doi:10.6041/j.issn.1000-1298.2026.22.001

人工智能气象预报发展范式评价与融合路径

刘文丰[1,2]　孙鸿盛[1,2]　李梦学[1,2]　孙　玮[1,2]　孟　莹[3]　赵　闯[4]

(1. 中国农业大学水利与智能工程学院, 北京 100083; 2. 农业水资源高效利用全国重点实验室, 北京 100083;
3. 大湾区大学先进工程学院, 东莞 523808; 4. 中国农业大学资源与环境学院, 北京 100193)

摘要: 随着再分析资料、深度学习方法和高性能计算技术的发展,人工智能(Artificial intelligence,AI)气象预报已由预报结果订正拓展至全球确定性与集合预报,以及观测资料进入预报前端的学习型预报系统。然而,现有研究多按照网络结构、预报尺度或特定任务对 AI 气象预报归纳,在方法分类与性能评价上仍缺少统一分析框架,难以全面反映不同技术路线的适用条件与应用范围。AI 预报模型已具备全球多变量、多时效预报和快速推理能力,但在数据与状态表征、预报生成统计平滑与物理约束、性能评价与业务验证衔接以及运行适配与版本管理等方面仍存在不足。针对上述问题,本文梳理 AI 气象预报的技术演进,提出一种以实际预报输入为主要依据、以 AI 承担的核心计算环节为辅助判据分类框架,将现有方法归纳为预报产品驱动的后处理与融合方法、分析场初始化状态预报模型和观测前端驱动的学习型预报系统 3 类发展范式;从总体预报技巧、极端事件、空间过程、物理一致性和概率可靠性等方面构建多维性能评价框架,并结合典型研究分析评价证据成立条件及其应用情况。在此基础上,进一步诊断数据、模型、评价与业务运行等环节的主要瓶颈,提出多源、跨尺度与多圈层资料协同,训练目标改进与分层物理耦合,条件一致多维检验,以及持续监测、模型适配与版本管理相衔接的融合路径,为 AI 气象预报的能力评价、方法改进和业务应用提供参考。

关键词: 人工智能气象预报; 数值天气预报; 发展范式; 多维性能评价; 融合路径

中图分类号: P456　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)22-0001-14　　**OSID**:

Development Paradigms Evaluation and Integration Pathways of AI Weather Forecasting

Liu Wenfeng[1,2]　Sun Hongsheng[1,2]　Li Mengxue[1,2]　Sun Wei[1,2]　Meng Ying[3]　Zhao Chuang[4]

(1. *College of Water Resources and Intelligence Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
2. *State Key Laboratory of Efficient Utilization of Agricultural Water Resources*, *Beijing* 100083, *China*
3. *Institute of Advanced Engineering*, *Greater Bay Area University*, *Dongguan* 523808, *China*
4. *College of Resources and Environmental Sciences*, *China Agricultural University*, *Beijing* 100193, *China*)

Abstract: Driven by advancements in reanalysis data, deep learning, and high-performance computing, AI-based weather forecasting has evolved from simple post-processing corrections to global deterministic and ensemble forecasting, and further to learning-based systems that ingest observational data directly. However, existing studies typically categorized these methods based on network architectures, forecast scales, or specific tasks. The absence of a unified framework for classification and performance evaluation hindered a comprehensive understanding of the applicability of different technical approaches. Although current AI models demonstrated capabilities in global multivariate, multi-lead-time forecasting and rapid inference, they faced challenges regarding data and state representation, statistical smoothing, physical constraints, and the alignment between performance evaluation and operational verification. To address these issues, the technological evolution of AI weather forecasting was reviewed and a classification framework was proposed based primarily on forecasting inputs and secondarily on the core computational roles of AI. The existing methods were categorized into three paradigms: forecast product-driven post-processing and fusion methods, analysis-initialized state forecasting models, and observation-driven learning-based forecasting systems. Furthermore, a multidimensional performance evaluation framework

收稿日期: 2026-08-07　修回日期: 2026-08-28

基金项目: 国家重点研发计划青年科学家项目(2024YFC3213700)和国家自然科学基金项目(52679062、52239002、32361143871)

作者简介: 刘文丰(1988—),男,教授,主要从事气候变化与农业水文过程研究,E-mail: wenfeng.liu@cau.edu.cn

covering overall skill, extreme events, spatial processes, physical consistency, and probabilistic reliability was constructed. By analyzing typical studies, the capability boundaries and application contexts of these methods were defined. Finally, key bottlenecks in data, modeling, evaluation, and operations were identified, proposing a convergence pathway that integrated multi-source data synergy, hierarchical physical coupling, consistent operational verification, and robust version management, thereby providing a reference for the advancement and operational deployment of AI weather forecasting.

Key words: AI weather forecasting; numerical weather prediction; developmental paradigms; multidimensional performance evaluation; convergence pathway

0 引言

气象预报是认识大气演变规律、开展气象灾害预警和保障经济社会安全运行的重要基础,预报能力直接影响农业生产管理和水资源调度等高度依赖气象信息的领域[1-3]。在全球气候变化背景下,极端天气事件频发,致灾风险上升,社会各领域对气象预报的时效性、精细化和稳定性提出了更高要求[4-9]。因此,发展更高精度、更高效率和更强业务适用性的气象预报方法,已成为当前研究的重要方向[10-12]。长期以来,数值天气预报(Numerical weather prediction,NWP)是现代气象预报的核心技术路径[13-15],在资料同化、物理参数化和集合预报等关键技术支撑下,NWP 推动了全球与区域中短期预报能力的发展,并为灾害性天气预警提供了重要支撑[16-18]。然而,随着社会经济发展,更精细化预报和高频更新需求的增强,传统 NWP 体系在计算成本、参数化误差累积以及业务流程复杂性等方面面临新的挑战,成为提升预报效率、准确性与稳定性的主要限制[19-21]。

近年来,在大规模气象资料积累、人工智能算法发展和高性能计算能力提升的推动下,气象预报研究正加速由基于物理过程模拟向数据驱动范式拓展[22-25]。不同于传统数值天气预报主要依赖显式物理方程求解,人工智能(Artificial intelligence,AI)气象预报模型主要通过学习历史观测、再分析资料或模式输出时空关系,建立大气状态变化与未来预报结果之间的映射关系,并据此生成未来预报结果[26-29]。以深度学习为代表的 AI 气象预报方法在全球中期预报、短时临近预报及预报后处理等方向发展迅速[30-32],AI 模型已具备全球多变量联合预报能力,并在推理速度和单次预报计算成本方面表现出明显优势[33-36],表明人工智能正在成为推动气象预报方法体系由单一物理驱动向数据驱动与物理驱动协同演进的重要力量。

然而,AI 气象预报的快速发展并不意味着对数值模式简单替代,而是气象预报体系在新数据、新算法和新设备下进一步拓展[37-39]。AI 模型能够有效提取复杂非线性关系并实现快速预报生成,但性能较大程度依赖训练资料和初始状态质量,在极端天气刻画、物理一致性、跨区域泛化和不确定性表达等方面仍存在不足[12,33,39-41],尤其在复杂天气过程和长期业务运行条件下,AI 模型是否具有稳定可靠的预报能力,仍需要通过系统评价和业务实践进一步验证。因此,当前研究需要由证明模型能否生成有效预报,进一步转向对不同方法任务边界、性能证据成立条件和业务适用范围系统分析。

现有 AI 气象预报综述主要从模型结构、时空尺度、资料利用方式和具体预报任务等角度归纳相关进展,但研究分析对象和分类依据存在差异,Ren 等[42]以深度学习气象预报为研究对象,神经网络架构、预报时空尺度、数据集与评价基准以及 AI 与 NWP 的关系等方面已有研究,重点研究不同算法结构及数据驱动模型与数值模式的结合。Zeng 等[43]聚焦降雨预报,围绕神经网络类型、训练过程、损失函数、数据集和评价指标总结技术进展,呈现了 AI 在特定预报要素中应用。资料同化相关综述主要讨论观测资料处理和初始状态生成[39],面向水文预报等下游应用研究则更重视气象预报订正、应用模型设置和预报检验等环节[44]。从整体上看,现有研究从算法结构、预报要素、技术环节和应用任务等不同层面深化了对 AI 气象预报的认识,但由于分析对象和分类依据不同,预报产品后处理、当前状态生成和未来状态预报等方法尚未在同一框架下得到统一比较,模型在实际预报中直接使用何种资料、AI 承担何种核心计算任务以及不同方法之间的功能边界也缺少系统梳理。此外,Zeng 等[43]对不同数据集和评价指标的总结还表明,模型性能比较需要结合具体的数据基础、评价方式和预报任务进行分析,难以脱离试验条件作统一判断。

基于上述研究基础,本文首先梳理 AI 气象预报的技术演进,并依据模型在实际预报中直接使用的输入数据和 AI 承担的核心计算环节,将现有方法归纳为预报产品驱动的后处理与融合方法、分析场初始化的状态预报模型和观测前端驱动的学习型预报系统;随后从总体预报技巧、极端事

件、空间过程、物理一致性和概率可靠性等方面构建多维性能评价框架，并结合典型案例研究分析评价证据的成立条件及其应用情况；在此基础上，从数据、模型、评价和业务运行 4 个层面归纳主要瓶颈，并讨论相应融合路径。通过上述分析，尝试形成由技术演进、方法分类、能力边界、性能证据与应用情况，再到融合路径相互衔接的分析框架，以期为 AI 气象预报的能力评价、方法改进和业务应用提供参考。

