中国科学: 地球科学 2024 年 第 54 卷 第 12 期: 3734 ~ 3747
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文

FuXi-Extreme: 利用扩散模型改进极端降雨和风速预报

仲晓辉†, 陈磊†, 刘俊, 林晨森, 漆远, 李昊*

复旦大学人工智能创新与产业研究院, 上海 200433
* 通讯作者, E-mail: lihao_lh@fudan.edu.cn
† 同等贡献

收稿日期: 2023-10-31; 收修改稿日期: 2024-08-19; 接受日期: 2024-09-10; 网络版发表日期: 2024-10-15
中国博士后科学基金项目(GZB20240154)资助

摘要 基于机器学习(Machine Learning, ML)的天气预报模型近些年取得了显著进展, 展示了优越的预报性能. 与欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF)的高分辨率预报(High-Resolution Forecasts, HRES)相比, FuXi等先进的基于ML的天气预报模型, 在统计预报指标上表现出色. 然而, 这些模型存在着共同的局限性, 即随着预报时间步长的增加, 预报结果趋于平滑, 导致极端天气事件强度的低估. 为了解决这一问题, 本文研发了FuXi-Extreme模型. 该模型采用去噪扩散概率模型(Denoising Diffusion Probabilistic Model, DDPM), 增强了FuXi模型在5天预报中的地表预报数据细节. 对极端总降水量(Total Precipitation, TP)、10m风速(10-meter Wind Speed, WS10)和2m温度(2-meter Temperature, T2M)的评估表明, FuXi-Extreme在性能上优于FuXi和HRES. 此外, 基于国际热带气旋最佳路径资料集(International Best Track Archive for Climate Stewardship, IBTrACS)的评估显示, 与HRES相比, FuXi和FuXi-Extreme在热带气旋(Tropical Cyclone, TC)路径预报方面表现优异, 但在TC强度预报方面仍有不足.

关键词 FuXi, 扩散模型, 天气预报, 极端天气

1 引言

气候变化导致包括强降雨和强风在内的极端天气事件频率和强度增加, 对人类社会产生了严重的负面影响(Ebi等, 2021; T. Parmesan和Morecroft, 2022). Kotz等(2022)进行了一项综合评估, 分析了强降雨对区域生产总值(Gross Regional Product, GRP)的影响. 研究表明, 雨天数量和极端日降雨量的增加导致全球宏观经济增长率明显下降. 此外, 极端强风可能导致风力发电机停机, 从而扰乱风电生产(Liu等, 2023). 为了减少这些极端事件造成的损失, 就需要准确预报提供早期预警和及时信息. 因此, 加强这些极端事件的预报应成为天气预报机构的一项首要任务. 基于物理和动力的数值天气预报(Numerical Weather Prediction,

中文引用格式: 仲晓辉, 陈磊, 刘俊, 林晨森, 漆远, 李昊. 2024. FuXi-Extreme: 利用扩散模型改进极端降雨和风速预报. 中国科学: 地球科学, 54(12): 3734–3747, doi: 10.1360/N072023-0274
英文引用格式: Zhong X, Chen L, Liu J, Lin C, Qi Y, Li H. 2024. FuXi-Extreme: Improving extreme rainfall and wind forecasts with diffusion model. Science China Earth Sciences, 67(12): 3696–3708, https://doi.org/10.1007/s11430-023-1427-x

© 2024 《中国科学》杂志社

NWP)模式已被广泛用于各种天气预报. 然而, 由于极端事件空间和时间尺度多变, 涵盖了从小尺度水滴相互作用到大尺度天气系统的多重尺度的相互作用, 这对NWP模式来说是一个重要挑战(Hess和Boers, 2022).

近年来, 机器学习(ML)模型作为传统NWP模式的替代方案取得了快速进展, 在统计指标上的表现优于世界领先的基于物理的NWP模式, 包括欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF)的高分辨率预报模式(High-Resolution Forecasts, HRES)(Haiden等, 2021; Bi等, 2023; Chen等, 2023b; Lam等, 2023). FuXi模型是优化的ML级联模型, 三个连续预报时间段(0~5、5~10和10~15天)分别称为FuXi-Short、FuXi-Medium和FuXi-Long (Chen等, 2023b). FuXi模型的研发基于ECMWF的39年ERA5再分析数据(Hersbach等, 2020), 能够产生未来15天的0.25°空间分辨率和逐6小时的全球天气预报, 其性能与ECMWF集合平均值相当. 然而, 基于ML的天气预报模型存在生成趋于平滑预报的问题, 且随着预报时间步长的增加, 问题变得更加明显, 从而低估了极端天气事件的强度(Pathak等, 2022; Rasp等, 2024).

扩散模型(Sohl-Dickstein等, 2015; Ho等, 2020; Saharia等, 2022; Croitoru等, 2023)在生成高精度图像方面能力卓越, 近年来在计算机视觉领域受到了极大关注. 扩散模型通过一个正向过程逐步向输入数据添加高斯噪声, 通过反向过程逐渐消除噪声数据, 使数据恢复至原始状态. 受扩散模型在高质量图像生成方面的启发, Chen等(2023a)开发了一种基于扩散模型的超分辨率模型, 该模型在设计上具有一种固有能力, 能够捕捉精细尺度结构, 成功地从分辨率为1.40525°的低分辨率预报数据中重建出0.25°的高质量预报. 扩散模型通过学习添加和去除噪声的随机过程, 能够从噪声中重建数据, 深入学习到数据分布的尾部, 而这里通常是极端值的位置. 因此, 扩散模型具备更准确预报极端事件的潜力.

我们引入了针对地表变量极端值的优化预报的FuXi-Extreme模型, 通过采用去噪扩散概率模型(Denoising Diffusion Probabilistic Model, DDPM)(Sohl-Dickstein等, 2015; Ho等, 2020)来恢复FuXi模型在5天预报中生成的地表预报的精细尺度的细节. 参照ERA5再分析数据集, 我们评估了极端总降水量(Total Precipitation, TP)、10m风速(10-meter Wind Speed, WS10)和2m温度(2-meter Temperature, T2M)的预报能力. 此外, 使用国际热带气旋最佳路径资料集(International Best Track Archive for Climate Stewardship, IBTrACS) (Knapp等, 2010, 2018)评估了热带气旋(Tropical Cyclone, TC)的预报性能.

