国际上公开发布预报数据并可获取历史数据的主要业务机构包括: NCSW (National Center for Space Weather)、SEPC(Space Environment Prediction Center)、美国 NOAA 空间天气预报中心 (SWPC)、英国气象局空间天气业务中心 (MOSWOC)、澳大利亚空间天气预报中心 (ASWFC)、比利时皇家天文台太阳影响数据分析中心 (SIDC) 以及日本气象厅 (JMA) 空间天气信息中心等.

预报评分是衡量预报水平的关键客观依据[11]. 国际主要空间天气业务中心已针对不同空间天气事件建立了系统化的检验体系. SWPC 采用临界成功指数 (CSI) 最大值对应概率作为阈值, 对第 23 太阳活动周

的业务化太阳耀斑预报模型展开技能评分分析[12]. 后续研究进一步基于 1998−2024 年长达 26 年的观测数据, 采用使 TS 值最大的概率作为阈值, 对太阳耀斑概率预报进行系统检验分析, 发现业务预报整体表现未显著优于基准模型且存在误报率偏高的问题[13]. 在太阳质子事件预报方面, SWPC 不仅选取 1986−2004 年期间的质子事件对业务模型进行性能检验分析, 还系统检验了第 23 和 24 太阳活动周的各类概率预报数据, 分析了预报性能与技巧的演变特征[14,15]. MOSWOC 则采用多维度检验方法, 通过可靠性图、ROC 曲线和区分图对 2015−2016 年发布的地磁暴和耀斑概率预报进行检验. 分析表明, 尽管两类预报具备一定的事件区分能力, 但普遍存在过度预报和虚报率偏高的问题. 研究进一步验证了人工干预对预报技能的提升作用: 经人工订正的预报结果在分级概率技巧评分 (RPSS) 上优于原始模式输出, 且随着预报提前时长的增加, 整体预报技巧呈逐步下降趋势[16,17]. SIDC 则采用 50% 固定概率阈值对太阳耀斑和地磁暴预报进行检验分析. 分析发现, 具备扎实太阳物理背景和丰富经验的预报员, 相较于数值模型能获得更高的技巧评分, 凸显了专业预报员在业务预报中的关键作用[18]. 国内 SEPC 对质子事件概率预报结果进行了统计性能检验[19]. 针对概率预报检验中的阈值判定问题, NCSW 提出一套判定方法, 并利用该方法对其地磁暴短期概率预报模型的预报效能开展了系统性检验分析[20].

然而, 当前业务化检验分析仍面临三方面挑战. 一是现有评分指标众多, 但业务服务用户难以全面理解和应用, 亟需从中选取关键指标构建简洁而全面的检验分析体系. 二是不同业务中心预报发布时间存在差异, 导致后发机构在信息获取上具备固有的时间优势; 同时, 各机构判定事件发生与否的概率阈值存在差异, 若直接采用固定阈值进行分类评分对比, 极易因阈值设定的差异产生偏差, 难以实现公平有效的国际对比. 三是概率预报的统计特性 (可靠性与锐度) 虽可用于表征预报系统整体性能, 但若按年分段统计, 往往因单一年份事件样本稀疏 (尤其是强事件), 导致逐年可靠性曲线波动剧烈、统计代表性差, 难以客观反映年度间的水平变化. 针对上述挑战, 本文依据检验对象的统计特征和业务需求, 选用现有成熟的评分标准, 并结合概率预报统计性能分析方法, 构建适用于业务化空间天气概率预报的综合检验方案. 具体策略如下.

(1) 单中心内部性能诊断. 从众多指标中选取简明多维指标集 POD(Probability of Detection) 反映事件捕捉能力, FAR(False Alarm Ratio) 衡量空报程度, Bias 揭示系统性偏差, TS 表征综合精准度, TSS 衡量判别能力), 兼顾检验全面性与业务用户可理解性. 鉴于第 24 太阳活动周整体活动水平偏低, 极端事件样本极度稀疏, 在这种极端不平衡样本条件下, TSS 虽有局限, 但相较于对样本比例极度敏感的指标, 仍能提供相对有效的判别信息.

(2) 跨机构公平对比. 针对各机构概率阈值不同导致的对比困难, 采用基于概率预报整体统计特性的可靠性–锐度分析方法. 该方法不依赖单一事件阈值判定, 而是通过长期累积的预报概率与观测结果的对应关系分析预报质量, 从而在各机构业务机制存在差异的条件下实现公平、客观的对比. 对于发布时间差异带来的时效性影响, 本文在分析中予以客观呈现, 不作人为抹平.

(3) 业务服务适配性. 将分类评分 (TS, TSS) 与统计特性分析相结合, 前者用于逐年能力分析和业务监测, 满足业务服务对年度预报水平跟踪的需求; 后者用于长期性能对比和系统校准检验, 二者互为补充.

本文聚焦于太阳耀斑、太阳质子事件和地磁暴三类关键事件, 基于其物理关联性强 (构成从太阳驱动源到近地空间响应的关键灾害链条[6]) 且为全球业务中心核心预报数据的特点, 对 NCSW 2005−2022 年长达 18 年的 1～3 天概率预报数据进行系统性检验, 并与 SWPC 同期数据进行对比分析.

1 数据与事件定义

分析 NCSW 对未来 1～3 天内三类空间天气事件的逐日概率预报: 太阳耀斑、太阳质子事件和地磁暴. NCSW 于 09:30 LT(01:30 UT) 发布预报, 起报时间为当日 00:00 UT. 本文将预报时效划分为三个时段: 第 1 日 (0～24 h), 对应世界时发布当日; 第 2 日 (24～48 h); 第 3 日 (48～72 h).

1.1 耀斑事件

太阳软 X 射线耀斑是指在 0.1～0.8 nm 波段观测到的太阳辐射爆发现象, 其强度以地球大气层外、

1 AU 处的峰值流量 ($W{\cdot}m^{-2}$) 衡量[21]. 根据 NOAA/GOES 卫星观测, 耀斑分为 A, B, C, M, X 五个等级 (见表 1)[22]. 本文重点关注可能引发灾害性空间天气的 M 级 ($\geqslant 10^{-5}$ $W{\cdot}m^{-2}$) 及以上耀斑. 分析以 GOES-13, GOES-14, GOES-15 卫星的 1～8 Å波段软 X 射线观测数据为基准, 检验 NCSW 对 M 级和 X 级耀斑的预报效果.

1.2 太阳质子事件

太阳质子事件是指太阳活动导致地球静止轨道处能量>10 MeV 的质子通量强度持续 ≥15 min 超过 10 pfu(1 pfu=1 proton$\cdot$cm$^{-2}\cdot$sr$^{-1}\cdot$s^{-1}) 的事件[23]. 本文以 GOES 系列卫星能量粒子传感器 (EPS/HEPAD) 探测的>10 MeV 质子通量为观测真值, 数据来自 NOAA/SWPC 官方发布的 5 min 平均通量[24]. 由于 2005－2022 年间质子事件仅 141 天 (气候概率 2.1%), 远低于概率预报检验所需的样本分布和事件/非事件平衡性要求 (事件样本量建议 ≥30)[25], 且以 10% 为阈值构建二分类列联表, SWPC 质子事件概率阈值与 NCSW 并不一致且发布时间不同, 导致无法基于标准二分列联表 (见表 4) 评分体系进行公平对比. 因此, 仅对 NCSW 的质子事件概率预报进行独立检验分析, 未与 SWPC 业务预报进行直接比较.

1.3 地磁暴事件

地磁暴表现为地球磁场水平分量的剧烈扰动, 其扰动幅度通常在几十至几百纳特之间, 极端情况下可超过 1000 nT[26]. 业务上, 常依据行星 3 h 磁情指数 (Kp) 划分地磁暴等级 (见表 2). 本文重点关注可能引发灾害性空间天气的小地磁暴 (Kp=5) 和中等以上地磁暴 ($Kp\geqslant 6$). 鉴于 NCSW 业务中采用 SWPC 的 Kp 估计值作为参考, 为确保检验分析与业务实践一致, 本文以 SWPC 发布的 Kp 指数估计值为观测基准, 分析 NCSW 对小地磁暴 (Kp=5) 与中等以上地磁暴 ($Kp\geqslant 6$) 的预报结果.

2 概率预报检验方法

2.1 空间天气事件分类检验指标

2.1.1 概率阈值与事件判定

依据行业标准《空间天气短期预报检验方法》(QX/T 295－2015)[27], 采用设定的概率阈值对预报数据进行检验. 当预报概率 (P) 超过阈值时, 即判定该日事件会发生 (见表 3).

2.1.2 检验指标定义

采用以下 6 个指标定量分析 NCSW 每日事件预报性能[25].