1　AI 气象预报技术演进与方法体系

本文按照 4 个层面递进逻辑逐步展开分析，总体分析框架如图 1 所示。基于结构安排，首先梳理 AI 气象预报的技术演进，并在此基础上建立统一的方法体系分类框架。AI 在气象预报中应用已由局地要素订正和短时预报，逐步发展到全球多变量天气状态预报、概率集合预报以及观测资料进入预报前端的学习型预报系统。随着模型使用资料和承担任务的增加，仅依据网络结构、预报尺度或应用场景进行归纳，容易造成不同类型之间的交叉。本文以模型在实际预报中直接使用的输入数据为主要分类依据，并结合 AI 直接完成核心计算的环节，将现有方法分为预报产品驱动的后处理与融合方法、分析场初始化的状态预报模型和观测前端驱动的学习型预报系统，并分析不同方法的主要特点、适用条件和能力边界。

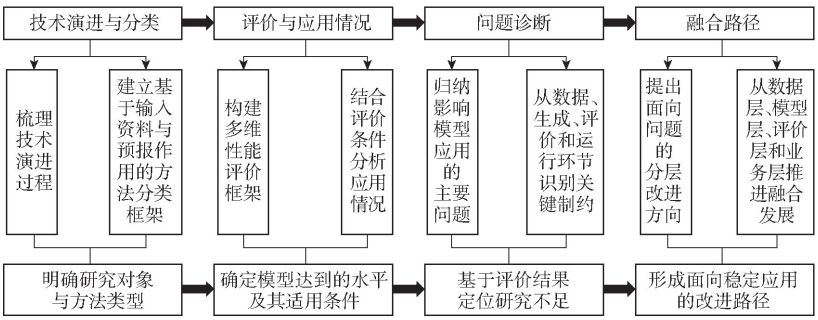


图 1　AI 气象预报总体分析框架

Fig. 1　Overall analytical framework for AI-based weather forecasting

1.1　AI 气象预报技术演进过程

气象资料的长期积累、深度学习方法的发展以及计算能力的提高，为 AI 气象预报研究提供了重要基础。综合考虑代表性模型、资料条件和预报任务的演进，相关研究大致经历了局部任务探索、全球状态预报发展和预报体系扩展 3 个阶段。各阶段之间并不存在严格的时间分界，相关时间节点、代表性模型、应用定位和能力边界如图 2 所示。图中年份按代表性模型对应论文首次发表年标注，不同节点表示代表性模型、预报系统或标志性研究。图中发展阶段由本文根据代表性研究的输入资料、AI 承担的核心计算任务、预报范围和应用方式综合归纳，各阶段之间不存在严格的时间分界；能力边界是在相应类别代表性模型研究基础上的概括。AIFS 为人工智能预报系统（Artificial intelligence forecasting system），ENS 为集合预报（Ensemble forecasting），

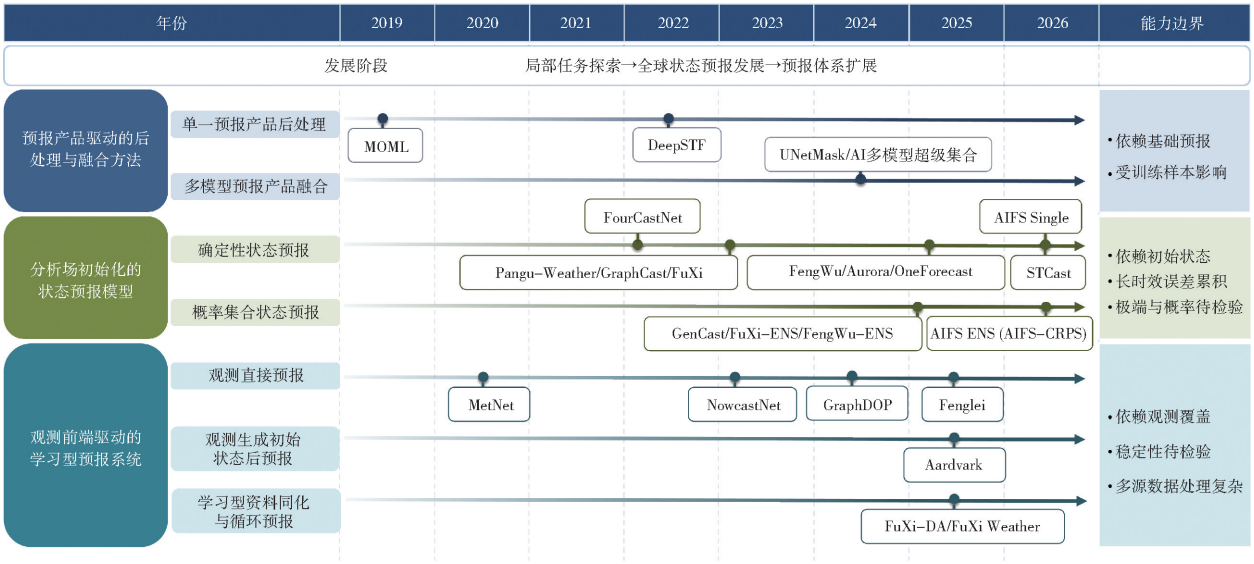


图 2　AI 气象预报技术演进、方法体系及能力边界

Fig. 2　Technological evolution, methodological framework and capability boundaries of AI-based weather forecasting

DA 为资料同化(Data assimilation);Pangu - Weather、FuXi、FengWu 和 Fenglei 分别为盘古气象、伏羲、风乌和风雷模型。

早期 AI 气象预报主要围绕特定地区或气象要素展开,数值天气预报能够提供较完整的预报产品,利用机器学习进一步订正模式误差、开展统计后处理,对原有预报流程改动较小,因而成为 AI 较早应用方向。MOML、DeepSTF 和 UNetMask 等方法分别用于格点温度订正、多站点后处理和短时降雨订正[45-47];雷达和卫星等高频观测资料能够连续记录天气变化,为深度学习开展短时降雨预报提供了较好的资料条件,MetNet 模型验证了深度学习在高时空分辨率短时降雨预报中应用潜力[48]。受早期资料范围、模型规模和计算条件限制,这一阶段的研究多面向具体任务,并逐渐形成预报产品后处理和观测驱动短时预报 2 个方向。

随着 ERA5 等全球再分析资料不断完善,AI 模型提供了时间连续、空间规则且包含多层次气象变量的训练数据,计算设备和深度学习方法的进步也提高了模型处理全球高分辨率气象场的能力[13,21,23]。在这些条件推动下,FourCastNet 于 2022 年实现 0.25°分辨率的全球数据驱动天气预报,随后 Pangu - Weather、GraphCast、FuXi 和 FengWu 等模型相继提出,AI 预报对象由局部地区或少数变量扩展到全球多变量大气状态[34-36,49-50]。这些模型以格点分析场表示当前大气状态为初始条件,直接生成未来多个时效预报,使 AI 由已有预报结果的辅助处理发展到承担未来天气状态的核心计算。AIFS Single 于 2025 年 2 月 25 日起投入业务运行,AIFS ENS 于 2025 年 7 月 1 日起业务运行,分别实现了确定性和集合 AI 预报的实时发布,表明分析场初始化的 AI 模型已经开始由研究试验进入实际预报系统[51-52]。

在全球确定性状态预报取得进展后,单一预报结果难以充分表达未来天气的不确定性,推动相关研究向概率集合预报发展。GenCast、FuXi - ENS、FengWu - ENS 和 AIFS ENS 等模型通过生成多个可能的未来天气状态,对预报不确定性进行表达[50,52-54]。与此同时,模型训练方式也开始由针对固定变量和任务的专用训练,向大规模预训练和面向具体任务的微调拓展,Aurora 利用多种地球系统资料进行预训练,并针对天气、海浪、空气质量和热带气旋等任务进行微调,体现了基础模型在不同地球系统预报任务中适配能力[55],天气预报模型仍以分析场为初始条件,属于分析场初始化确定性状态预报模型,预训练和微调主要反映训练方式变化。

观测资料进入预报前端方式也逐渐增多,NowcastNet 和 Fenglei 等模型继续利用雷达等高频观测资料开展临近预报[56-57];Aardvark Weather 尝试由多源观测资料生成当前格点大气状态并继续开展未来预报[58];FuXi - DA 和 FuXi Weather 将卫星观测与背景预报结合,用于生成分析场和开展循环预报[59-60];GraphDOP 则直接利用地球系统观测资料建立模型内部状态并预报未来气象变量,不以传统分析场作为实际预报输入[61]。这些研究使观测驱动方法由局地短时预报进一步扩展到当前状态估计、学习型资料同化和中期预报。

全球预报与区域天气细节之间的衔接也受到关注,OneForecast 利用全球背景信息支持区域高分辨率预报,STCast 进一步针对全球与区域之间的空间关系和边界信息进行调整[62-63]。这类模型仍以格点大气状态为起点,主要变化在于将大尺度天气背景与区域天气信息结合,以改善区域预报中边界信息和局地细节描述。

总体来看,AI 气象预报演进受到资料条件、模型方法和预报需求的共同推动。其应用由预报产品后处理和观测驱动短时预报,逐步发展到分析场初始化的全球多变量状态预报,并进一步拓展到概率集合预报、预训练与任务微调、观测前端学习型预报系统以及全球与区域预报衔接。

1.2 AI 气象预报主要方法体系

为避免训练资料、输出形式和空间范围等不同维度相互交叉,以模型在实际预报中直接使用的输入数据为主要分类依据,并结合 AI 对输入数据所完成的主要处理,将现有方法分为预报产品驱动的后处理与融合方法、分析场初始化的状态预报模型和观测前端驱动的学习型预报系统。第 1 类以已经生成的 1 个或多个预报产品为输入,AI 主要负责订正、后处理或融合。第 2 类以格点分析场为初始状态,由 AI 直接预报未来 1 个或多个时效的大气状态。第 3 类则将雷达、卫星、地面站点和探空等观测资料直接输入学习系统,由 AI 开展未来量预报、当前状态估计或资料同化。