2　数据集

2.1　ERA5

FuXi-Extreme模型与FuXi模型一样, 采用ECMWF的ERA5再分析数据集进行训练, 空间分辨率为0.25°, 时间分辨率为6小时. FuXi-Extreme模型训练选用了2012年至2017年间的ERA5数据子集. FuXi模型可同时预报高空和地表的气象变量的预报, 而FuXi-Extreme着重预报5个地表气象要素, 包括T2M、10m高度的u风分量(10-meter U Wind Component, U10)、10m高度的v风分量(10-meter V Wind Component, V10)、平均海平面气压(Mean Sea-Level Pressure, MSL)和TP. 本文还采用了2018年的ERA5数据集评估和对比了ECMWF HRES、FuXi和FuXi-Extreme的预报性能.

2.2　热带气旋数据集

我们采用美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)提供的IBTrACS(Knapp等, 2010, 2018)数据集对TC预报进行了评估. IBTrACS汇集了全球范围内所有可获得的TC最佳路径数据, 每条记录代表TC的6小时时间序列, 通过纬度和经度坐标表示中心位置, 并包括特定时间和强度等其他相关特征. 根据已建立的TC预报评估标准(Magnusson等, 2021), 在FuXi、FuXi-Extreme和HRES同时识别出TC时, 进行TC路径评估, 以确保所有模型均在统一事件集上进行评估.

为了与ECMWF HRES进行比较, 我们使用了THORPEX交互式全球大集合档案(TIGGE)(THORPEX Interactive Grand Global Ensemble, TIGGE) (Bougeault等, 2010; Swinbank等, 2016), 该档案中包含了基于ECMWF采用的TC路径提取算法估算的TC路径. ECMWF的TC路径数据以XML格式存储, 涵盖了来自HRES和ECMWF集合预报的TC预报结果, 我们根据“forecast”标签从中提取了HRES的预报数据.

除了IBTrACS数据集，我们还将ECMWF的TC追踪算法应用于ERA5数据集，以提取用于TC预报评估的路径和强度信息.

3 方法

3.1 FuXi-Extreme模型

如图1所示，FuXi-Extreme由FuXi模型和DDPM模型构成.

3.1.1 基础FuXi模型

FuXi是一种自回归模型, 通过结合前一时间步X^{k-1}和当前时间步的X^{k}数据来预报未来一时间步X^{k+1}. 输入数据涵盖高空及地表变量，维度为2×70×721×1440. FuXi基础模型包含三个主要组成部分：立方体嵌入(Cube Embedding)用于减少输入数据的空间和时间维度; U-Transformer由一个下采样块、48 个重复的Swin Transformer V2(Liu等, 2022)模块以及一个上采样模块组成, 用于处理嵌入后的数据; 最后通过全连接层输出预报结果. 在此项工作中, 用于0~5天预报的FuXi模型参数已冻结.

3.1.2 去噪扩散概率模型(DDPM)

去噪扩散模型(Denoising Diffusion Probabilistic Models, DDPM)因其能够生成具备精细细节的高质量图像而受到认可，在图像和视频生成领域中得到广泛应用. 尤其是在图像超分辨率和图像恢复等任务中(Saharia等, 2022), DDPM能有效地恢复图像中的精细尺度结构，即使这些图像可能呈现出平滑或模糊的特征. 在这项研究中，我们利用DDPM模型优化FuXi的预报, 特别是极端值的预报.

DDPM模型(Ho等, 2020)包括正向和反向两个扩散过程. 给定一个符合真实数据分布的样本$\mathrm{X}_0 \sim q(\mathrm{X}_0)$, 正向扩散过程以马尔可夫链的形式, 按照预定义的方差序列, 在T个步骤上逐步将高斯噪声引入样本: $\beta_1, \cdots, \beta_T$. 方差序列对于扩散模型的有效运行至关重要，因为它会显著影响生成图像的质量和模型的收敛.

$$q(\mathrm{X}_t \mid \mathrm{X}_{t-1}) = N\left(\mathrm{X}_t; \sqrt{1-\beta_t}\,\mathrm{X}_{t-1}, (\beta_t) I\right), \tag{1}$$

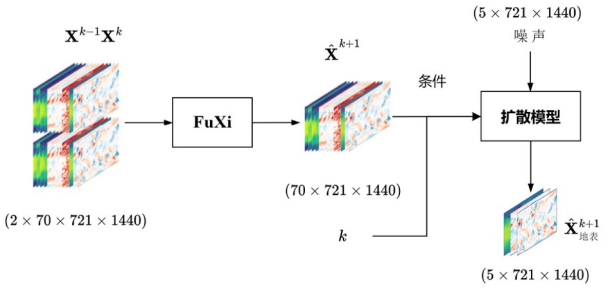


图 1　FuXi-Extreme模型的结构示意图
模型由FuXi模型和扩散模型组成

其中，I表示单位方差. 当每一步添加的噪声幅度β_t足够小且总步数T足够大时，生成的输出X_T非常接近各向同性的高斯分布. 因此，若学习反向分布$q\left(\mathrm{X}_{t-1} \mid X_t\right)$, 我们可以从高斯噪声分布$\mathcal{N}(0, I)$(即$\mathrm{X}_T$)采样, 并通过反向过程复原出原始$q(\mathrm{X}_0)$的样本. 然而, 计算后验分布$q\left(\mathrm{X}_{t-1} \mid \mathrm{X}_t\right)$具有挑战性，因为它需要依赖整个数据集. 因此，$q\left(\mathrm{X}_{t-1} \mid \mathrm{X}_t\right)$通过一个参数化模型$p_\theta$来近似该后验分布. 由于$q\left(\mathrm{X}_{t-1} \mid \mathrm{X}_t\right)$是高斯分布, 对于足够小的$\beta_t$，$p_\theta$也是高斯分布，并且参数化只涉及均值和方差:

$$p_\theta\left(\mathrm{X}_{t-1} \mid \mathrm{X}_T\right) = N\left(\mathrm{X}_{t-1}; \mu_\theta(X_t, t), \Sigma_\theta(\mathrm{X}_t, t)\right), \tag{2}$$