(1) 命中率 (POD, δ_{POD}) 反映对实际发生事件的捕捉能力, 定义为命中数 (N_k^{A})(表 4) 与观测事件总数 (N_k^{C})(表 4) 之比, 其表达式为

$$\delta_{\mathrm{POD}}=\frac{N_k^{\mathrm{A}}}{N_k^{\mathrm{A}}+N_k^{\mathrm{C}}}. \tag{1}$$

其中, k=1 代表 M 级耀斑或小地磁暴, k=2 代表 X 级耀斑或中等以上地磁暴; 质子事件仅设 k=3.

(2) 空报率 (FAR, δ_{FAR}) 衡量空报严重程度, 即空报数占所有预报事件的比例, 其表达式为

表 1 太阳软 X 射线耀斑强度等级划分

Table 1 Classification of solar soft X-ray flare intensity

强度级别	流量范围 (W·m–2)	强度级别	流量范围 (W·m–2)
A	$F_{\mathrm{X}}<1.00\times10^{-7}$	B	$1.00\times10^{-7}\leqslant F_{\mathrm{X}}<1.00\times10^{-6}$
C	$1.00\times10^{-6}\leqslant F_{\mathrm{X}}<1.00\times10^{-5}$	M	$1.00\times10^{-5}\leqslant F_{\mathrm{X}}<1.00\times10^{-4}$
X	$F_{\mathrm{X}}\geqslant1.00\times10^{-4}$	–	–

表 2 基于 *Kp* 指数的地磁暴强度等级划分

Table 2 Classification of geomagnetic storm intensity based on *Kp* index

强度级别	*Kp*指数范围	强度级别	*Kp*指数范围	强度级别	*Kp*指数范围
平静	$Kp<4$	活跃	$Kp=4$	小地磁暴	$Kp=5$
中等地磁暴	$Kp=6$	大地磁暴	$Kp=7$	严重地磁暴	$Kp=8$
极端地磁暴	$Kp=9$	–	–	–	–

表 3　空间天气事件预报概率阈值

Table 3　Probability thresholds for space weather event forecasting

事件	M级耀斑	X级耀斑	小地磁暴	中等以上地磁暴	太阳质子事件
概率阈值	$P\geqslant 35\%$	$P\geqslant 20\%$	$P\geqslant 35\%$	$P\geqslant 20\%$	$P\geqslant 20\%$

注　M级耀斑和小地磁暴采用35%阈值, X级耀斑、中等以上地磁暴和太阳质子事件采用20%阈值.

表 4　空间天气事件概率预报检验列联表

Table 4　Contingency table for probabilistic space weather forecast verification

预报 观测	有	无
有	命中次数 (N_k^{A})	漏报次数 (N_k^{C})
无	虚报次数(N_k^{B})	正确否定次数(N_k^{D})

$$\delta_{\mathrm{FAR}}=\frac{N_k^{\mathrm{B}}}{N_k^{\mathrm{A}}+N_k^{\mathrm{B}}}. \tag{2}$$

(3) 正确率 (PC, δ_{PC}) 计算预报总体正确比例, 其表达式为

$$\delta_{\mathrm{PC}}=\frac{N_k^{\mathrm{A}}+N_k^{\mathrm{D}}}{N_k^{\mathrm{A}}+N_k^{\mathrm{B}}+N_k^{\mathrm{C}}+N_k^{\mathrm{D}}}. \tag{3}$$

(4) 预报偏差 (Bias, δ_{Bias}) 从统计总量上判断预报频率与观测频率的比率, 揭示系统性高估 ($\delta_{\mathrm{Bias}}>1$) 或低估 ($\delta_{\mathrm{Bias}}<1$), 其表达式为

$$\delta_{\mathrm{Bias}}=\frac{N_k^{\mathrm{A}}+N_k^{\mathrm{B}}}{N_k^{\mathrm{A}}+N_k^{\mathrm{C}}}. \tag{4}$$

其中, 若分母为零, δ_{Bias}标记为 Inf; 若分子分母均为零, δ_{Bias}则标记为 NaN.

(5) TS 评分 (Threat Score, δ_{TS}) 综合命中、漏报和空报, 直接衡量对目标事件的预报精准度, 其表达式为

$$\delta_{\mathrm{TS}}=\frac{N_k^{\mathrm{A}}}{N_k^{\mathrm{A}}+N_k^{\mathrm{B}}+N_k^{\mathrm{C}}}. \tag{5}$$

(6) TSS 技巧评分 (True Skill Statistic, δ_{TSS}) 通过命中率与误报概率之差量化模型判别事件是否发生的能力, 其计算不受事件气候频率影响, 在样本不平衡情况下更为稳健[28], 其表达式为

$$\delta_{\mathrm{TSS}}=\frac{N_k^{\mathrm{A}}\cdot N_k^{\mathrm{D}}-N_k^{\mathrm{B}}\cdot N_k^{\mathrm{C}}}{\left(N_k^{\mathrm{A}}+N_k^{\mathrm{C}}\right)\left(N_k^{\mathrm{B}}+N_k^{\mathrm{D}}\right)}. \tag{6}$$

2.2　概率预报特性诊断指标

为克服单一阈值分析的局限, 引入基于概率分布特性的综合分析框架, 从可靠性、锐度与分辨率三个维度对预报系统进行深层诊断. 可靠性衡量预报概率与事件实际发生频率的一致性, 反映系统的校准水平. 锐度描述预报概率向 0 或 1 两端集中的趋势, 体现预报的确定性程度. 分辨率表征系统将样本划分为具有不同发生频率子集的能力, 即区分事件发生与否的统计效能. 本文采用可靠性图、相对作用特征 (ROC) 曲线及区分图作为核心分析工具.

可靠性图将预报概率划分为若干区间, 以区间中值为横坐标, 区间内事件实际发生频率为纵坐标绘图. 理想情况下, 数据点应落在对角线上; 曲线位于对角线下方表明过度预报, 反之则为预报不足. 曲线的平坦程度反映分辨率高低[28]. 相对作用特征 (ROC) 曲线: 通过选取一系列概率阈值, 绘制命中率 (POD) 随虚警率变化的轨迹. 曲线下面积 (AUC) 是量化判别能力的核心指标, 介于 0.5(无技巧) 至 1.0(完美) 之间. ROC 曲线对预报的整体偏差 (Bias) 不敏感, 常被用于衡量预报系统的潜在有用性[29]. 区分图: 叠加显示事件发生与事件未发生两类样本中预报概率的频率分布密度, 直接呈现系统对两类状态的分离程度. 分布重叠面积越小, 表明系统的判别效能越高[25]. 这三种方法互为补充, 可靠性图诊断校准质量, ROC 曲线与区分图量化判别潜力, 共同构成对概率预报系统的全方位性能分析.

3　事件预报检验分析

对 NCSW 自 2004 年开展业务以来的空间天气事件预报水平进行了分析. 图 1 给出了 2005－2022 年业务期间观测到的耀斑与地磁暴事件的累积天数, 以及同期太阳黑子数的变化曲线. 该时段主要涵盖了第 23 太阳活动周的下降期 (2005－2008 年)、第 24 太阳活动周的完整周期 (2009－2019 年) 以及第 25 太阳活动周的上升期 (2020－2022 年). 太阳黑子数与耀斑发生频次具有显著正相关性, 而地磁暴的发生不仅与 CME 相关, 还受冕洞高速太阳风以及由此

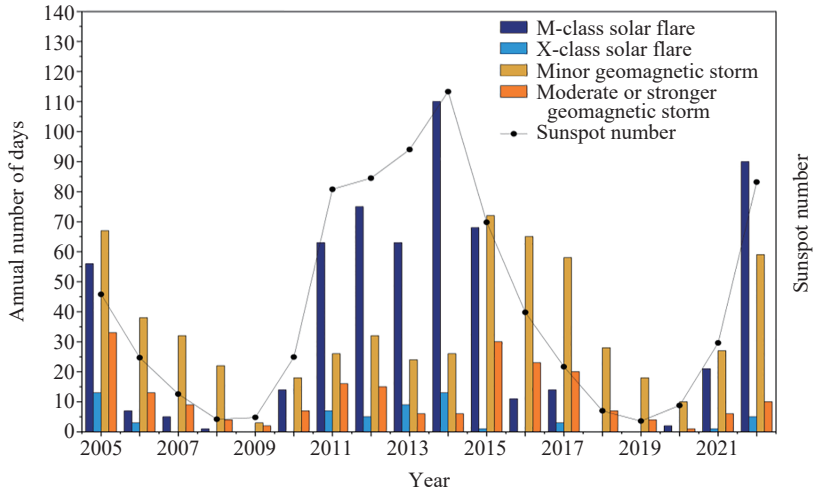


图 1 2005−2022 年耀斑及地磁暴事件天数与太阳黑子数的变化

Fig. 1 Variations of flare and geomagnetic storm event days and sunspot number from 2005 to 2022

产生的共转相互作用区 (CIR) 影响, 后者在太阳活动下降年更为频繁, 使得地磁暴呈现高年和下降年双峰值分布特征[30–32].