1.2.1 预报产品驱动后处理与融合方法

预报产品驱动的后处理与融合方法以已经生成的 NWP 或 AI 预报结果及其相关变量为输入,利用机器学习进一步开展误差订正、统计后处理或产品融合。AI 不直接从当前大气状态预报未来天气,而是在预报产品形成后,根据历史预报结果与观测之间的关系进行修正或组合。按照输入产品的数量和处理方式,可分为单一预报产品后处理和多模型预报产品融合。

单一预报产品后处理主要针对某一模式、地区或气象要素建立订正关系。Li 等[45]提出模式输出机器学习方法 MOML，将 NWP 输出、历史预报信息和空间位置等作为输入，用于北京地区格点温度订正。Kong 等[46]提出深度时空预报模型 DeepSTF，在考虑各站点时间变化和空间联系的基础上，对多个站点的温度、湿度和风速进行统一后处理。这类方法通常围绕特定预报对象开发，便于接入现有预报流程，但订正效果会受输入预报质量和训练样本条件的影响。

多模型预报产品融合同时使用 2 个或多个预报结果，通过规则或学习方法形成新的综合预报。Zheng 等[47]提出 UNetMask 方法将 GRAPES 模式结果与深度学习降雨预报结果相结合，通过掩膜方法改善短时定量降雨预报。Purohit 等[64]提出基于 AI 的多模型超级集合方法，以 Pangu－Weather、SFNO 等模型的预报结果为输入，利用 XGBoost 学习不同模型结果之间的组合关系，形成 AI 多模型超级集合。由此可见，这类方法共同特征是在已有预报产品的基础上进行后处理或融合，效果取决于输入产品之间是否具有可利用的误差差异。

1.2.2　分析场初始化状态预报模型

分析场是观测资料和背景预报经过资料同化后形成的格点大气状态。再分析资料主要为模型训练和历史试验提供较为连续的数据，业务分析场则用于实际预报起始时刻的初始化。分析场初始化的状态预报模型不以未来时刻 NWP 结果作为主要输入，而是根据当前分析场直接生成未来 1 个或多个时效的气象变量。根据输出形式，可分为确定性状态预报和概率集合状态预报。

确定性状态预报模型在 1 次起报中给出 1 组未来天气结果，FourCastNet、Pangu－Weather、GraphCast、FuXi 和 FengWu 等模型均以全球格点大气状态为输入，直接开展多变量中期预报[34－36,49－50]；AIFS Single 进一步将此类方法用于业务确定性预报，并通过训练资料扩充、变量调整和输出约束等方式持续更新模型版本[51]；Aurora 虽然采用多源资料预训练和面向具体任务微调的训练方式，但天气预报版本仍由分析场初始化并生成确定性预报[55]；OneForecast 和 STCast 同样以格点大气状态为起点，通过利用全球背景信息支持区域预报，归入分析场初始化的确定性状态预报模型[62－63]。

确定性模型通常只给出一种未来天气结果，难以完整反映大气演变的不确定性。GenCast、FuXi－ENS 和 FengWu－ENS 等模型通过生成多个可能的未来状态开展概率集合预报[50,53－54]；ECMWF 基于 AIFS－CRPS 模型构建并业务运行 AIFS ENS，其中 AIFS－CRPS 采用基于连续分级概率评分的概率训练，可生成多个集合成员并表达预报不确定性[52]。确定性和概率集合模型虽然输出形式不同，但均以分析场作为预报起点，并由 AI 承担未来大气状态预报的主要计算。

1.2.3　观测前端驱动学习型预报系统

观测前端驱动的学习型预报系统将雷达、卫星、地面站点和探空等观测资料直接输入 AI 模型，而不是只接收外部资料同化系统已经生成的分析场。根据观测资料进入系统后处理方式，以及形成当前格点状态时是否需要背景预报，可分为观测直接预报、观测生成初始状态后预报，以及学习型资料同化与循环预报 3 种形式。

观测直接预报直接根据连续观测资料预报未来结果，MetNet 利用雷达和卫星资料输出未来数小时概率降雨分布[48]；NowcastNet 结合降雨变化规律和生成式方法，开展极端降雨临近预报[56]；Fenglei 则根据雷达组合反射率预报未来 3 h 雷达回波[57]；GraphDOP 则将这一思路拓展到中期预报，直接利用地球系统观测建立模型内部状态，并在观测空间和模型内部表示中预报未来气象变量，不以传统分析场作为实际预报输入[61]。这些模型预报时效和输出变量不同，但共同特点是由观测资料直接形成未来预报结果。

观测生成初始状态后预报是由 AI 先根据多源观测资料估计当前格点大气状态，再以该状态为起点预报未来天气。Allen 等[58]提出 Aardvark Weather 系统，将卫星、地面站、船舶和探空等资料输入模型，先估计当前全球大气状态，再生成格点和站点天气预报。在实际预报中不以 NWP 背景预报作为必要输入，因而不同于需要观测资料和前一时次短期预报共同形成分析场的资料同化系统。

学习型资料同化与循环预报同时使用观测资料和背景预报，由 AI 生成当前分析场，并以该分析场启动未来预报，FuXi－DA 利用卫星观测和背景场生成格点分析结果，FuXi Weather 进一步将资料同化与天气预报连接，使前次短期预报成为下次状态更新的背景，从而形成循环运行的分析－预报系统[59－60]。这种形式保留了观测、背景预报和分析场之间的联系，与 Aardvark Weather 仅由观测估计初始状态以及 GraphDOP 直接在观测和模型内部表示中预报未来结果存在明显区别。

上述 3 类方法区别主要在于实际预报输入和 AI 所完成的处理环节。预报产品驱动方法处理已经生成的预报结果，分析场初始化模型由当前格点

状态直接预报未来天气,观测前端驱动系统则从观测资料开始完成直接预报、状态估计或资料同化。

1.3 不同方法能力边界

3 类 AI 气象预报方法在实际预报输入、AI 承担的计算环节和输出形式等方面存在差异,性能评价需要结合预报区域、气象要素、预报时效和应用任务开展,不能仅依据单一总体误差判断模型表现。

预报产品驱动的后处理与融合方法便于接入现有预报流程,适合针对特定地区、站点和气象要素开展误差订正、统计后处理或多模型融合[45-47,64],这类方法改进效果取决于输入预报产品的质量、不同产品之间的误差差异以及训练样本的代表性。对于单一产品后处理,应在相同验证资料下比较处理前后的结果;对于多模型产品融合,还应分别与各输入模型及简单平均等基准方法比较,在不同变量、预报时效和天气条件下检验融合增益是否保持一致。

分析场初始化的状态预报模型已具备全球多变量和多时效预报能力,并在推理速度和单次预报计算成本方面表现出明显优势[34-36,49-50],但结果仍受初始分析场、训练资料和连续预报误差累积的影响,确定性模型还可能在极端强度和空间结构方面存在不足,概率集合模型则需要检验集合离散度和预报概率是否可靠[33,58-59],AIFS 业务运行和版本更新进一步说明,训练资料、变量设置和模型约束的变化均可能影响预报性能,模型更新后仍需在相同条件下重新评价[51-52]。因此,这类模型不仅需要比较总体预报技巧,还应结合极端事件、空间过程、物理一致性、概率可靠性和不同版本长期表现进行评价。

观测前端驱动的学习型预报系统能够直接利用雷达、卫星、地面站点和探空等资料开展临近预报、当前状态估计或学习型资料同化[48,56,58-59],其表现与观测资料的覆盖范围、时间更新频率、质量和变量匹配程度密切相关。对于直接由观测生成未来结果的模型,应检验在不同观测覆盖和资料条件下预报表现[48,56,61];对于先由观测生成当前格点状态再开展预报的系统,应分别评价当前状态估计和后续预报[58];对于同时使用观测资料和背景预报的循环资料同化系统,还需比较背景预报对分析结果的影响,并检验多次循环后的稳定性[59-60]。

总体而言,3 类方法能力边界主要受输入资料、处理环节和输出形式共同影响。评价时应首先明确模型实际使用的资料和完成的任务,再从总体预报技巧、极端事件、空间过程、物理一致性和概率可靠性等方面选择相应指标,并在相同或可说明差异的条件下开展比较。

2 AI 气象预报性能评价

随着 AI 气象预报方法由预报产品订正与融合逐渐发展到基于分析场初始化的状态预报以及观测前端驱动的学习型预报系统,不同方法在输入资料、初始化方式、输出形式和预报任务等方面形成了明显差异。同时,模型评价目标也由单纯比较预报误差,逐渐转向分析不同方法能够解决的问题、优势形成条件以及可能存在的限制。因此,AI 气象预报性能评价不仅需要关注总体预报精度,还需要结合极端天气、空间结构、物理一致性和概率信息等方面,判断模型能力是否具有稳定性和可推广性。