其中, $\mu_\theta(\mathrm{X}_t, t)$和$\Sigma_\theta(\mathrm{X}_t, t)$分别表示均值和方差.

q和p的组合与变分自动编码器(Variational Auto-encoder, VAE)类似(Kingma and Welling, 2013), 可以采用相似的损失函数优化训练数据的负对数似然值$(log(p_\theta(\mathrm{X}_0)))$. 此外, 可以采用称为“证据下界(Evidence Lower Bound, ELBO)”的变分下界. Ho等(2020)通过简化目标, 即预报每一步的噪声, 获得了更好的结果, 如下所示:

$$L_t = E_{t\sim[1,T],\mathrm{X}_t,\epsilon_t\sim\mathcal{N}(0,I)}\left[\| \epsilon_t - \epsilon_\theta(\mathrm{X}_t, t) \|^2\right], \tag{3}$$

这里t均匀分布于1到T之间, $\mathrm{X}_t \sim q\left(\mathrm{X}_t \mid \mathrm{X}_0\right)$ 通过向X_0添加高斯噪声ϵ而获得, ϵ_θ是预报添加噪声的模型. 该损失函数实质上是前向过程中添加的噪声与模型预报的噪声之间的均方误差(Mean Squared Error, MSE).

DDPM模型将FuXi模型生成的所有变量作为条

件. 在训练过程中, FuXi模型实时预报X^k, 其预报精度随预报时间的延长而降低. 扩散模型可结合额外的输入为条件(Dhariwal和Nichol, 2021; Nichol等, 2021; Ho等, 2022; Ramesh等, 2022; Whang等, 2022). 在本研究中, DDPM模型进一步根据FuXi模型的时间步长<k为条件, 这使得我们能够得到以下后验:

$$p_\theta\left(\mathrm{Y}_{t-1}^k \mid \mathrm{Y}_t^k, \mathrm{X}^k, t, k\right), \tag{4}$$

其中, Y^k代表模型预报的X^k相对应的真值. 我们直接预报原始目标X^k, 而非采用ϵ-预报方法, 参照Ramesh等(2022). 该模型作为去噪函数, 并使用MSE损失进行训练:

$$L = E_{t\sim[1,T],\mathrm{Y}_t^{(k)}\sim q_t}\left[\| \mathrm{Y}^k - f_\theta\left(\mathrm{Y}_t^{(k)}, \mathrm{X}^k, t, k\right)\|^2\right]. \tag{5}$$

3.2　FuXi-Extreme模型训练

在使用Pytorch对FuXi-Extreme模型进行训练过程中(Paszke等, 2017), 只对DDPM模型的参数进行更新, FuXi模型的参数则保持固定. 训练过程需要持续24小时, 完成120000次参数模型更新, 运行在配备了8块Nvidia A100 GPU的集群上. 每块GPU上设置的批处理大小为1. 采用AdamW(Kingma和Ba, 2014; Loshchilov和Hutter, 2017)优化器进行优化, 参数配置: β_1=0.9和β_2=0.95, 初始学习率为2.5×10^{-5}, 权重衰减系数为0.1. 为了防止过拟合, 采用了了Scheduled DropPath (Larsson等, 2016)技术, 设定丢弃比率为0.2.

3.3　评估方法

3.3.1　临界成功指数(CSI)

临界成功指数(Critical Success Index, CSI)(Schaefer, 1990; Wilks, 2011), 也称为风险评分(Threat Score, TS), 是预报评估中广泛使用的指标, 用于量化命中率预报的成功率. CSI的计算公式为$CSI = \frac{TP}{TP+FP+FN}$, 其数值范围从0到1. 数值为0表示没有预报技巧, 而值1代表完美的预报结果. 这里, TP、FP、FN和TN分别表示真阳性(命中)、假阳性(误报)、假阴性(未命中)和真阴性.

在计算TP和WS10的CSI指标时, 需要设置特定的阈值. 对于6小时累计降水量, 我们参考中国国家级降水分级标准(GB/T 28592-2012)的24小时累积降水量, 将其除以4, 以获得6小时累积降水量的阈值分别为2.5、6.25、25、62.5和70mm. 关于风速等级, 我们采用由英国海军上将弗朗西斯·蒲福特(Sir Francis Beaufort)于1805年创建的蒲福风级(Beaufort Scale), 将风速与海上或陆地上的观察条件联系起来. 这些风速阈值设定为13.9、17.2、20.8、24.5、28.5和32.7m/s.

3.3.2　对称极值依赖指数(SEDI)

极端事件通常定义为在变量分布尾部发生的事件, 通常是上尾. 这些事件可以通过超过某个物理变量的特定绝对值或超过气候分布中的某个百分位数来识别. 绝对阈值对于评估损害更为相关, 使用基于百分位的阈值可以帮助生成在不同地区和季节间更具可比性的评分, 也更具适用性. CSI指标采用绝对阈值计算, 不考虑地理位置和时间. 尽管CSI指标对于评估预报质量很有帮助, 但对于极端事件, 其得分往往趋于零值(Stephenson等, 2008). 为了解决这一问题, Ferro和Stephenson(2011)开发了对称极值依赖指数(Symmetric Extremal Dependency Index, SEDI), 用于衡量预报和观测到的极端事件之间关联. SEDI分数的计算公式如下:

$$SEDI = \frac{\log F - \log H - \log(1-F) + \log(1-H)}{\log F + \log H + \log(1-F) + \log(1-H)}, \tag{6}$$

其中, F是误报率($F = \frac{FP}{FP+TN}$), H是命中率($H = \frac{TP}{TP+FN}$). SEDI值的范围从–1到1, 0表示无预报技巧, 1表示完美预报.