第 24 太阳活动周被观测记录为近百年来最弱的太阳周之一, 其太阳黑子数峰值 (约 116) 较第 23 周下降约 30%, 且经历了 2008−2009 年极深的太阳活动极小期, 因此该周期内 X 级和 M 级耀斑的总体发生频次较第 23 周明显减少[33]. 相比之下, 地磁暴在 2015−2017 年 (第 24 周下降期) 仍维持较高频次, 这与 CIR 驱动的地磁暴在该时段的贡献增强有关. 特别是在 2009 年, 全年未发生 M 级及以上耀斑, 仅记录到三次小地磁暴和二次中等及以上地磁暴, 反映出极端低活动年份的事件特征.

3.1 M 级与 X 级耀斑事件检验

基于 NCSW 2005−2022 年 M 级 (见图 2) 和 X 级 (见图 3) 耀斑概率预报数据, 采用 Bias, POD, FAR, PC(Proportion of Correct), TS 和 TSS 等指标, 对 1～3 天预报时效的预报性能进行统计分析. 图 2 和图 3 中橙、蓝、绿色依次表示第 1, 2, 3 天预报时效, 图 2 中灰色表示该年份因 M 级耀斑发生天数为 0 而无法计算分析指标. 图 3 中灰色表示该年份因 X 级耀斑发生天数为 0 而无法计算分析指标, 白色方形方块表示 Bias 为无穷大 (Inf).

样本特征. 根据预报统计稳定性要求, 设定 M 级耀斑年发生天数≥50 天为样本充足, 1～49 天为样本不足, 0 天为零事件年; X 级耀斑≥5 天为样本相对充足, 1～4 天为样本不足, 0 天为零事件年. 据此, M 级充足年份为 2005, 2011−2015, 2022(56～110 天), 不足年份为 2006−2010, 2016−2021(1～21 天), 零事件年份为 2009, 2018−2019; X 级相对充足年份为 2005, 2011−2014, 2022(5～13 天), 不足年份为 2006, 2015, 2017, 2021(1～3 天), 零事件年份为 2007−2010, 2016, 2018−2020.

M 级耀斑预报性能. 第 1 天预报在样本量较充足的年份 (2005, 2011−2015, 2022) 表现良好, TS 为 0.23～0.46, TSS 为 0.23～0.60, POD 为 0.29～0.70, FAR 为 0.39～0.56, Bias 为 0.52～1.20. 第 2～3 天预报技巧随时效延长有所降低, TS 为 0.08～0.36, TSS 为 0.02～0.46, POD 降至 0.12～0.57, FAR 升至 0.36～0.77. 在样本量不足的年份 (2006−2010, 2016−2021), 部分年份 (2008, 2010, 2020)TS 为 0, TSS 甚至出现负值, 表明极低年基本无预报能力. 零事件年份 (2009, 2018−2019) 不存在空报. PC 与发生天数呈显著负相关, 样本充足年份 PC 为 0.73～0.88, 不足年份达 0.94～1.00, 反映 PC 的基础率依赖性.

X 级耀斑预报性能. 第 1 天预报在发生天数相对较多的年份, 2005 年 (13 天)TSS 为 0.74, TS 为 0.13; 2013 年 (9 天)TSS 为 0.66, TS 为 0.09; 2014 年 (13 天)TSS 为 0.43, TS 为 0.08. 第 2～3 天预报技巧呈现波动, 2005 年第 1～2 天 TSS 稳定在 0.74, 第 3 天降至 0.58; 2013 年第 3 天 TSS(0.69) 高于第 1 天 (0.66). 2007−2008, 2016, 2020 年存在空报. NCSW 采用宁可空报不漏报策略, 有效检验年份 POD 为 0.87～

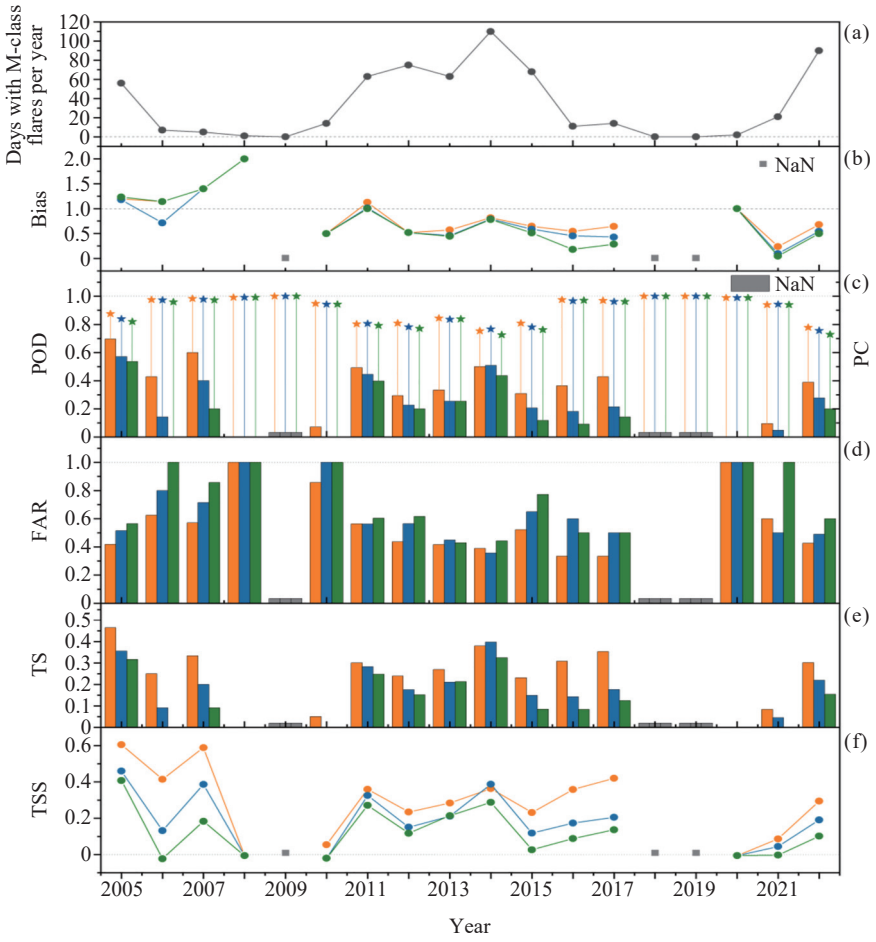


图 2　国家空间天气监测预警中心 (NCSW)M 级耀斑概率预报检验分析 (2005−2022 年). (a) M 级耀斑发生天数, (b) 预报偏差 (Bias), (c) 命中率 (POD, 柱状) 与正确率 (PC, 星号), (d) 空报率 (FAR), (e) 威胁评分 (TS), (f) 真实技巧评分 (TSS)

Fig. 2　Verification of M-class solar flare probability forecasts from the National Center for Space Weather (NCSW) during 2005−2022. (a) Annual number of days with M-class flares, (b) Frequency bias (Bias), (c) Probability of Detection (POD, bars) and Percent Correct (PC, asterisks), (d) False Alarm Ratio (FAR), (e) Threat Score (TS), (f) True Skill Statistic (TSS)

1.00, FAR 为 0.75～0.98.

M 级与 X 级对比. M 级 Bias 为 0.52～1.20, 除个别年份 (2005−2008 年) 外均小于或接近 1, 反映大多数年份预报不足; X 级 Bias 为 5.5～14.5(2015 年样本量为 1 时 Bias 达 47.5), 明显大于 1, 显示严重过度预报. 该差异源于两方面: 一是 X 级发生天数极少 (5～13 天), 概率预报易产生过度预报; 二是业务策略差异, M 级采取保守策略, X 级采取宁可虚报不可漏报原则.

M 级保守策略导致 POD 较低 (0.29～0.70)、漏报较多; X 级激进策略导致 POD 极高 (0.87～1.00)、漏报少, 但 FAR 极高. M 级 TS 与 TSS 变化趋势基本一致, X 级二者差距显著且趋势分离: 第 1 天预报 2005 年 TSS 与 TS 相差 0.61, 2011 年二者相差 0.51, 2014 年二者相差 0.35, TSS 为 0.30～0.97, TS 仅为 0.01～0.23. 该分离现象源于 X 级极低样本量: TS 仅依赖命中、漏报和空报, 在总样本 365 天中, 命中和空报占比极低, 对发生日判别能力不敏感; TSS 通过命中率与虚警率 (POFD) 之差剔除随机预报基准贡献, 纳入正确否定信息后可反映真实预报技巧. 例如 2005 年 POD 为 1.00, FAR 为 0.87, TS 仅 0.13, 而 TSS 达 0.74.