2.1 不同方法评价需求与比较基础

不同类型 AI 气象预报方法承担的任务不同,评价重点也存在差异。预报产品驱动方法主要利用已有预报结果开展订正、后处理或融合,其性能评价需要关注 AI 处理后结果相对于原始预报产品的改进程度。分析场初始化的状态预报模型直接根据当前大气状态生成未来天气状态,评价时需要明确初始状态来源、验证资料以及比较基准,以判断模型自身预报能力,而非将初始状态质量差异纳入模型性能比较。观测前端驱动的学习型预报系统则进一步涉及观测资料利用、当前状态估计和未来预报生成,其评价需要区分不同环节对最终预报结果的贡献。WeatherBench 基于 ERA5 再分析资料设置了统一的预报任务、基准模型和评价指标,为数据驱动天气预报模型比较提供了较为规范的测试基础[65];Pangu-Weather、GraphCast 和 FuXi 等通过多变量、多时效评价与传统数值模式开展比较,使全球 AI 天气预报的性能分析更加系统[34-36]。不同研究之间性能结果仍受试验条件影响,模型训练资料、初始状态来源、验证数据、空间分辨率、变量范围以及比较基准的差异,均可能导致评价结果不同。因此,已有研究关于 AI 模型性能提升或优于传统方法的结论,通常对应特定资料条件和验证尺度,不能简单推广至不同区域和业务环境。

2.2 面向能力识别的多维性能评价方法

AI 气象预报模型的能力不仅表现为平均误差降低,还体现在对天气、空间结构、大气过程以及未来不确定性的表达能力。针对不同能力特征,可从总体预报技巧、极端事件与空间过程、物理一致性和概率可靠性等方面开展综合评价。不同评价维度对应的主要能力及常用指标见表 1。

总体预报技巧是当前 AI 气象预报研究中最常采用的评价方式。均方根误差(Root mean square error,RMSE)、平均绝对误差(Mean absolute error,

表 1　不同评价维度对应的 AI 气象预报能力及常用指标

Tab. 1　AI weather forecasting capabilities and common metrics across evaluation dimensions

评价维度	主要指标	对应能力	主要作用及局限
总体预报技巧	RMSE、MAE、Bias、ACC	平均预报水平和变量变化一致性	反映整体误差水平，便于模型间比较，但难以体现局地结构和极端事件差异
极端事件	POD、FAR、CSI、极端分位数偏差、事件频率和强度偏差	高影响天气刻画能力	评价异常天气过程，但受事件阈值和样本数量影响
空间过程	邻域检验、对象检验、空间尺度分析、天气系统追踪	天气系统位置、范围和结构表达能力	补充逐格点误差不足，但依赖评价尺度和参数设置
物理一致性	变量协调关系、典型过程诊断、空间结构	变量关系和天气演变合理性	识别平均误差无法反映的问题，但缺少统一量化标准
概率可靠性	CRPS、Brier 评分、可靠性、集合离散度	未来状态不确定性表达能力	评价概率预报质量，但受集合构建方式影响

MAE）用于衡量预报结果与观测或分析资料之间的误差水平，平均偏差（Bias）用于反映系统性高估或低估，距平相关系数（Anomaly correlation coefficient，ACC）则用于评价预报变化趋势与实际变化的一致程度[34,36,65]。该类指标能够反映模型整体预报能力，但结果通常综合不同区域、天气类型和强度等级的误差，因此不能单独用于判断模型对局地天气结构和极端事件的预报能力。

针对高影响天气过程，极端事件评价进一步关注模型对暴雨、高温、强风和热带气旋等异常状态的刻画能力。通过命中率（Probability of detection，POD）、空报率（False alarm ratio，FAR）、临界成功指数（Critical success index，CSI）以及极端分位数偏差、事件频率和强度偏差等指标，可以分析模型对事件发生概率、强度和持续过程的预报能力[33,40]。对于具有明显空间结构的天气系统，逐格点误差可能受到位置偏移影响，因此还需要结合对象检验、邻域检验和空间尺度分析方法评价天气系统的位置、范围和演变特征[14,56,66]。

除误差水平外，AI 预报结果是否能够保持合理的大气结构和变量关系，也是判断模型可靠性的重要方面。已有研究表明，部分 AI 模型虽然具有较好的总体误差表现，但仍可能存在天气系统结构偏差和局地强度刻画不足等问题[33,66]。对于概率集合模型，还需要利用连续分级概率评分（Continuous ranked probability score，CRPS）、Brier 评分以及可靠性分析等方法，判断模型输出的不确定性是否与实际预报误差相匹配[50,53-54]。

因此，不同评价方法反映模型能力的不同方面，单一平均误差指标不能全面代表 AI 气象预报水平，需要结合模型任务和输出形式开展综合评价。

2.3　典型评价证据与应用情况

现有模型采用的输入资料、初始化方式、空间分辨率、预报任务和比较基线差异较大，单纯汇总模型是否优于某一对照，难以说明相关优势能够适用于何种预报条件。因此，表 2 在选择评价条件和限制说明相对完整的研究基础上，进一步兼顾方法类型、预报尺度和国内外案例的代表性。

已有研究的性能优势均建立在特定评价条件下，预报产品融合方法需要以输入模型和简单集合结果为比较基线，其性能增益受到输入产品质量、目标变量、区域和样本时段的影响。Purohit 等[64]的结果仅能说明该融合方法在德国地区 1 月近地面温度和风场预报中改进效果。分析场初始化模型的评价则与训练资料、初始状态来源和验证基准密切相关，Pangu－Weather 在 ERA5 初始化和验证条件下表现出较强的全球确定性预报技巧，但与不同对照系统的初始资料和验证设置并非完全一致[34]；AIFS Single 1.1.0 的性能变化表明，业务分析资料更新和训练时段扩展也会影响模型技巧[51]；OneForecast 利用全球背景支持区域预报，但相关结论仍受到全球资料分辨率、背景输入方式和试验区域的约束[62]。观测前端驱动系统还同时受到观测覆盖、状态估计和背景预报质量的影响，Fenglei 评价建立在我国雷达资料和 SWAN 光流基线之上，主要反映雷达回波临近预报能力[57]；Aardvark Weather 与 HRES、GFS 比较建立在共同的 1.50°网格上，只能说明在该尺度下具有一定竞争力，不能推断已达到业务原始分辨率和完整变量体系下预报水平[58]；FuXi Weather 分析与预报表现还依赖背景预报质量[59]。因此，优于某一基线仅说明模型在相应资料、任务和评价尺度下具有优势，不能直接外推为对其他模型、区域或业务系统的全面替代。

总体误差、极端事件、空间结构和概率可靠性反映不同层面的预报能力，单项指标的改善并不意味着其他能力同步提高。AIFS Single 1.1.0 总体技巧和轻中等降雨评分有所提高，但较高降雨阈值下仍存在事件频率低估[51]；Pangu－Weather 在 RMSE 和

表 2 典型 AI 气象预报研究的评价条件、性能证据与应用情况

Tab. 2 Evaluation conditions, performance evidence, and applications of typical AI weather forecasting studies

典型研究	关键试验条件	验证与比较方式	主要性能证据	应用情况
AI 多模型超级集合[64]	融合 Pangu－Weather 和 SFNO 产品；德国地区 1 月；2 m 气温和 10 m 风；最长 15 d	ERA5 验证；与输入模型及简单集合平均比较；RMSE	RMSE 低于单一模型和简单平均，长时效改进较明显	仅涉及特定区域、季节和近地面变量；研究性试验
Pangu－Weather[34]	ERA5 训练和初始化；0.25°全球 69 变量；最长 7 d	ERA5 和 IBTrACS 验证；与 IFS、FourCastNet 和 HRES 比较；RMSE、ACC 及路径误差	多数变量总体技巧优于对照，热带气旋路径预报具有竞争力	基于 ERA5 试验；预报场相对平滑；部分对照的初始资料不一致；研究模型
AIFS Single[51]	ERA5 预训练，业务分析场微调和初始化；约 0.25°全球多变量；1～10 d	分析场及独立观测验证；与早期 AIFS 和业务 IFS 比较；RMSE、ACC 及降雨评分	多数变量总体技巧提高，轻中等降雨预报有所改善	受训练资料和业务分析场更新影响，较强降雨仍存在低估；已业务运行
FengWu/FengWu－ENS[50]	ERA5 训练并适配业务分析场；0.25°全球确定性与集合预报；最长 10 d	ERA5 和业务分析资料验证；与 IFS、Pangu－Weather、GraphCast 等比较；RMSE、ACC、CRPS	确定性和概率技巧均具有竞争力，部分热带气旋预报得到改善	长时效存在平滑现象；集合中偶有物理不一致或反事实成员；实时试验阶段
OneForecast[62]	WeatherBench2/ERA5 训练；1.5°全球及区域嵌套；变量 69；最长 10 d	ERA5 验证；与 Pangu－Weather、GraphCast、FuXi、Graph－EFM 及区域基线比较；RMSE、ACC	全球和区域预报具有竞争力，神经嵌套改善区域预报表现	区域结果依赖全球背景；全球试验分辨率为 1.5°；结论受试验区域和个例限制；研究模型
Fenglei[57]	2016—2022 年全国雷达资料训练；1 km、6 min；最长 3 h	2023 年全国检验及 2024 年强对流个例评估；与 SWAN 光流比较；TS、Bias	各时效 TS 优于 SWAN，强回波 TS 提高 33%，能够刻画部分对流生消演变	依赖初始场和对流尺度；长时效趋势、回波强度和结构仍有偏差；已发布并开展业务检验
Aardvark Weather[58]	卫星、站点和探空等观测，不使用 NWP 背景；1.5°全球；最长 10 d	ERA5 和站点观测验证；与重网格化 HRES、GFS 及站点产品比较；RMSE、MAE	部分变量、区域和预报时效达到或优于对照结果	对照模式均重网格至 1.50°，不能外推至业务原始分辨率和完整变量体系；研究原型
FuXi Weather[59]	卫星观测与背景预报共同输入；6 h 循环同化；0.25°全球；最长 10 d	ERA5 和分析场验证；与 HRES 及无背景方案比较；RMSE、ACC	背景预报可改善分析精度和后续预报，部分变量在中长时效具有优势	依赖背景预报，预报场相对平滑，在地面观测密集区优势较弱；连续循环试验