不同于计算CSI指标时所用的固定阈值, SEDI的计算考虑了特定地点和时间的阈值. 根据Magnusson等(2014)的方法, 参考气候态是针对每个变量、每月、每日时段以及纬度/经度坐标分别得出的. 这种方法通过减少个别空间位置的日变化和季节性周期的影响, 改善了极端事件的检测. 举例来说, 考虑在6小时内超过5mm降水的情况, 这在干旱的撒哈拉沙漠中可能被视为极端事件, 而在世界上最多雨潮湿的地方之一——夏威夷的大沼泽地(Big Bog, Hawaii)则不然. 在本研究中, 参考气候态包括从1993年至2016年的ERA5再分析数据得出的第90、第95、第98和第99.5百分位数. 如网络版附图S1~S3所示(http://earthcn.scichina.com), 这些百分位数在不同地理区域表现出显著的差异. 关于T2M, 百分位数还显示出一年中不同月份的显著差异.

3.3.3 能量谱

通过对预报模型的能量谱分析，量化其预报结果的平滑度，能量谱揭示了在不同尺度上能量的分布. 在较小尺度上具有更大能量表明预报场中精细尺度结构的主导地位. 根据Rasp等(2024)概述的方法，我们计算了纬度带30°<|*lat*|<60°之间的平均经向能量谱.

3.3.4 热带气旋追踪方法

在本研究中，我们应用了由van der Grijn(2002)提出并由ECMWF开发的TC追踪器算法的改进版本，以应用于FuXi和FuXi-Extreme的预报结果. 根据Lam等(2023)的建议，我们对该算法进行了多项增强，旨在减少FuXi和FuXi-Extreme预报中错误提取TC的情况. 首先，我们概述了标准的ECMWF TC追踪方法，然后详细说明了我们的修改及其依据.

ECMWF的TC追踪算法首先利用TC观测数据来确定TC位置的初步估计. 追踪算法的主要功能是在初始场中通过在观察到的TC位置周围445km半径内找到最低MSL来识别TC. 然后，追踪算法在后续的预报数据中继续进行搜索，直到TC不再被检测到. 具体而言，追踪算法通过结合两个向量的平均值来更新TC位置的估计: (1) 基于当前和先前追踪位置之间位移的进行线性外推, (2) 基于风力引导的平流向量. 该平流向量通过在当前TC位置的200、500、700和850hPa气压层上的纬向(U)和经向(V)风分量的平均值计算得出. 然后，将平均值乘以预报时间步长(dt, 即6小时)($\frac{U200+U500+U700+U850}{4}$, $\frac{V200+V500+V700+V850}{4}$) × d$t$).

在初始预报步骤中(本研究中涉及6小时的预报时长)，只有一个估计位置可用. 因此，线性外推不可行，只能采用风力引导的平流. 在计算出下一步的TC位置估计后，追踪算法评估该估计位置445km半径内的所有局部最小MSL. 然后，寻找满足表1中列出的三个条件的最近候选最小值. 如果没有任何最小值满足所有条件，或者TC位于海拔超过1000m的地形上，则追踪算法认定在预报数据中不存在TC并终止TC追踪.

在Lam等(2023)提出的改进ECMWF追踪算法的基础, 并结合我们对数据集中误追踪情况的分析, 我们对标准的ECMWF追踪算法进行了以下增强.

由于基于ML的天气预报相比NWP模式生成的预报更为平滑，ML预报中相邻点之间的差异在也相应减小. 因此，ML预报的固有平滑性常常导致假局部最小值. 为了缓解这一问题，我们通过使用步长和内核大小为5的平均池化, 将FuXi和FuXi-Extreme预报的空间分辨率从0.25°降低至1.25°. 然后，根据原始5×5网格的平均池化值选择中心候选点, 候选点位于网格的中心位置.

除了利用局部最小MSL值外，我们还加入了10m高度涡度局部最大值，以增加中心候选点数量. 一旦选择了中心候选点，数据将恢复到其原始分辨率. 除了表1中列出的三个条件外，还须满足一个额外条件才能被选为最终的中心点: 通过步长为1、核大小为4的平均池计算出的局部最小Z850. 根据我们的测试，这一调整最为显著. 为了防止因大幅跳跃而导致的误追踪，我们探索了限制位移估计值不超过TC从上一个时间步长移动距离的3倍. 然而, 我们没有对前后6个小时的候选点之间的夹角进行约束，因为有些情况下TC路径的方向会发生显著变化(见图2). 这些调整对于提高我们TC追踪算法的准确性和可靠性至关重要，特别是在应对基于ML的预报带来的挑战和TC追踪的复杂性方面. 此外, 由于FuXi-Extreme专门优化了地表变量, 因此在FuXi-Extreme的预报数据，追踪算法结合使用了FuXi预报的高空大气变量和FuXi-Extreme预报的地表变量来检测TC.

3.3.5 TC路径和强度评估

TC的预报评估包括路径和强度两个方面. 路径的

表 1 ECMWF热带气旋(TC)追踪器中使用的参数、相关水平和相应阈值的概述

参数	高度	半径	阈值	描述
10m风速	地面	278km	>8m/s	仅当TC登陆陆地上需要满足此条件
涡度	850hPa	278km	绝对值≥5×10^{-5}/s	–
位势厚度	850和200hPa	278km		只有在热带气旋转变为温带气旋后, 才需要达到最大厚度