M 级 Bias 小于 1 而 X 级 Bias 大于 1, 导致 M 级 FAR 低于 X 级, 但 M 级 TSS 反而低于 X 级, 揭

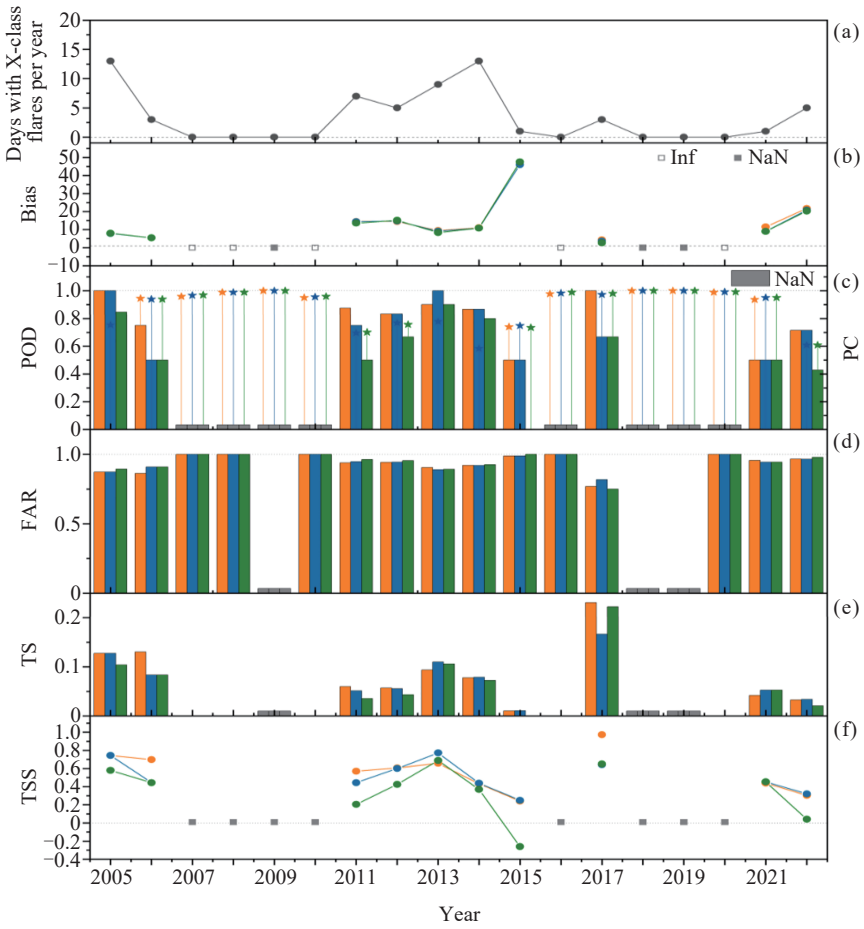


图 3 国家空间天气监测预警中心 (NCSW)X 级耀斑概率预报检验评分析 (2005－2022 年). (a) X 级耀斑发生天数, (b) 预报偏差 (Bias), (c) 命中率 (POD, 柱状) 与正确率 (PC, 星号), (d) 空报率 (FAR), (e) 威胁评分 (TS), (f) 真实技巧评分 (TSS)

Fig. 3 Verification of X-class solar flare probability forecasts from the National Center for Space Weather (NCSW) during 2005－2022. (a) Annual number of days with X-class flares, (b) Frequency Bias (Bias), (c) Probability of Detection (POD, bars) and Percent Correct (PC, asterisks), (d) False Alarm Ratio (FAR), (e) Threat Score (TS), (f) True Skill Statistic (TSS)

示预报策略与分析指标的复杂关系. M 级保守策略降低空报率, 但漏报增加导致命中率下降, TSS 受损; X 级激进策略虽然空报率极高, 但漏报极少, 命中率接近 1, TSS 得到提升. 然而, X 级 TS 极低, 表明在命中、漏报、空报构成的评价框架下, 其预报技巧实际较低, TSS 的高值部分源于正确否定的贡献.

综上, NCSW M 级耀斑第 1 天预报在样本量充足年份具备中等预报技巧, TSS 达 0.23～0.60, 呈现低 Bias、低命中率、低空报率、中等 TSS 特征, 保守策略导致漏报较多; X 级预报在发生天数相对较多的年份 TSS 可达 0.43～0.74, 呈现高 Bias、高命中率、高空报率、高 TSS、低 TS 特征, 激进策略导致空报严重. TSS 较 TS 更能消除基础率影响, 适合作为稀有事件预报的核心检验指标. 整体而言, NCSW 耀斑概率预报系统对 M 级和 X 级事件均具备业务可用水平, 但两类事件的预报策略和检验特征存在显著差异, M 级需提升命中率以减少漏报, X 级需在保持高命中率的同时适当降低空报率.

3.2 质子事件检验与分析

质子事件与 X 级耀斑均采用逐日概率预报模式, 但二者在预报策略与检验方式上存在显著差异. 2005－2022 年间, X 级耀斑年发生天数约 60 天, 质子通量达到质子事件水平累计天数约 141 天. X 级耀斑预报以宁可虚报不可漏报为原则, FAR 为 0.75～

0.98; 质子事件概率预报则执行零虚报策略, FAR 为 0. 该差异由两类事件的影响效应、预报基础及业务决策需求决定.

质子事件直接威胁航天员安全、在轨航天器运行以及航空通信质量, 针对强质子事件, 需要采取航线调整、航天器规避、中止航天员出舱等高代价应对措施, 空报带来的业务与安全成本极高. 同时, 质子事件涉及 CME 传播、行星际磁场连接性、粒子加速与输运等多环节耦合过程, 提前 1～3 天预报不确定性较大. 因此业务预报采用保守预警策略, 仅在观测到明确太阳爆发及高置信度粒子加速信号时发布警报, 从而实现 FAR 为 0.

预报检验方式的差异主要源于零虚报约束. X 级耀斑因存在空报, 各检验指标物理意义独立, 可逐年统计并分析年际变化. 质子事件概率预报 FAR 为 0 时, POD, Bias, TS 和 TSS 在数学上等价, 指标体系退化, 难以通过逐年样本有效刻画预报性能. 为此, 质子事件预报采用 2005－2022 年全时段样本进行整体统计分析, 分析结果以综合指标表格给出.

表 5 中的 2005－2022 年分析结果显示, 在零虚报条件下, POD, Bias, TS, TSS 取值一致, 预报能力随时效延长快速衰减: 24, 48, 72 h TSS 分别为 0.70, 0.43, 0.19. 预报技巧下降的主要原因为漏报率随时效增长显著上升, 这与 CME 及粒子传播过程的预报不确定性累积密切相关. 各时效 PC 分别为 0.994, 0.988, 0.983, 均接近 0.99, 随时效延长略有下降但变化幅度较小. 该特征在 FAR 为 0 条件下, PC 主要受事件基础率 (约 2.1%) 主导, 质子事件的稀有性使得正确否定样本占比极高, 导致 PC 对预报技巧的实际变化不敏感. 结果表明, 当前质子事件 24 h 预报具备一定技巧, 而 48～72 h 时效受物理过程认知与预报能力限制, 漏报问题突出, 预报性能明显降低.

3.3 地磁暴事件检验与分析

基于 NCSW 2005－2022 年地磁暴概率预报数据, 采用 Bias, POD, FAR, PC, TS 和 TSS 等指标, 对 1～3 天预报时效的性能进行统计分析.

样本特征. 小地磁暴年发生天数 3～72 天, 第 23 周下降段 (2005－2008 年) 为 22～67 天, 2009 年极小年骤降至 3 天; 第 24 周上升段 (2009－2014 年) 为 18～26 天, 峰段 (2012－2014 年) 升至 24～32 天; 第 24 周下降段 (2015－2019 年) 仍保持 18～72 天; 第 25 周上升段初期 (2020－2022 年) 回落至 10～59 天. 中等及以上地磁暴年发生天数 1～33 天, 约为小地磁暴年发生天数的 1/3, 第 24 周峰段 (2012－2014 年) 为 6～15 天, 第 23 周下降段末期 (2007－2009 年), 第 24 周上升段初期 (2010 年) 及第 25 周初期 (2020 年) 普遍低于 10 天, 2020 年仅 1 天, 第 24 周下降段初期 (2015－2017 年) 为 20～33 天.