ACC 等总体指标上表现突出，但预报场相对平滑，小尺度空间结构仍需结合过程检验进行判断[34]；Fenglei 的 TS 评分优于光流方法，并能刻画部分对流系统的生消趋势，但个例评价仍发现长时效趋势判断以及回波强度、形态和位置存在偏差[57]；FuXi Weather 在部分变量和较长预报时效上 RMSE 具有优势，但同时表现出预报场活动度下降和空间结构平滑[59]；FengWu－ENS 在 CRPS 和集合离散度等概率指标上表现较好，但长时效下仍可能出现物理不一致或反事实集合成员[50]。因此，RMSE 和 ACC 主要反映总体误差及大尺度变化的一致性，CRPS 反映集合概率分布的综合质量，但这些指标不能单独证明模型在极端强度、空间细节、物理合理性和单个集合成员可信度等方面均得到改善，AI 气象预报性能评价应结合模型输出形式和应用任务，将总体技巧与极端事件、空间结构、预报活动度及概率可靠性等评价结果综合考虑。

综上，AI 气象预报模型性能受到输入资料、初始化方式、预报任务、空间分辨率、验证资料、比较基线和评价指标等条件共同影响。AI 方法已在预报产品融合、全球确定性与概率预报、全球区域衔接、短时临近预报和观测前端系统等方面取得进展，但不同研究结果只有在评价条件基本一致时才具有直接可比性。当试验条件存在明显差异时，应重点说明各项评价结果所证明的具体能力及其尚未验证的范围，而不宜形成脱离任务、尺度和评价条件的模型排序。明确性能证据的成立条件和结论适用范围，是避免模型能力被过度解释，并判断其业务应用价值的重要基础。

3 AI 气象预报发展瓶颈与研究不足

前述方法梳理与性能评价表明，AI 气象预报已经具备预报产品优化、全球多变量状态预报、概率集合生成和实时观测利用等能力，但相关表现仍受到输入资料、初始状态、预报任务和评价条件的限制。从预报流程看，现阶段的主要问题分别存在于资料与当前状态表征、未来状态生成、性能评价与业务验证以及长期运行管理等环节。

3.1　数据与状态表征不足

不同 AI 气象预报方法依赖的资料基础存在明显差异，预报产品驱动方法以 NWP 或 AI 预报结果为输入，分析场初始化模型通常使用再分析资料或业务分析场，观测前端驱动系统则直接利用雷达、卫星、地面站点和探空等资料。再分析资料和业务分析场具有时空连续、变量较完整等特点，但同时包含基础模式和资料同化系统的影响；直接观测则受到空间覆盖不均、资料缺测、仪器变化和质量差异等限制，这些问题均会影响当前大气状态的表征并传递至后续预报[58-59,61]。全球模型能够提供大尺度天气背景，但对复杂地形和局地天气细节的描述仍相对有限，区域高分辨率预报又需要全球信息提供边界条件，OneForecast 通过嵌套网格连接全球与区域预报，STCast 进一步利用动态空间关联提取区域边界信息，表明全球区域协同已成为区域精细化预报的重要路径，但不同尺度资料在网格、变量和时间频率上衔接仍需完善[62-63]。此外，当前多数 AI 天气模型仍以大气状态为主要预报对象，Aurora 和 NeuralGCM 等研究虽然已将任务扩展至海浪、大气成分或气候模拟，但尚未完整描述大气、海洋、陆面和冰冻圈之间的相互作用，面向中长期预报的多圈层状态表征仍处于探索阶段[55,67-68]。

3.2　统计平滑与物理约束不足

当前多数确定性 AI 气象预报模型采用均方误差（Mean squared error，MSE）或平均绝对误差（MAE）等回归损失学习未来大气状态，当给定初始状态对应多种可能的天气演变时，MSE 和 MAE 最优点估计分别对应条件均值和条件中位数，因而会将不同位置、强度和发展路径的可能状态压缩为单一结果，表现为小尺度结构减弱、预报场活动度下降和局地极值偏弱；在自回归预报中，平滑后的结果继续作为下一时次输入，还会使小尺度信息损失随预见期延长而累积[53,69]。统计平滑是极端事件频率和强度被低估的重要原因，极端样本数量有限、空间分辨率不足以及破纪录事件超出训练资料分布等因素同样会限制模型表现。Zhang 等[40]发现，部分 AI 模型对破纪录高温、低温和强风事件的发生次数及强度存在系统性低估，且误差随事件超出历史记录的程度增加，表明训练域内外推能力不足仍是极端天气预报的重要瓶颈。物理一致性不足则是另一个问题，较低的 RMSE 或较高的 ACC 不能保证变量关系、天气系统结构和长时效演变均符合大气规律，Saleem 等[70]和 Yu 等[71]分别通过热力、动能和守恒关系构建物理启发损失，但这类方法通常属于软约束，其效果受到物理关系简化程度和损失权重影响。Hu 等[72]对 Aurora 大气化学模型的分析表明，较高的总体技巧并不必然意味着模型掌握了相应的物理或化学机制。

3.3　性能评价与业务验证不足

现有基准测试主要在固定历史时段和预设资料条件下比较模型的总体误差、极端事件、空间结构和概率可靠性，但不同研究采用的训练资料、初始状态、空间分辨率、验证资料和比较基线并不完全一致，相关结论只能在相应试验条件下成立[33,65]。与此同时，多数研究仍以报告评价指标为主，较少进一步识别误差来源。预报产品驱动方法的性能增益可能来自不同输入产品之间的互补，分析场初始化模型受到初始状态质量影响，观测前端系统的最终误差则同时包含观测处理、状态估计和未来预报等多个环节[58-59]。业务验证还需要关注模型能否连续、及时地生成预报，以及在资料缺测、输入变化和不同天气条件下能否保持稳定。AIFS 业务运行和 FuXi Weather 的 1 周年循环试验已经开始涉及连续运行问题，但相较于大量基于历史样本的性能比较，面向计算时效、异常资料条件和持续稳定性的系统检验仍然有限[51,59]。因此，评价工作需要由模型间精度比较进一步转向误差来源识别、异常条件测试和运行风险判断，使评价结果能够表明模型在何种条件下具备实际应用价值。

3.4　运行适配与版本管理不足

AI 气象预报系统的长期运行包括状态更新、参数适配、业务版本发布和性能监测等相互衔接的环节。每次起报时利用最新观测和背景预报生成当前分析状态属于状态更新或循环资料同化，通常不改变模型参数；随着新增资料持续积累，可以采用在线或增量学习对参数进行周期性调整；当资料来源、分析系统或应用任务发生较大变化时，则需要开展定期再训练或迁移学习；参数或结构调整完成后，还需通过统一回归检验形成正式业务版本；长期性能监测则用于识别输入分布漂移和持续性能下降，并为是否调整模型提供依据。FuXi Weather 同时包含上述不同过程，其中每 6 h 开展的循环资料同化用于更新当前大气状态，而基于回放样本对 FuXi-DA 进行的月度调整属于增量学习[59]。AIFS Single 和 AIFS ENS 由 v1 系列升级至 v2 则属于经过测试发布的业务版本变化[73]，不应与常规起报过程混淆。现有研究更多关注单次参数调整或新版本带来的性能变化，对长期监测如何触发参数适配、候选版本应通过哪些回归检验、关键指标变化幅度以及新版本异常时如何回退等问题讨论仍然不足，尚需建立更加完整、可追溯的运行管理方法。

总体来看,AI 气象预报主要瓶颈依次涉及资料与状态表征、未来状态生成、性能评价与业务验证以及长期运行管理。资料和初始状态决定模型能够获得的信息,预报生成机制影响天气过程和极端事件的表达,性能评价用于判断这些能力在何种条件下成立,持续监测和版本管理则决定模型能否长期稳定使用。

4 AI 气象预报分层融合路径与未来展望

针对资料与状态表征、预报结果生成、性能评价及长期运行中存在的问题,AI 气象预报需要从数据、模型、评价和业务 4 个层面形成相互衔接的融合路径,如图 3 所示。

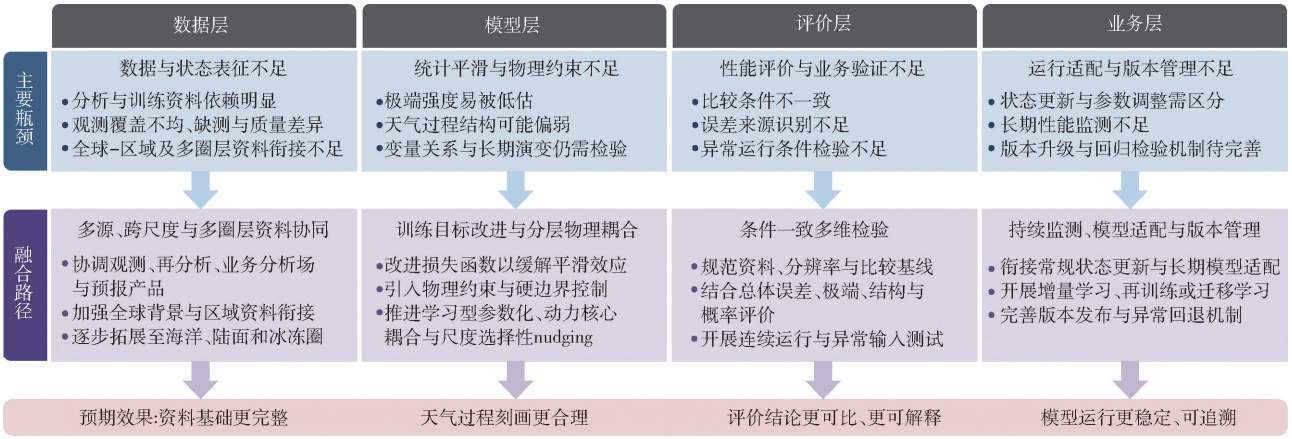


图 3 AI 气象预报在不同层次主要瓶颈、融合路径及预期效果

Fig. 3 Major bottlenecks, integration pathways, and expected outcomes of AI-based weather forecasting across different forecasting levels