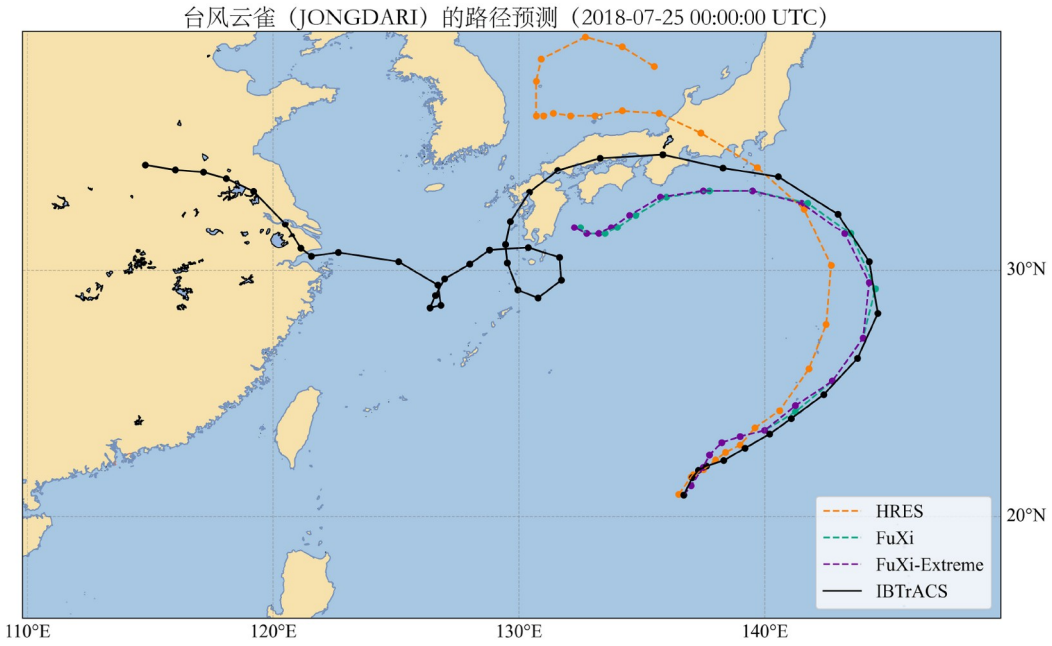


图 2　2018年台风云雀(JONGDARI)的路径预报

HRES、FuXi、FuXi-Extreme的追踪结果分别用橙色线、深绿线和浅紫线表示, IBTrACS的台风路径用黑色线表示, 预报初始时间为2018年7月25日00时UTC

精度通过平均绝对误差(Mean Absolute Error, MAE)量化, 具体计算观测和预报的TC中心位置之间的距离. 强度评估通过TC中心附近的最大WS10和中心的MSL来量化, 并通过均方根误差(Root Mean Square Error, RMSE)来确定TC强度预报的准确性. 该评估涵盖了2018年的90个TC, 仅当这些TC同时记录在IBTrACS(或ERA5)、HRES、FuXi和FuXi-Extreme数据集中时才被纳入评估范围.

4　结果

4.1　极端天气预报的整体统计评估

图3显示了HRES、FuXi和FuXi-Extreme在不同阈值下评估的WS10和6小时累计TP方面的CSI的5天时间序列. 在所有模型中, 随着阈值增加, CSI值逐渐降低. 图3表明, 对于较低的阈值(如2.5mm的TP), FuXi和FuXi-Extreme之间的CSI差异相对较小, 但在更极端的阈值下, 这一差异显著放大. 此外, 在预报初期, FuXi和FuXi-Extreme的CSI值非常接近, 随着预报时长的增加, FuXi-Extreme的表现逐渐超过FuXi. 总体而言, FuXi-Extreme在WS10和TP方面都取得了最高的CSI评分. 关于TP, 在小到中等阈值上, FuXi表现优于HRES, 但在极端阈值(如62.5和70mm的TP)上, FuXi落后于HRES. 在WS10的方面, FuXi-Extreme的整体表现优于FuXi和HRES.

图4展示了HRES、FuXi和FuXi-Extreme在5天预报中基于不同百分位数计算的T2M、TP和WS10的SEDI分数. 正如预期, 随着百分位数的增加, SEDI评分下降, 表明更极端事件的预报挑战更大. 在评估的三个变量中, TP的评分最低, 而T2M的评分最高, 其SEDI值在整个5天预报时长内始终保持在0.6以上. 比较FuXi和FuXi-Extreme的SEDI评分, 在第90百分位时差异微小, 但在更高百分位数时差异显著, 表明FuXi-Extreme在更高百分位数上显著优于FuXi. FuXi在第90和95百分位数时超过了HRES, 而在第98百分位时表现与HRES相近. 然而, 在第99.5百分位数时, FuXi的表现不如HRES. 总体而言, FuXi-Extreme在所有三个变量和预报时长上都取得了最高评分, 表明其在5天预报中对极端T2M、WS10和TP的预报能力更强.