小地磁暴预报性能. 第 1 天预报技巧随太阳活动周演进显著提升. 第 23 周下降段 (2005－2008 年) 样本较充足 (22～67 天), 但技巧偏低且波动大, TSS 为 0.00～0.19, TS 为 0.00～0.19, 2007 年 (22 天)TSS 为 –0.003, 反映早期能力有限. 第 24 周上升段及峰段 (2010－2014 年) 样本 18～32 天, 技巧回升, TSS 为 0.16～0.25, TS 为 0.13～0.26. 第 24 周下降段 (2015－2019 年), 由于样本较为丰富 (2015 年 72 天), 预报技巧达到峰值, TSS 为 0.25～0.37, TS 为 0.23～0.34; 技能评分并非随样本量单调递增, 2015 年 Bias, POD, FAR, TS, TSS 分别为 0.54, 0.31, 0.44, 0.25, 0.25, 而 2019 年 (19 天) 分别为 0.83, 0.39, 0.53, 0.27, 0.37. 第 25 周上升段 (2020－2022 年) 样本减少, 但技巧保持较高水平, TSS 为 0.14～0.23, TS 为 0.15～0.23, FAR 为 0.50～0.67, 优于早期同等样本量年份. 剔除前期 (2006－2009 年) 性能极差年份, 第 1 天 TSS 为 0.17～0.39, TS 为 0.14～0.32, POD 为 0.19～0.47, FAR 为 0.39～0.88, Bias 为 0.13～0.84. 第 2～3 天技巧随时效延长降低, TSS 为 0.02～0.30, TS 为 0.03～0.28, POD 降至 0.04～0.36, FAR 升至 0.00～

表 5　NCSW 太阳质子事件概率预报性能分析 (2005－2022 年)

Table 5　Performance of NCSW solar proton event forecasts (2005－2022)

太阳质子事件预报 / h	TSS评分	TS评分	报准率(POD)	偏差(Bias)	虚报率(FAR)	正确率(PC)
24	0.70	0.70	0.70	0.70	0	0.994
48	0.43	0.43	0.43	0.43	0	0.988
72	0.19	0.19	0.19	0.19	0	0.983

0.86. PC 与发生天数呈负相关 (3～72 天对应 0.82～0.99), 反映基础率依赖性, 不宜作为独立评价指标. 图 4 为国家空间天气监测预警中心 (NCSW) 小地磁暴 (Kp=5) 概率预报检验分析 (2005－2022 年),图中酒红、深蓝、深青依次表示第 1, 2, 3 天预报时效, 灰色表示该年份因小地磁暴发生天数过少导致检验指标无法计算.

中等及以上地磁暴预报性能. 第 1 天预报技巧随时效延长系统性衰减, 且在不同太阳活动阶段差异显著. 发生天数≤10 天 (2007－2010 年, 2013－2014 年, 2018－2022 年), 除 2013 年 (6 天)POD, TS, TSS 分别为 0.33, 0.22, 0.32 外, 其他年份 TS 和 TSS 均<0.1. 样本>10 天 (主要为 20～30 天) 的第 23 周下降段中期 (2005－2006 年) 和第 24 周下降段初期 (2015－2017 年), 第 1 天 TS 和 TSS 为 0.07～0.20, POD 为 0.08～0.20, 总体>0.1; 第 2～3 天 TS 为 0～0.17, TSS 为−0.003～0.18, POD 为 0～0.20, 总体<0.1. PC 维持在 0.92～0.99 高位, 源于无事件日占比极高 (>90%) 的基础率效应, 无法真实反映捕捉能力. TS 总体为 0.00～0.20. 2005－2022 年尚未形成稳定预报能力 (TSS≤0.2). 图 5 为国家空间天气监测预警中心 (NCSW) 的中等及以上地磁暴 (Kp≥6) 概率预

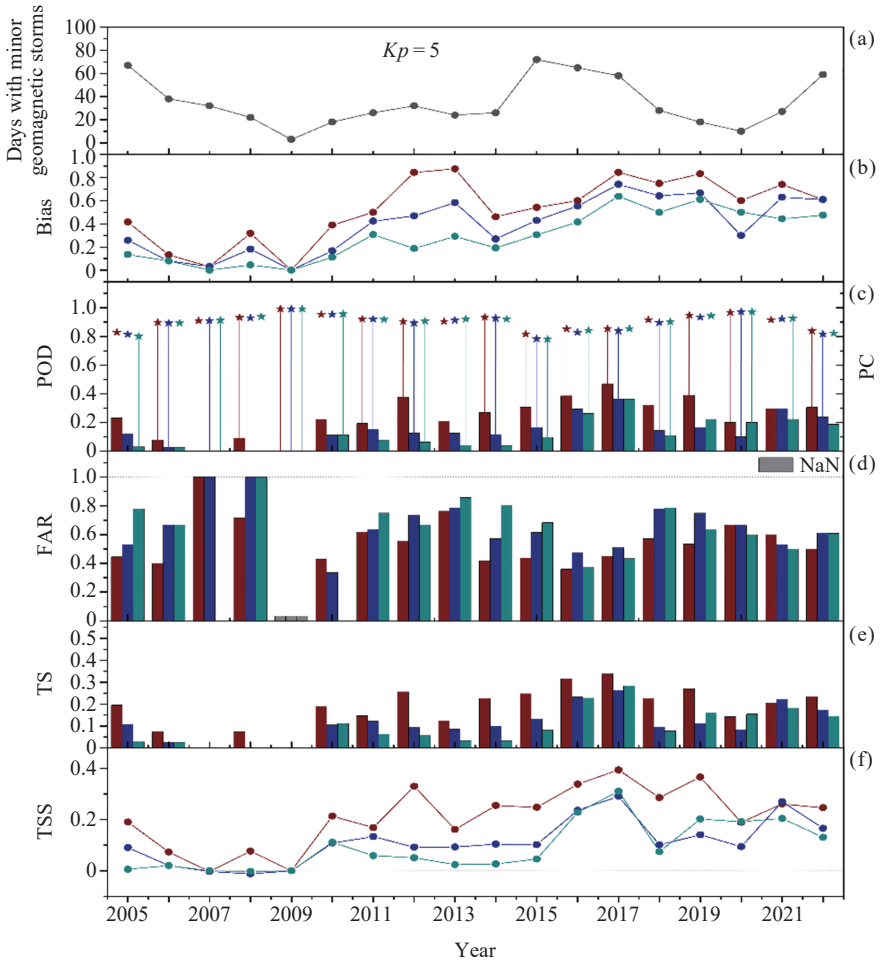


图 4 国家空间天气监测预警中心 (NCSW) 小地磁暴 (Kp=5) 概率预报检验分析 (2005－2022 年). (a) 小地磁暴发生天数, (b) 预报偏差 (Bias), (c) 命中率 (POD, 柱状) 与预报准确率 (PC, 星号), (d) 空报率 (FAR), (e) 威胁评分 (TS), (f) 真实技巧评分 (TSS)

Fig. 4 Verification of minor geomagnetic storm (Kp=5) probability forecasts from the National Center for Space Weather (NCSW) during 2005－2022. (a) Annual number of days with minor geomagnetic storms, (b) Frequency Bias (Bias), (c) Probability of Detection (POD, bars) and Percent Correct (PC, asterisks), (d) False Alarm Ratio (FAR), (e) Threat Score (TS), (f) True Skill Statistic (TSS)