4.1 数据层:多源、跨尺度与多圈层资料协同

数据层应协调使用观测资料、再分析资料、业务分析场和预报产品,提高历史训练资料与实际运行输入的一致性,并加强不同资料在时间、空间和变量上匹配。针对区域精细化不足,可利用全球模型提供大尺度环流和边界信息,并结合区域雷达、站点、地形资料补充局地特征,OneForecast 和 STCast 在全球背景信息与区域高分辨率预报之间的嵌套衔接、边界条件传递与注意力融合机制提供了相应探索[62-63]。随着预报时效延长,资料范围还应由大气内部变量扩展至海洋、陆面和冰冻圈,推动 AI 气象预报向地球系统基础模型发展,Aurora 通过异质地球系统资料预训练,并面向天气、海浪和大气成分等任务进行微调,体现了跨资料和跨任务迁移的潜力[55];NeuralGCM 则在统一框架中连接中期天气预报与气候模拟,为天气-气候一体化建模提供了探索[67]。后续应进一步建立大气、海洋、陆面和冰冻圈资料的动态衔接机制,明确不同圈层信息对区域精细化预报和中长期预报的实际贡献。

4.2 模型层:训练目标改进与分层物理耦合

对确定性回归引起的统计平滑,应首先改进未来状态的训练目标和生成方式,MSE 和 MAE 以单一确定性结果逼近未来状态,易在天气演变存在不确定性时输出条件均值或中位状态。扩散模型和流匹配等生成式方法则通过学习给定初始状态下未来状态分布,生成多种具有不同位置、强度和空间结构的可能结果,因而能够减少单一回归结果的平均化。GenCast 采用扩散模型生成天气轨迹,ArchesWeatherGen 利用流匹配学习确定性预报与实际状态之间的残差分布,为缓解预报场平滑提供了新的实现方式[53,69]。后续应将生成式目标与极端样本强化、空间结构保持和物理约束协同设计,使模型在缓解统计平滑的同时,更好地保持局地强度和天气系统结构。

针对物理约束不足,应根据物理信息进入预报过程的位置区分不同技术路径,输出订正和产品级融合主要作用于已有预报结果,能够修正系统偏差或利用不同产品的互补信息,但不能改变基础模型内部的动力演变。物理启发损失通过在训练目标中加入热力、动能、质量或水汽关系等惩罚项,对不合理结果形成软约束,其效果受到物理关系简化程度和权重设置的影响[70-71];硬约束或物理编码架构则将非负性、取值范围、对称性或特定守恒关系直接写入输出层或网络结构,能够保证预先设定的条件,但不能代表完整大气过程均满足物理规律。更深层的结合是将 AI 引入模式积分过程,学习型参数化利用神经网络描述云、辐射、湍流和降雨等未解析过程,动力核心-AI 模块耦合则使学习型模块与动力方程共同积分。NeuralGCM 将可微分大气动力核心与学习型物理参数化联合训练,属于动力核心-AI 模块耦合[67];AIFS-IFS 混合试验通过谱 nudging 使 IFS 大尺度虚温和涡度向 AIFS 预报靠拢,同时保留小

尺度过程在 IFS 动力和物理作用下演变，属于尺度选择性耦合[74]。后续应依据输入资料、耦合位置、物理约束强度和预报任务，选择与具体问题相匹配的融合方式，避免将结果订正、产品组合和动力过程耦合笼统归为同一技术层级。因此，预报生成层应形成由训练目标改进、物理损失约束和结构约束，逐步深入到学习型参数化、动力核心－AI 模块耦合及尺度选择性 nudging 的分层路径，并根据统计平滑、变量关系和动力演变等具体问题选择相应的结合方式。

4.3　评价层：条件一致多维检验

评价层应明确训练资料、初始状态、空间分辨率、预报变量、预报时效、验证资料、重网格方式和比较基线，只有试验条件基本一致时，结果才适合直接比较[33,65]。评价内容应与任务对应，预报产品订正关注相对于原始产品的增益；分析场初始化模型综合检验总体技巧、极端事件和空间结构；观测前端系统区分状态估计与未来预报误差；概率集合则同时检验 CRPS、集合离散度和概率可靠性。面向业务应用，还应增加观测资料缺失、异常输入、输入延迟、分析系统变化和连续运行等条件下检验，并结合站点、探空和卫星等独立观测开展补充验证。通过建立条件一致、任务对应和覆盖异常运行情景的评价方法，可将评价结果由单一精度比较转化为对模型适用条件、误差来源和运行风险的综合判断。

4.4　业务层：持续监测、模型适配与版本管理

业务层应建立常规起报与长期维护相衔接的运行机制，每次起报时，利用最新观测和背景预报更新当前大气状态或开展循环资料同化，通常不改变模型参数，FuXi Weather 每 6 h 开展循环资料同化，而基于历史回放样本对 FuXi－DA 进行月度调整属于增量学习，两者分别作用于当前状态和模型参数[59]；在长期运行中，应持续监测输入资料分布以及模型在不同区域、天气类型和预报时效下表现；当新增资料逐步积累且输入体系变化较小时，可采用在线或增量学习进行参数调整，当资料来源、业务分析系统或应用任务发生较大变化时，则需通过定期再训练或迁移学习重新适配，如 FengWu 由 ERA5 向业务分析资料的适配[50]。参数、结构或预报变量调整后，应形成候选版本，并在统一条件下开展回归检验，综合评价总体技巧、极端事件、概率可靠性、计算时效及异常输入条件下的稳定性，检验通过后再发布为正式业务版本。ECMWF 将 AIFS Single v1.1 和 AIFS ENS v1 升级至 v2，并新增波浪分量和数据驱动积雪覆盖预报[73]，体现了 AI 预报系统在模型方法、预报变量和产品配置调整后进行正式版本发布的业务实践。后续应进一步明确长期监测、模型适配和版本发布之间的触发关系，并保留上一稳定版本，以保障业务运行的连续性和版本变化的可追溯性。

数据层、模型层、评价层和业务层分别对应资料与状态表征、预报结果生成、性能判断和长期运行。通过多源、跨尺度和多圈层资料协同，发展生成式概率建模和不同深度的物理耦合，在条件一致的基础上开展多维业务检验，并建立持续学习与版本管理机制，可推动 AI 气象预报由单次模型性能提升转向预报系统整体能力的持续改进，为农业生产、水资源管理和灾害防御提供更加稳定可靠的气象信息。

5　结束语

AI 气象预报已经由数值模式结果订正和统计后处理，逐步拓展至全球确定性与集合预报、观测资料驱动的临近预报，以及连接状态估计、资料同化和未来预报的学习型预报系统。本文依据模型在实际预报中直接使用的输入资料及 AI 承担的核心计算环节，将现有方法归纳为预报产品驱动的后处理与融合方法、分析场初始化的状态预报模型和观测前端驱动的学习型预报系统。3 类方法输入条件、预报任务和误差来源存在差异，其性能不能脱离初始状态、空间分辨率、验证资料、比较基线和评价指标进行解释。总体预报技巧、极端事件、空间过程、物理一致性和概率可靠性分别反映模型能力的不同方面，单一 RMSE、ACC 或 CRPS 结果不能证明模型在极端强度、天气系统结构、物理合理性和业务稳定性等方面同步改善。现有研究表明，AI 模型已具备全球多变量、多时效预报和快速推理能力，但在数据与状态表征、确定性回归造成的统计平滑、物理约束、评价结果向业务条件延伸以及长期运行管理等方面仍有待完善。未来应加强多源、跨尺度和多圈层资料的协同利用，针对统计平滑改进训练目标和状态生成方式，并根据物理信息进入预报过程的位置，合理选择物理损失约束、学习型参数化、动力核心－AI 模块耦合和尺度选择性 nudging 等路径；同时建立条件一致、任务对应的多维业务检验方法，明确状态更新、增量学习、再训练与迁移学习、回归检验、版本发布和长期性能监测之间的关系。AI 气象预报的发展不宜简单理解为对 NWP 的替代或预报结果的直接叠加，而是资料输入、预报生成、性能评价和业务运行全过程的协同完善。

参 考 文 献

[1] 董林, 贾莉, 王月冬, 等. 气候变化背景下极端天气的新特征与防御技术对策[J]. 中国工程科学, 2026, 28(2): 275 - 286.
Dong Lin, Jia Li, Wang Yuedong, et al. New characteristics of extreme weather in the context of climate change and defense countermeasures[J]. Strategic Study of CAE, 2026, 28(2): 275 - 286. (in Chinese)

[2] 毕馨元, 代刊. 人工智能在气象预报中的突破性进展与科学价值[J]. 科学通报, 2025, 70(27): 4638 - 4645.
Bi Xinyuan, Dai Kan. Breakthrough progress and scientific value of artificial intelligence in weather forecasting[J]. Chinese Science Bulletin, 2025, 70(27): 4638 - 4645. (in Chinese)

[3] 王会军, 孙建奇, 陈活泼, 等. 全球变暖加速和气候极端化:2024 年中国气候研究重大进展速评[J]. 大气科学学报, 2025, 48(1): 1 - 7.
Wang Huijun, Sun Jianqi, Chen Huopo, et al. Accelerated global warming and climate extremization: a brief review of major advances in climate research in China in 2024[J]. Transactions of Atmospheric Sciences, 2025, 48(1): 1 - 7. (in Chinese)

[4] Swain D L, Prein A F, Abatzoglou J T, et al. Hydroclimate volatility on a warming Earth[J]. Nature Reviews Earth & Environment, 2025, 6(1): 35 - 50.