图5展示了三个地表变量在三个预报时长的能量谱, 表明模型在不同空间尺度和预报时长的能量水平, 从而展示模型的有效性. 作为基准的ERA5的能量谱不

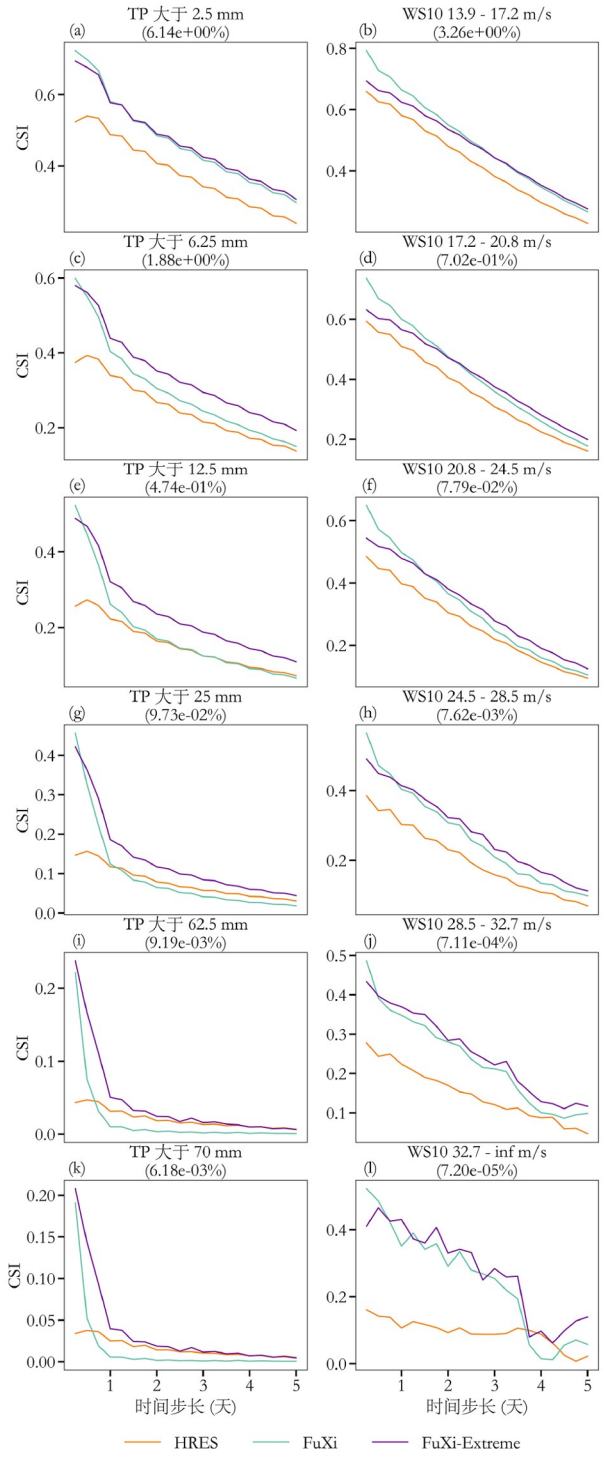


图 3　总降水量(TP)和10m风速(WS10)CSI指标对比

HRES(橙色线)、FuXi(深绿线)和FuXi-Extreme(浅紫线)在预报总降水量(TP, 左列)和10m风速(WS10, 右列)方面在预报总降水量(TP, 左列)和10m风速(WS10, 右列)方面对于各种阈值的CSI对比. 其中, 6小时总降水量阈值在图(a)、(c)、(e)、(g)、(i)和(k)中的阈值分别为2.5、6.25、12.5、25、62.5和70mm, 而10m风速在图(b)、(d)、(f)、(h)、(j)和(l)中的阈值分别为13.9、17.2、20.8、24.5、28.5和32.7m/s. 对比采用的是2018年的测试数据. 极端情况相对于整个测试集的比例在括号中标出. 所有预报数据均与ERA5再分析数据集进行了评估

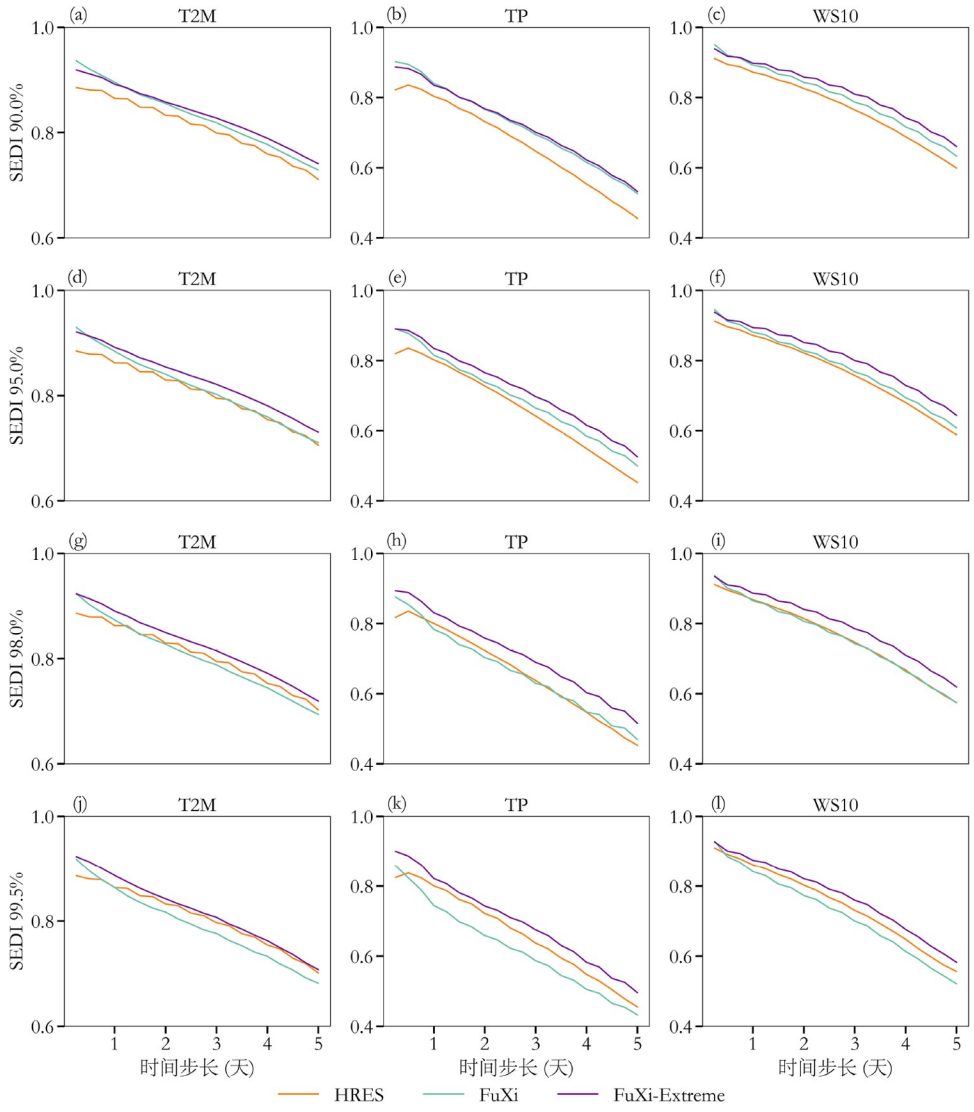


图 4　2m温度(T2M)、总降水量(TP)和10m风速(WS10)SEDI指标对比

HRES(橙色线)、FuXi(深绿线)和FuXi-Extreme(浅紫线)的2m温度(T2M, 左列)、总降水量(TP, 中列)和10m风速(WS10, 右列)在不同分位数下的SEDI指标对比. 其中, (a)~(c)、(d)~(f)、(g)~(i)和(j)~(l)分别对应的是第90百分位(第一行)、第95百分位(第二行)、第98百分位(第三行)和第99.5百分位(第四行)的SEDI, 对比采用的是2018年的测试数据. 极端情况相对于整个测试集的比例分别是10%(第一行)、5%(第二行)、2%(第三行)和0.5%(第四行). 所有预报数据均与ERA5再分析数据集进行了评估

随预报时长而改变. 与HRES预报相比, FuXi和FuXi-Extreme与ERA5数据更为一致. 然而, 随着预报时长的增加, 这两个模型在较小尺度上的表现均有所下降, 这一趋势在FuXi模型的TP预报中尤为显著. 这种在所有尺度上的下降, 突显了准确捕捉TP复杂细节的挑战. 与ML模型不同, HRES预报随着预报时长的增加保持稳定的性能, 展示了基于物理模型的固有优势.