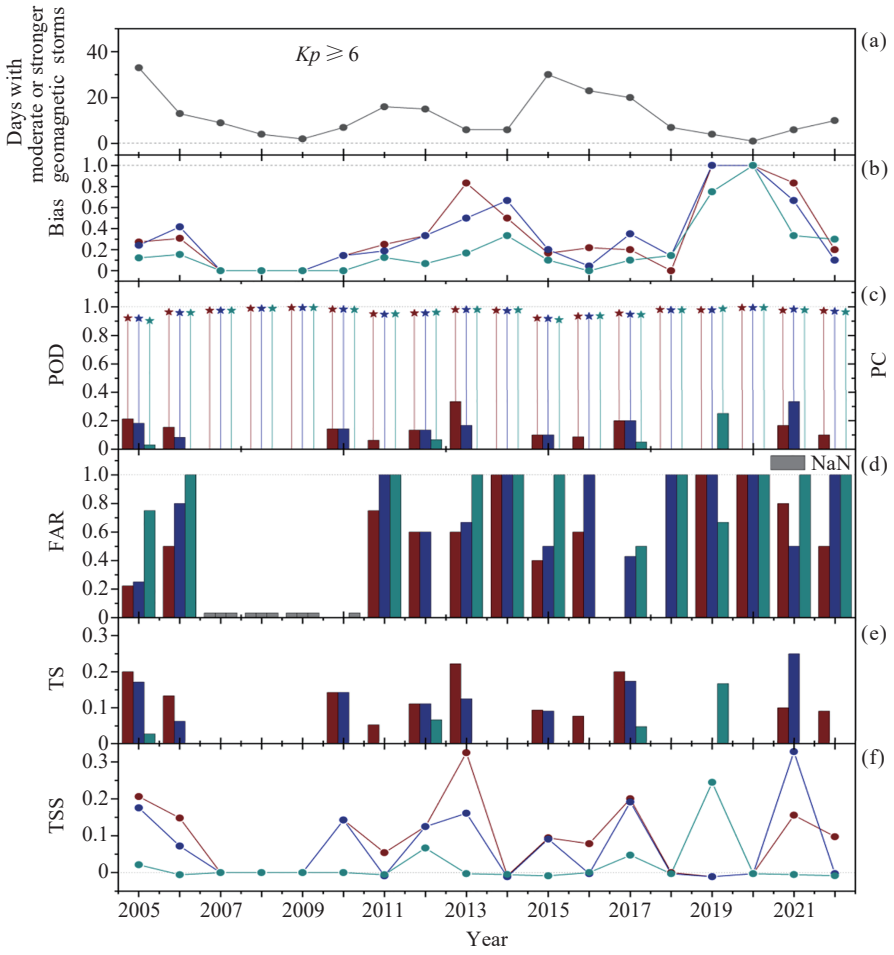


图 5　国家空间天气监测预警中心 (NCSW) 的中等及以上地磁暴 ($Kp \geqslant 6$) 概率预报检验分析 (2005－2022 年). (a) 中等及以上地磁暴发生天数, (b) 预报偏差 (Bias), (c) 命中率 (POD, 柱状) 与预报准确率 (PC, 星号), (d) 空报率 (FAR), (e) 威胁评分 (TS), (f) 真实技巧评分 (TSS)

Fig. 5　Verification of moderate or stronger geomagnetic storm probability forecasts from the National Center for Space Weather (NCSW) during 2005－2022. (a) Annual number of days with moderate or stronger geomagnetic storms, (b) Frequency Bias (Bias), (c) Probability of Detection (POD, bars) and Percent Correct (PC, asterisks), (d) False Alarm Ratio (FAR), (e) Threat Score (TS), (f) True Skill Statistic (TSS)

报检验分析 (2005－2022 年), 图中酒红、深蓝、深青依次表示第 1, 2, 3 天预报时效, 灰色表示该年份因中等及以上地磁暴发生天数过少导致检验指标无法计算.

小地磁暴与中等及以上地磁暴对比: 小地磁暴 Bias 为 0.13～0.84, 以<1 为主, 显示预报不足; 中等及以上地磁暴 Bias 为 0～0.83, 均<1, 预报不足较为严重. 小地磁暴 FAR 为 0.39～0.88, 中等及以上地磁暴大多 FAR<0.6, 未出现因追求命中率导致的 FAR 飙升 (FAR=1.0 仅在无命中时出现). 差异源于样本量与预报策略: 小地磁暴样本相对充足 (3～72 天), 策略相对积极; 中等及以上地磁暴样本稀少 (1～33 天), 策略较为保守, 因未能充分捕捉真实事件 (POD 为 0.2～0.4) 导致虚报数相对减少. 小地磁暴 TSS 与 TS 变化趋势基本一致, 中等及以上地磁暴二者均处于低值且波动大.

跨周期能力演变. 控制样本量 (20～30 天) 进行跨相位对比, 第 23 周下降段 2008 年 (22 天)TSS 为 0.08, TS 为 0.07; 第 24 周上升段 2011 年 (26 天)TSS 为 0.17, TS 为 0.15; 第 24 周下降段 2018 年 (28 天) TSS 为 0.29, TS 为 0.23; 第 25 周上升段 2021 年 (27 天)TSS 为 0.26, TS 为 0.21. 该序列表明, 排除样本

量差异后, 同等样本条件下 2015 年后较 2010 年前 TSS 和 TS 均有所提升, 证实小地磁暴预报能力得到实质性提升, 而非依赖于太阳活动高年的样本红利.

综上所述, NCSW 小地磁暴第 1 天预报的 TSS 均值为 0.25, 第 2～3 天该指标呈现下降趋势, 但整体表现出长期的改进特征. 对于中等及以上级别的地磁暴, 其预报 TSS 值未超过 0.2, 且在低发年及 2～3 天预报时效下, 该指标趋近于 0. 在偏差 (Bias) 方面, 两类事件的 Bias 均<1, 表明当前预报存在漏报现象. 此外, 由于 PC 指标受基础率影响在低发年出现虚高, 而 TSS 指标能有效消除基础率的依赖效应, 因此 TSS 更适合作为地磁暴预报的检验指标.

4 概率预报统计性能对比

NCSW 与 SWPC 预报数据的发布时刻存在明显差异. SWPC 在前一日 22:30 UT 发布数据, NCSW 于当日 01:30 UT 发布数据; 二者预报数据对应的时段起始时刻 (起报时间) 均为当日 00:00 UT. 该时间差使 NCSW 可利用更接近预报时段的观测资料, 在信息时效性上具有潜在优势; 同时两机构在概率阈值设定上亦存在差异. 若采用基于具体事件是否发生的事件检验分析方法, 则难以区分性能差异究竟来源于预报能力本身, 还是源于概率阈值与发布时间的不同设置. 因此, 本文采用前文所述基于概率分布特性的分析方法, 从可靠性、锐度与分辨率三个维度, 利用可靠性图、ROC 曲线及区分图, 仅对发生率较高的 M 级耀斑和小地磁暴 (均约为 9.1%) 开展对比分析, 而未对 X 级耀斑、质子事件和中等及以上地磁暴 (均低于 5%, 分别约为 0.9%, 2.1%, 3.1%) 进行分析.

M 级耀斑分析结果 (图 6) 表明, 在可靠性方面 (图 6a), NCSW 的预报曲线与对角线贴合更紧密, 尤其在中等概率区间 (0.2～0.6) 匹配更好; SWPC 呈现轻微系统性偏差, 在低概率区间 (<0.3) 倾向于预报不足, 在中高概率区间偶有过度预报. 在分辨能力方面 (图 6c), 两机构均能有效区分事件是否发生, SW-

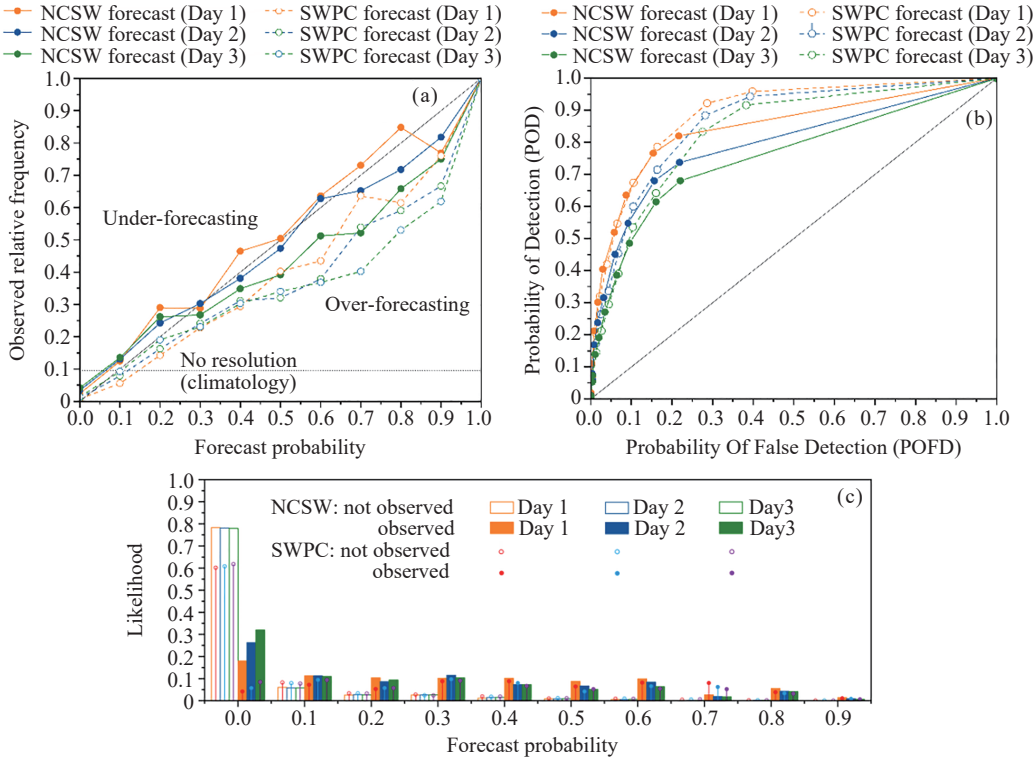


图 6 国家空间天气监测预警中心 (NCSW) 与美国空间天气预报中心 (SWPC) 的 M 级耀斑概率预报统计性能对比 (2005－2022 年). (a) 可靠性图, (b) ROC 曲线, (c) 区分图

Fig. 6 Statistical performance comparison of M-class solar flare probability forecasts between the National Center for Space Weather (NCSW) and the Space Weather Prediction Center (SWPC) for the period 2005－2022. (a) Reliability diagram, (b) ROC curve, (c) discrimination diagram

PC 对高概率事件 (>0.7) 的集中度略高, 显示其预报锐度较强; NCSW 在低概率区间 (<0.2) 的集中度更显著, 且两类样本分布重叠更小, 表明其在控制虚报、提升判别清晰度方面表现更好. ROC 曲线 (图 6b) 显示两机构技巧均显著高于无技巧线; 第 1 天预报中 NCSW 的 ROC 曲线更靠近左上角, 综合判别能力略优于 SWPC, 这可能与其晚发布 3 h、能够获得更多临近观测信息有关. 图 6 中实心与空心标记分别对应 NCSW 与 SWPC 预报; 橙、蓝、绿色依次表示第 1, 2, 3 天预报时效. 随着预报时效延长至第 2～3 天, 两机构性能差异逐渐缩小, 表明晚发布所带来的信息优势随预报时效增加而减弱.