[5] 祝亚丽, 刘洋, 孔祥慧, 等. 中国高温、干旱及其复合事件的研究进展和展望[J]. 大气科学学报, 2025, 48(1): 26 - 36.
Zhu Yali, Liu Yang, Kong Xianghui, et al. Research progress and prospects of high-temperature, drought and compound events in China[J]. Transactions of Atmospheric Sciences, 2025, 48(1): 26 - 36. (in Chinese)

[6] Jiang L, Zhang J, Bai L, et al. Increased frequency and severity of global compound dry and heat wave events in a daily scale [J]. Journal of Hydrology, 2025, 654: 132857.

[7] Zhang Y, Hao Z, Chen Y, et al. Increases in global hot droughts across multiple timescales[J]. Science China Earth Sciences, 2025, 68(10): 3189 - 3199.

[8] Guo J, Wang F, Wen Y, et al. Rising compound hot-dry extremes engendering more inequality in human exposure risks[J]. NPJ Natural Hazards, 2025, 2(1): 66.

[9] Tang Y, Luo M, Wu S, et al. Increasing synchrony of extreme heat and precipitation events under climate warming[J]. Geophysical Research Letters, 2025, 52(8): e2024GL113021.

[10] 蒋云钟, 刘珂, 杨明祥. 黄河流域气象水文预报模型技术研究综述与展望[J]. 人民黄河, 2025, 47(11): 13 - 19.
Jiang Yunzhong, Liu Ke, Yang Mingxiang. Review and prospects of meteorological and hydrological forecasting model technologies in the Yellow River Basin[J]. Yellow River, 2025, 47(11): 13 - 19. (in Chinese)

[11] Chen J, Zhu Y, Duan W, et al. A review on development, challenges, and future perspectives of ensemble forecast[J]. Journal of Meteorological Research, 2025, 39(3): 534 - 558.

[12] 张小玲, 金荣花, 代刊, 等. 中央气象台人工智能气象应用发展及思考[J]. 大气科学学报, 2025, 48(3): 353 - 365.
Zhang Xiaoling, Jin Ronghua, Dai Kan, et al. Development and reflections on artificial intelligence applications in meteorological operations at the National Meteorological Centre[J]. Transactions of Atmospheric Sciences, 2025, 48(3): 353 - 365. (in Chinese)

[13] 王开存. 全球大气再分析及其在中国的适用性[J]. 大气科学学报, 2025, 48(4): 529 - 555.
Wang Kaicun. Global atmospheric reanalyses and their applicability in China[J]. Transactions of Atmospheric Sciences, 2025, 48(4): 529 - 555. (in Chinese)

[14] 陈昊明, 李建, 李普曦, 等. 中国数值模式评估方法的研究进展综述[J]. 气象学报, 2025, 83(3): 566 - 581.
Chen Haoming, Li Jian, Li Puxi, et al. Review of research progress in evaluation methods for numerical models in China[J]. Acta Meteorologica Sinica, 2025, 83(3): 566 - 581. (in Chinese)

[15] 于丹, 李萍, 李婧华, 等. ECMWF 成立 50 周年:气象预报旗舰机构半个世纪发展的解读[J]. 气象科技进展, 2026, 16(2): 21 - 34.
Yu Dan, Li Ping, Li Jinghua, et al. ECMWF's 50th anniversary: interpreting half a century of development of a flagship weather forecasting institution[J]. Advances in Meteorological Science and Technology, 2026, 16(2): 21 - 34. (in Chinese)

[16] Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction[J]. Nature, 2015, 525(7567): 47 - 55.

[17] 马雷鸣, 鲍旭炜. 数值天气预报模式物理过程参数化方案的研究进展[J]. 地球科学进展, 2017, 32(7): 679 - 687.
Ma Leiming, Bao Xuwei. Research progress on physical process parameterization schemes in numerical weather prediction models[J]. Advances in Earth Science, 2017, 32(7): 679 - 687. (in Chinese)

[18] 马刚, 黄静, 巩欣亚, 等. 数值预报中气象卫星资料同化前处理技术进展[J]. 应用气象学报, 2024, 35(2): 142 - 155.
Ma Gang, Huang Jing, Gong Xinya, et al. Review of preprocessing techniques for meteorological satellite data assimilation in numerical prediction[J]. Journal of Applied Meteorological Science, 2024, 35(2): 142 - 155. (in Chinese)

[19] 汪海涛. 国内 AI 中期天气预报大模型进展及面临的挑战[J]. 计算机科学与应用, 2024, 14(12): 196 - 206.
Wang Haitao. Progress and challenges of large-scale AI medium-range weather forecasting models in China[J]. Computer Science and Application, 2024, 14(12): 196 - 206. (in Chinese)

[20] 潘林林, 靳双龙, 宋宗朋, 等. 风能和太阳能短中期气象预报技术及其最新进展[J]. 地球科学进展, 2026, 41(1): 73 - 86.
Pan Linlin, Jin Shuanglong, Song Zongpeng, et al. Short- and medium-term meteorological forecasting technologies for wind and Solar energy and their latest advances[J]. Advances in Earth Science, 2026, 41(1): 73 - 86. (in Chinese)

[21] 石宜衡, 王琦祎, 孙婧, 等. 气象大模型计算资源评估技术方法[J]. 数据与计算发展前沿(中英文), 2025, 7(4): 182 - 195.
Shi Yiheng, Wang Qiyi, Sun Jing, et al. Technical methods for computing resource assessment of large meteorological models [J]. Frontiers of Data & Computing, 2025, 7(4): 182 - 195. (in Chinese)

[22] 周代数，魏杉汀. 人工智能驱动的科学研究第五范式：演进、机制与影响[J]. 中国科技论坛，2024(12)：97－107.
Zhou Daishu，Wei Shanting. The fifth paradigm of scientific research driven by artificial intelligence：evolution，mechanisms and impacts[J]. Forum on Science and Technology in China，2024(12)：97－107.（in Chinese）
[23] 黄小猛，林岩銮，熊巍，等. 数值预报 AI 气象大模型国际发展动态研究[J]. 大气科学学报，2024，47(1)：46－54.
Huang Xiaomeng，Lin Yanluan，Xiong Wei，et al. International development trends of AI weather foundation models for numerical weather prediction[J]. Transactions of Atmospheric Sciences，2024，47(1)：46－54.（in Chinese）
[24] 黄菲妮，蒋世杰，上官微，等. 加强融合：可解释人工智能推动地球系统科学发展[J]. 中国科学：地球科学，2026，56(6)：2025－2043.
Huang Feini，Jiang Shijie，Shangguan Wei，et al. Bridging the gap：explainable AI for advancing Earth system science[J]. Scientia Sinica Terrae，2026，56(6)：2025－2043.（in Chinese）
[25] 黄建平，陈斌. 人工智能技术在未来改进天气预报中的作用[J]. 科学通报，2024，69(17)：2336－2343.
Huang Jianping，Chen Bin. How artificial intelligence is transforming weather forecasting for the future[J]. Chinese Science Bulletin，2024，69(17)：2336－2343.（in Chinese）
[26] Ebert-Uphoff I，Hilburn K. The outlook for AI weather prediction[J]. Nature，2023，619(7970)：473－474.
[27] Hu L，Ma Z，Fu H. Artificial intelligence-based large weather models：advances，comparisons，and prospects[C]// Proceedings of the 2025 International Conference on Embodied Intelligence and Large Models. Association for Computing Machinery，2026：191－198.
[28] Luo J J，Xia J，Pan B，et al. AI for atmosphere-ocean sciences：advancements，challenges and ways forward[J]. National Science Review，2026，13(5)：nwag063.
[29] Waqas M，Humphries U W，Chueasa B，et al. Artificial intelligence and numerical weather prediction models：a technical survey[J]. Natural Hazards Research，2025，5(2)：306－320.
[30] 田伟，秦子航，乔建权，等. 深度学习在气象数据挖掘中的应用[J]. 中国科技论文，2025，20(4)：277－286.
Tian Wei，Qin Zihang，Qiao Jianquan，et al. Application of deep learning in meteorological data mining[J]. China Sciencepaper，2025，20(4)：277－286.（in Chinese）
[31] 陈鹤，蔡荣辉，陈静静，等. 基于深度学习方法的气温预报技术应用与评估[J]. 气象，2022，48(11)：1373－1383.
Chen He，Cai Ronghui，Chen Jingjing，et al. Application and evaluation of air temperature forecasting technology based on deep learning[J]. Meteorological Monthly，2022，48(11)：1373－1383.（in Chinese）
[32] 杨绚，代刊，朱跃建. 深度学习技术在智能网格天气预报中的应用进展与挑战[J]. 气象学报，2022，80(5)：649－667.
Yang Xuan，Dai Kan，Zhu Yuejian. Progress and challenges of deep learning technology in intelligent grid weather forecasting[J]. Acta Meteorologica Sinica，2022，80(5)：649－667.（in Chinese）
[33] Ben Bouallègue Z，Clare M C A，Magnusson L，et al. The rise of data-driven weather forecasting：a first statistical assessment of machine learning-based weather forecasts in an operational-like context[J]. Bulletin of the American Meteorological Society，2024，105(6)：E864－E883.
[34] Bi K，Xie L，Zhang H，et al. Accurate medium-range global weather forecasting with 3D neural networks[J]. Nature，2023，619(7970)：533－538.
[35] Chen L，Zhong X，Zhang F，et al. FuXi：a cascade machine learning forecasting system for 15-day global weather forecast[J]. NPJ Climate and Atmospheric Science，2023，6(1)：190.
[36] Lam R，Sanchez-Gonzalez A，Willson M，et al. Learning skillful medium-range global weather forecasting[J]. Science，2023，382(6677)：1416－1421.
[37] 陶天友，邓鹏，王浩，等. 基于机器学习的极端风场短时预测研究进展与思考[J]. 空气动力学学报，2025，43(5)：78－91.
Tao Tianyou，Deng Peng，Wang Hao，et al. Progress and perspectives on short-term prediction of extreme wind fields based on machine learning[J]. Acta Aerodynamica Sinica，2025，43(5)：78－91.（in Chinese）
[38] 欧阳文宇，张弛，马昊然，等. 人工智能驱动水文预报与水库调度研究的探索与思考[J]. 水利学报，2025，56(9)：1119－1131.
Ouyang Wenyu，Zhang Chi，Ma Haoran，et al. Exploration and consideration of AI-driven hydrological forecasting and reservoir operation[J]. Journal of Hydraulic Engineering，2025，56(9)：1119－1131.（in Chinese）
[39] 王悟信，倪维成，袁太康，等. 资料同化中的人工智能技术：新兴方法、关键挑战与未来展望[J]. 中国科学：地球科学，2026，56(4)：1221－1240.
Wang Wuxin，Ni Weicheng，Yuan Taikang，et al. Artificial intelligence in data assimilation：emerging methods，key challenges and future prospects[J]. Scientia Sinica Terrae，2026，56(4)：1221－1240.（in Chinese）
[40] Zhang Z，Fischer E，Zscheischler J，et al. Physics-based models outperform AI weather forecasts of record-breaking extremes[J]. Science Advances，2026，12(18)：eaec1433.
[41] Su Y，Cheng Q，He Y，et al. An Evolution-Unet-ConvNeXt approach based on feature fusion for enhancing the accuracy of short-term precipitation forecasting[J]. Atmospheric Research，2025，317：107984.
[42] Ren X，Li X，Ren K，et al. Deep learning-based weather prediction：a survey[J]. Big Data Research，2021，23：100178.
[43] Zeng Y，Wang J，Wu J. A review of neural networks in precipitation prediction[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2026，19：12956－12986.
[44] 赵铜铁钢，张弛，田雨，等. 全球气象预报驱动流域水文预报研究进展与展望[J]. 水科学进展，2024，35(1)：156－166.
Zhao Tongtiegang，Zhang Chi，Tian Yu，et al. Research progress and prospects of basin hydrological forecasting driven by global weather forecasts[J]. Advances in Water Science，2024，35(1)：156－166.（in Chinese）
[45] Li H，Yu C，Xia J，et al. A model output machine learning method for grid temperature forecasts in the Beijing area[J]. Advances in Atmospheric Sciences，2019，36(10)：1156－1170.
[46] Kong W，Li H，Yu C，et al. A deep spatio-temporal forecasting model for multi-site weather prediction post-processing[J].