此外, FuXi-Extreme模型显著提高了各个尺度上TP预报的一致性, 与ERA5数据的吻合度明显优于FuXi模型. 随着预报时长的增加, FuXi-Extreme相较于FuXi的优势更加明显, 表明其在改进小尺度气象要素(如TP)预报方面的有效性. 对于变量, 如WS10、FuXi和FuXi-Extreme的预报结果相似, 随着预报时间的延长差异可以忽略不计. 另一方面, FuXi-Extreme的T2M预报在较小尺度上的能量谱低于FuXi, 可能是由于FuXi-Extreme的优化方法导致的结果. 这种方法侧重

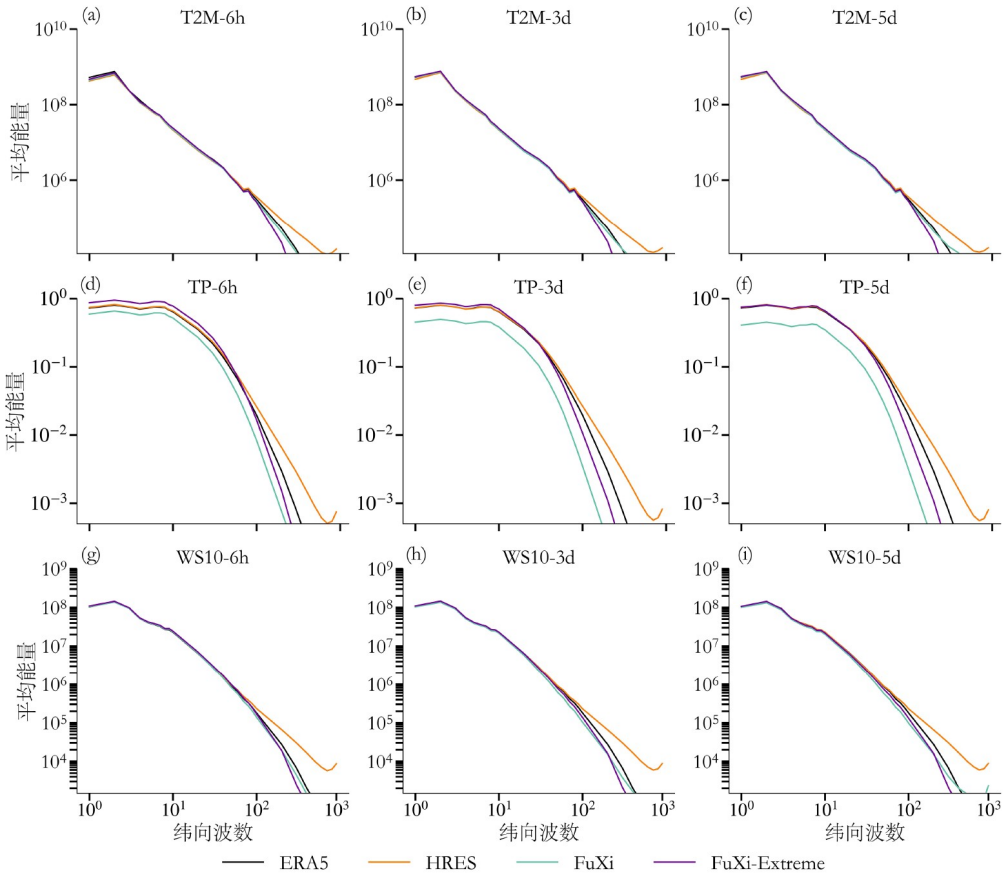


图 5　2m温度(T2M)、总降水量(TP)和10m风速(WS10)的能量谱对比

ERA5(黑线)、HRES(橙色线)、FuXi(深绿线)和FuXi-Extreme(浅紫线)的能量谱对比. 对比采用的是2018年测试数据, 在预报时长为6小时(第一列)、3天(第二列)和5天(第三列)时的2m温度(T2M, 第一行)、总降水量(TP, 第二行)和10m风速(WS10, 第三行)的能量谱

于中小尺度数据(如U10、V10和TP), 并以MSE为驱动, 会优先考虑中小尺度的信息, 但以牺牲大尺度能量分布为代价.

图6和图7展示了2018年8月15日12时UTC的初始化预报. 选择这一特定时间段是因为2018年8月15日至21日期间台风“温比亚”(RUMBIA)在中国东部和中部内陆移动时引发了创纪录的极端降雨. 这两幅图表明, 与FuXi相比, FuXi-Extreme能够生成更为细致的预报. 尤其是在60小时和120小时的预报中, 这一区别更加明显. 此外, 这两幅图还显示, HRES数据相比ERA5数据包含了更细尺度的细节, 因为它是源自空间分辨率为0.1°的原始HRES数据集.