小地磁暴分析结果 (图 7) 呈现类似特点. 在可靠性方面 (图 7a), NCSW 在各时效下曲线与对角线贴合较紧, 尤其在 0.3～0.6 概率区间内一致性较好; SWPC 呈现一定系统性偏差, 在低概率段多位于对角线上方 (预报不足), 中高概率段偶有过度预报. 在分辨能力方面 (图 7c), SWPC 对高概率事件的预报更为集中, 体现一定的锐度优势; NCSW 则在低概率区间表现出更明显的集中趋势, 且发生与未发生样本的重叠区域较小, 说明其在控制虚报、提升分类清晰度方面更具优势. ROC 曲线 (图 7b) 证实两机构均具有较高预报技巧, 第 1 天预报中 NCSW 表现略优, 随着预报时效延长, 两机构性能差异减小. 图 7 中实心与空心标记分别对应 NCSW 与 SWPC 预报; 酒红、深蓝、灰色依次表示第 1, 2, 3 天预报时效.

综上所述, NCSW 在概率校准可靠性 (尤其在中等概率区间)、低概率区间的集中度及短期 (24 h) 综合判别能力上表现更优 (可能有 3 h 信息优势), 预报风格较稳健; SWPC 则在对高概率信号 (>0.7) 的捕捉上锐度突出, 但在概率校准上存在一定系统性偏差. 两机构在短期预报性能上的差异, 部分可归因于发布时间不同导致的数据基础差异; 随着预报时效延长, 初始条件的影响减弱, 双方表现更为接近.

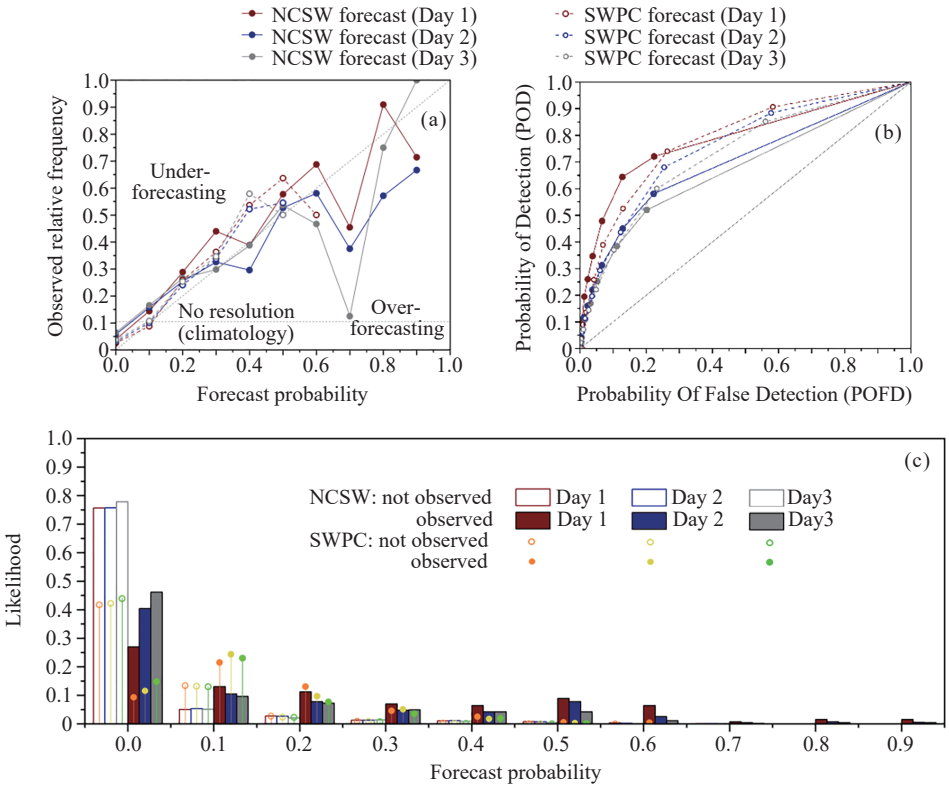


图 7 国家空间天气监测预警中心 (NCSW) 与美国空间天气预报中心 (SWPC) 的地磁暴 ($Kp\geqslant5$) 概率预报统计性能对比 (2005－2022 年). (a) 可靠性图, (b) ROC 曲线, (c) 区分图

Fig. 7 Statistical performance comparison of geomagnetic storm ($Kp\geqslant5$) probability forecasts between the National Center for Space Weather (NCSW) and the Space Weather Prediction Center (SWPC) for the period 2005－2022. (a) Reliability diagram, (b) ROC curve, (c) discrimination diagram

5 结论

针对国家空间天气监测预警中心 (NCSW) 2005—2022 年太阳耀斑、太阳质子事件和地磁暴三类关键空间天气事件的 1～3 天概率预报数据, 构建了综合检验方案并开展系统性分析. 主要结论如下.

(1) 综合检验方案采用多维指标集 (POD, FAR, Bias, PC, TS 和 TSS) 进行单中心内部性能诊断, 针对各机构概率阈值差异引入可靠性–锐度分析方法实现跨机构公平对比, 二分类评分适用于逐年业务监测, 统计特性分析适用于长期性能分析.

(2) NCSW 预报数据性能方面, M 级耀斑第 1 天技巧较好 (威胁评分 0.23～0.46, 真实技巧评分 0.23～0.60), X 级耀斑第 1 天在多发年具备一定技巧 (真实技巧评分 0.43～0.74), 但第 2～3 天技巧明显下降; 太阳质子事件 24 h 预报具备一定技巧 (真实技巧评分≈0.70), 48～72 h 快速衰减 (真实技巧评分降至 0.19); 小地磁暴第 1 天真实技巧评分均值约为 0.25, 中等及以上地磁暴尚未达到有效门槛 (真实技巧评分≤0.2), 两类地磁暴均处于预报不足状态 (预报偏差远小于 1), 漏报率偏高.

(3) NCSW 与 SWPC 对比表明, NCSW 在概率校准可靠性及短期判别能力上略优, SWPC 在高概率信号捕捉上锐度突出, 随预报时效延长二者差异逐渐缩小.

上述结论的适用性需作如下说明. 真实技巧评分在太阳耀斑预报领域被广泛采用[34], 较威胁评分更能消除基础率影响, 但极端稀疏事件中趋近于命中率[35], 对空报区分能力受限, 本文质子事件和 X 级耀斑分析结果印证了这一局限性, 故需结合多指标综合判断. 逐年分类评分受样本量波动影响显著, 极低年份指标存在无法计算或波动剧烈的现象; 逐年可靠性分析中概率分箱内样本稀疏, 统计代表性受限, 此系样本量约束而非方法缺陷[36, 1]. 跨机构对比中, 各中心因业务需求存在发布时间差异, NCSW 可利用更接近预报时段的观测资料, 本文旨在检验业务发布的实际预报水平, 未对此差异进行处理, NCSW 第 1 天略优部分可能与此信息优势有关, 随预报时效延长该差异逐渐缩小. 未来研究可结合机器学习识别预报误差源, 为模式改进与业务化应用提供参考.

致谢 美国国家海洋和大气管理局空间天气预报中心 (NOAA/SWPC) 预报及观测数据来源于以下公开渠道: 太阳耀斑、太阳质子事件和地磁暴概率预报数据 (RSGA) 通过 SWPC 网站 (https://www.swpc.noaa.gov/products/3-day-forecast) 获取太阳耀斑观测数据 (DGD) 来源于 GOES 卫星软 X 射线观测数据 (https://www.swpc.noaa.gov/products/goes-x-ray-flux); 太阳质子事件观测数据 (DPD) 来源于 GOES 卫星质子通量观测数据 (https://www.swpc.noaa.gov/products/goes-proton-flux); 地磁暴观测数据 (DSD) 来源于 *Kp* 指数观测数据 (https://www.swpc.noaa.gov/products/planetary-k-index).

医学伦理 研究不涉及人体或动物实验.