Communications in Computational Physics, 2022, 31(1): 131 - 153.
[47] Zheng J, Ling Q, Li J, et al. Improving the short-range precipitation forecast of numerical weather prediction through a deep learning-based mask approach[J]. Advances in Atmospheric Sciences, 2024, 41(8): 1601 - 1613.
[48] Sønderby C K, Espeholt L, Heek J, et al. MetNet: a neural weather model for precipitation forecasting[PP/OL]. V1. arXiv (2020 - 03 - 24)[2026 - 08 - 05]. https://doi.org/10.48550/arXiv.2003.12140.
[49] Pathak J, Subramanian S, Harrington P, et al. FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators[PP/OL]. V1. arXiv(2022 - 02 - 22)[2026 - 08 - 05]. https://doi.org/10.48550/arXiv.2202.11214.
[50] Chen K, Han T, Ling F, et al. The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time[J]. Communications Earth & Environment, 2025, 6(1): 518.
[51] Moldovan G, Pinnington E, Prieto Nemesio A, et al. AIFS Single 1.1.0: an update to ECMWF's machine-learned weather forecast model AIFS[J]. Geoscientific Model Development, 2026, 19(10): 4703 - 4724.
[52] Lang S, Alexe M, Clare M C A, et al. AIFS - CRPS: ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score[J]. NPJ Artificial Intelligence, 2026, 2(1): 18.
[53] Price I, Sanchez-Gonzalez A, Alet F, et al. Probabilistic weather forecasting with machine learning[J]. Nature, 2025, 637(8044): 84 - 90.
[54] Zhong X, Chen L, Li H, et al. FuXi - ENS: a machine learning model for efficient and accurate ensemble weather prediction [J]. Science Advances, 2025, 11(44): eadu2854.
[55] Bodnar C, Bruinsma W P, Lucic A et al. A foundation model for the Earth system[J]. Nature, 2025, 641(8065): 1180 - 1187.
[56] Zhang Y, Long M, Chen K, et al. Skilful nowcasting of extreme precipitation with NowcastNet[J]. Nature, 2023, 619(7970): 526 - 532.
[57] 盛杰, 金荣花, 张小雯, 等. 临近气象预报大模型"风雷"V1 版本检验及个例评估[J]. 气象, 2025, 51(4): 389 - 399.
Sheng Jie, Jin Ronghua, Zhang Xiaowen, et al. Verification and case evaluation of Fenglei V1, a large model for weather nowcasting[J]. Meteorological Monthly, 2025, 51(4): 389 - 399. (in Chinese)
[58] Allen A, Markou S, Tebbutt W, et al. End-to-end data-driven weather prediction[J]. Nature, 2025, 641(8065): 1172 - 1179.
[59] Sun X, Zhong X, Xu X, et al. A data-to-forecast machine learning system for global weather[J]. Nature Communications, 2025, 16(1): 6658.
[60] Xu X, Sun X, Han W, et al. FuXi - DA: a generalized deep learning data assimilation framework for assimilating satellite observations[J]. NPJ Climate and Atmospheric Science, 2025, 8(1): 156.
[61] Alexe M, Boucher E, Lean P, et al. GraphDOP: towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations[PP/OL]. V1. arXiv(2024 - 12 - 20)[2026 - 08 - 05]. https://doi.org/10.48550/arXiv.2412.15687.
[62] Gao O Y, Wu H, Shu R, et al. OneForecast: a universal framework for global and regional weather forecasting[C]// Proceedings of Machine Learning Research, 2025, 267: 18658 - 18697.
[63] Chen N H, Han T, Zhang G J, et al. STCast: adaptive boundary alignment for global and regional weather forecasting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2026: 20586 - 20596.
[64] Purohit K, Sinha M, Nanjundiah R S. AI-based multimodel superensemble for improved weather prediction[C]//2024 IEEE 31st International Conference on High Performance Computing, Data and Analytics Workshop. IEEE, 2024: 47 - 51.
[65] Rasp S, Dueben P D, Scher S, et al. WeatherBench: a benchmark data set for data-driven weather forecasting[J]. Journal of Advances in Modeling Earth Systems, 2020, 12(11): e2020MS002203.
[66] Li H, Du Y. Complementary error structures of AI and numerical models in forecasting boundary-layer jets over the South China Sea[J]. NPJ Climate and Atmospheric Science, 2026, 9: 141.
[67] Kochkov D, Yuval J, Langmore I, et al. Neural general circulation models for weather and climate[J]. Nature, 2024, 632(8027): 1060 - 1066.
[68] Wang B, Zhu S, Wu L, et al. Global weather forecasting through deep learning: applications and challenges[J]. Technology Review for Carbon Neutrality, 2025, 1: 9550021.
[69] Couairon G, Singh R, Charantonis A, et al. ArchesWeatherGen: skillful and compute-efficient probabilistic weather forecasting with machine learning[J]. Science Advances, 2026, 12(17): eadx2372.
[70] Saleem H, Salim F, Purcell C. PINN-Cast: exploring the role of continuous-depth NODE in transformers and physics-informed loss as soft physical constraints in short-term weather forecasting[C]// Neumann P, Puma M J, Lees M H, et al. Computational Science: ICCS 2026, 2026, 16784: 109 - 122.
[71] Yu H, Lin L, Zhang Z, et al. Physics-informed spatiotemporal deep learning for multivariate atmospheric forecasting[J]. Expert Systems with Applications, 2026, 315: 131721.
[72] Hu J Y, Higuera-Mendiola I, Sturm P O, et al. Toward mechanistic interpretability of an AI foundation model fine-tuned for atmospheric chemistry[PP/OL]. V1. arXiv(2026 - 07 - 22)[2026 - 08 - 05]. https://doi.org/10.48550/arXiv.2607.20778.
[73] European Centre for Medium-Range Weather Forecasts. AIFS machine learning data[DS/OL]. (2026 - 05 - 12)[2026 - 08 - 05]. https://www.ecmwf.int/en/forecasts/datasets/aifs-machine-learning-data.
[74] Polichtchouk I, Clare M, Chantry M, et al. Hybrid forecasting: nudging large scales of the IFS to the deterministic AIFS[J]. ECMWF Newsletter, 2025(184): 12 - 13.