利益冲突 所有作者均声明不存在利益冲突.

参考文献

[1] WANG Jingxiu. Explosive solar activity and disastrous space weather[J]. *Publications of Purple Mountain Observatory*, 2003, **22**(1): 48-57 (汪景琇. 太阳剧烈活动与空间灾害天气 [J]. 紫金山天文台台刊, 2003, **22**(1): 48-57)

[2] WANG Shengguo, XIE Zhihui, ZHAO Zengliang, *et al*. Influence on national security of disastrous space weather[C]//National Security Geophysics Series (9): Disaster Prevention and Reduction and National Security. Xi'an: Xi'an Map Publishing House, 2013: 37-42 (王胜国, 谢志辉, 赵增亮, 等. 灾害性空间天气对国家安全的影响与对策 [C]//国家安全地球物理丛书 (九)——防灾减灾与国家安全. 西安: 西安地图出版社, 2013: 37-42)

[3] LI Xiaoming. Research and Implement of Space Weather Power Grid Geomagnetic Storm Monitoring System[D]. Baoding: North China Electric Power University, 2013 (李小明. 空间天气电网磁暴灾害监测系统的研究与实现 [D]. 保定: 华北电力大学, 2013)

[4] SCHRIJVER C J, KAURISTIE K, AYLWARD A D, *et al*. Understanding space weather to shield society: a global road map for 2015-2025 commissioned by COSPAR and ILWS[J]. *Advances in Space Research*, 2015, **55**(12): 2745-2807

[5] ZHENG Y H, GANUSHKINA N Y, JIGGENS P, *et al*. Space radiation and plasma effects on satellites and aviation: quantities and metrics for tracking performance of space weather environment models[J]. *Space Weather*, 2019, **17**(10): 1384-1403

[6] WANG Jinsong, JIAO Weixin. Space Weather Disasters [M]. Beijing: China Meteorological Press, 2009 (in Chinese) (王劲松, 焦维新. 空间天气灾害 [M]. 北京: 气象出版社, 2009)

[7] CALDWELL B, MCCARRON E, JONAS S. An abridged history of federal involvement in space weather forecasting[J]. *Space Weather*, 2017, **15**(10): 1222-1237

[8] LILENSTEN J, BELEHAKI A. Developing the scientific basis for monitoring, modelling and predicting space weather[J]. *Acta Geophysica*, 2009, **57**(1): 1-14

[9] WANG Jinsong. Space weather operations in CMA[J]. *Advances in Meteorological Science and Technology*, 2011, **1**(4): 6-12 (王劲松. 中国气象局的空间天气业务 [J]. 气象科技进展, 2011, **1**(4): 6-12)

[10] LIU S Q, GONG J C. Operational space weather services in national space science center of Chinese academy of sciences[J]. *Space Weather*, 2015, **13**(10): 599-605

[11] MOU Weifeng. Review of forecast verification methods and proposals for schemes[J]. *Meteorological Monthly*, 1986, **12**(2): 45-49 (牟惟丰. 预报评分方法述评和方案建议 [J]. 气象, 1986, **12**(2): 45-49)

[12] CROWN M D. Validation of the NOAA Space Weather Prediction Center's solar flare forecasting look-up table and forecaster-issued probabilities[J]. *Space Weather*, 2012, **10**(6): 2011SW000760

[13] CAMPOREALE E, BERGER T E. Verification of the NOAA Space Weather Prediction Center solar flare forecast (1998-2024)[J]. *Space Weather*, 2025, **23**(10): e2025SW004546

[14] BALCH C C. Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model[J]. *Space Weather*, 2008, **6**(1): 2007SW000337.

[15] BAIN H M, STEENBURGH R A, ONSAGER T G, *et al*. A summary of National Oceanic and Atmospheric Administration Space Weather Prediction Center proton event forecast performance and skill[J]. *Space Weather*, 2021, **19**(7): e2020SW002670.

[16] SHARPE M A, MURRAY S A. Verification of space weather forecasts issued by the Met Office Space Weather Operations Centre[J]. *Space Weather*, 2017, **15**(10): 1383-1395

[17] MURRAY S A, BINGHAM S, SHARPE M, *et al*. Flare forecasting at the Met Office Space Weather Operations Centre[J]. *Space Weather*, 2017, **15**(4): 577-588

[18] DEVOS A, VERBEECK C, ROBBRECHT E. Verification of space weather forecasting at the Regional Warning Center in Belgium[J]. *Journal of Space Weather and Space Climate*, 2014, **4**: A29

[19] CUI Y M, LIU S Q, ERCHA A, *et al*. Verification of SPE probability forecasts at the Space Environment Prediction Center (SEPC)[J]. *Science China Earth Sciences*, 2016, **59**(6): 1292-1298

[20] LIU Dandan, ZHAO Haijuan, TANG Wei. Verification of short-term geomagnetic storm forecasts[J]. *Chinese Journal of Geophysics*, 2026, **69**(7): 2843-2857 (刘丹丹, 赵海娟, 唐伟. 地磁暴短期概率预报水平的检验评估 [J]. 地球物理学报, 2026, **69**(7): 2843-2857)

[21] LIN Yuanzhang. Introduction to Solar Physics[M]. Beijing: Science Press, 2000 (in Chinese) (林元章. 太阳物理导论 [M]. 北京: 科学出版社, 2000)

[22] 全国卫星气象与空间天气标准化委员会空间天气监测预警分技术委员会. 太阳软X射线耀斑强度分级: GB/T 31157—2014[S]. 北京: 中国标准出版社, 2014: 9

[23] TU Chuanyi, Zong Qiugang, He Jiansen, *et al*. Space Physics (2nd Edition, Vol. 1): Heliospheric Physics[M]. Beijing: Science Press, 2020 (涂传诒, 宗秋刚, 何建森等. 日地空间物理学 (第二版)(上册): 日球层物理 [M]. 北京: 科学出版社, 2020)

[24] ONSAGER T, GRUBB R, KUNCHES J, *et al*. Operational uses of the GOES energetic particle detectors[C]// Proceedings Volume 2812, GOES-8 and Beyond. Denver: International Society for Optics and Photonics, 1996, 2812: 281-290

[25] WILKS D S. Statistical Methods in the Atmospheric Sciences[M]. 2nd ed. Amsterdam: Elsevier, 2006

[26] TU Chuanyi, ZONG Qiugang, ZHOU Xuzhi. Space Physics (Interplanetary and Magnetosphere) (Vol. 2)[M]. Beijing: Science Press, 2020 (涂传诒, 宗秋刚, 周煦之. 日地空间物理学 (行星际与磁层)(下册)[M]. 北京: 科学出版社, 2020)

[27] 全国卫星气象与空间天气标准化委员会空间天气监测预警分技术委员会. 空间天气短期预报检验方法: QX/T 295—2015[S]. 北京: 中国气象出版社, 2015

[28] JOLLIFFE I T, STEPHENSON D B. Forecast Verification: A Practitioner's Guide In Atmospheric Science[M]. 2nd ed. Chichester: Wiley Blackwell, 2012: 274

[29] FAWCETT T. An introduction to ROC analysis[J]. *Pattern Recognition Letters*, 2006, **27**(8): 861-874

[30] GONZALEZ W D, JOSELYN J A, KAMIDE Y, *et al*. What is a geomagnetic storm?[J]. *Journal of Geophysical Research: Space Physics*, 1994, **99**(A4): 5771-5792

[31] RICHARDSON I G, CANE H V. Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996-2009): Catalog and summary of properties[J]. *Solar Physics*, 2010, **264**(1): 189-237

[32] KILPUA E K J, HIETALA H, TURNER D L, *et al*. Unraveling the drivers of the storm time radiation belt response[J]. *Geophysical Research Letters*, 2015, **42**(9): 3076-3084

[33] HATHAWAY D H. The solar cycle[J]. *Living Reviews in Solar Physics*, 2015, **12**(1): 4

[34] BLOOMFIELD D S, HIGGINS P A, MCATEER R T J, *et al*. Toward reliable benchmarking of solar flare forecasting methods[J]. *The Astrophysical Journal Letters*, 2012, **747**(2): L41

[35] DOSWELL III C A, DAVIES-JONES R, KELLER D L. On summary measures of skill in rare event forecasting based on contingency tables[J]. *Weather and Forecasting*, 1990, **5**(4): 576-585

[36] FERRO C A T, FRICKER T E. A bias-corrected decomposition of the Brier score[J]. *Quarterly Journal of the Royal Meteorological Society*, 2012, **138**(668): 1954-1960

作者简介



刘丹丹 女, 博士, 1979 年出生于湖北省随州市, 现为国家卫星气象中心 (国家空间天气监测预警中心) 高级工程师, 主要研究方向为空间天气预报和预报模式研究. E-mail: liudd@cma.gov.cn