4.2　热带气旋路径和强度预报性能

TC预报的评估包括对路径和强度预报的评估. 在本研究中, 我们分析了90个TC的总体预报技巧. 图8展示了HRES、FuXi和FuXi Extreme在5天预报中的统计比较, 涵盖了与IBTrACS数据集对比的TC路径和强度预报的评估. 在TC路径预报方面, FuXi-Extreme在0~2天的预报中的表现略差于HRES和FuXi. 然而, 在随后的预报中, FuXi和FuXi-Extreme均优于HRES, 且随着预报时长的增加, 这一优势更加明显. 值得注意的是, FuXi和FuXi-Extreme在路径误差上的差异微乎其微, 因为FuXi-Extreme本质上是FuXi的增强版, 专为近地面气象要素的极端预报优化的. 它们在高空气象变量(包括200、500、700和850hPa压力层的U和V分量以及位势高度)上具有相同的值, 这些变量对于确定TC中心至关重要. 在TC强度预报方面, FuXi-Extreme在最大WS10上比FuXi有明显改进, 这体现在更小的WS10 RMSE值上. 在MSL预报方面, FuXi-Extreme和

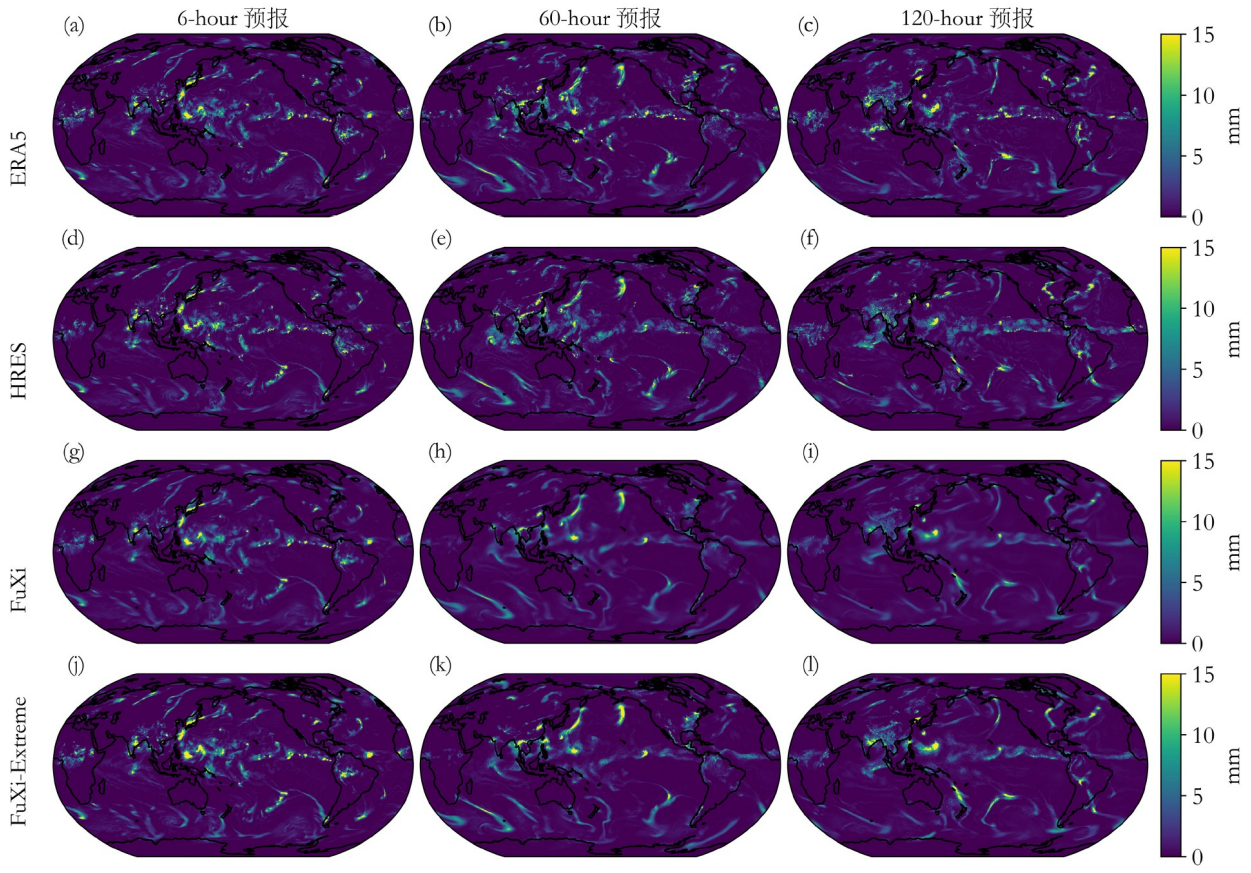


图 6　总降水量(TP)的快照示例对比

2018年8月15日12时UTC初始化的预报结果对比. 其中, 第一、二和三列分别为6、60和120小时预报中, 总降水量(TP)的快照示例比较. ERA5(真实值, 第一行)、HRES(第二行)、FuXi(第三行)和FuXi-Extreme(第四行)

FuXi之间的差异不显著. HRES在TC强度预报中表现最佳, 在整个预报期间始终保持最低的RMSE值. 总的来说, FuXi和FuXi-Extreme在TC路径预报上优于HRES, 而HRES在TC强度预报中超过FuXi和FuXi-Extreme.

同样, 如图8所示, 图9显示了使用ERA5数据集评估的TC预报比较. TC路径预报的比较结果与之前的结论一致: FuXi和FuXi-Extreme的表现优于HRES. 在TC强度预报方面, FuXi和FuXi-Extreme的RMSE值小于HRES.

这可能归因于以下因素: (1) 从ERA5数据集中提取的TC强度数据相比于IBTrACS数据集具有显著较小的TC强度, ERA5数据集中的MSL值较高而WS10值较低; (2) FuXi和FuXi-Extreme使用ERA5数据集进行训练的, 该数据集相比HRES更为平滑(见图6和图7). TC强度预报比较中的矛盾发现表明, 需要转向更高分辨率的数据集作为训练基于ML的天气预报模型的真值, 从而提升预报性能(Zhong等, 2024).

5　结论与未来工作

尽管基于ML的天气预报模型取得了显著成就, 但随着预报时长的增加, 它们往往会生成越来越平滑的预报. 这种平滑的预报常常导致ML模型对极端天气事件的低估. 在本研究中, 我们通过训练DDPM开发了FuXi-Extreme, 以增强FuXi模型生成的地面气象要素预报的细节. 基于CSI和SEDI指标的评估表明, FuXi-Extreme在预报TP、WS10和T2M的极值方面优于FuXi和HRES. 能量谱分析表明, FuXi-Extreme在捕捉TP预报中的精细结构方面更为有效. 此外, TC路径

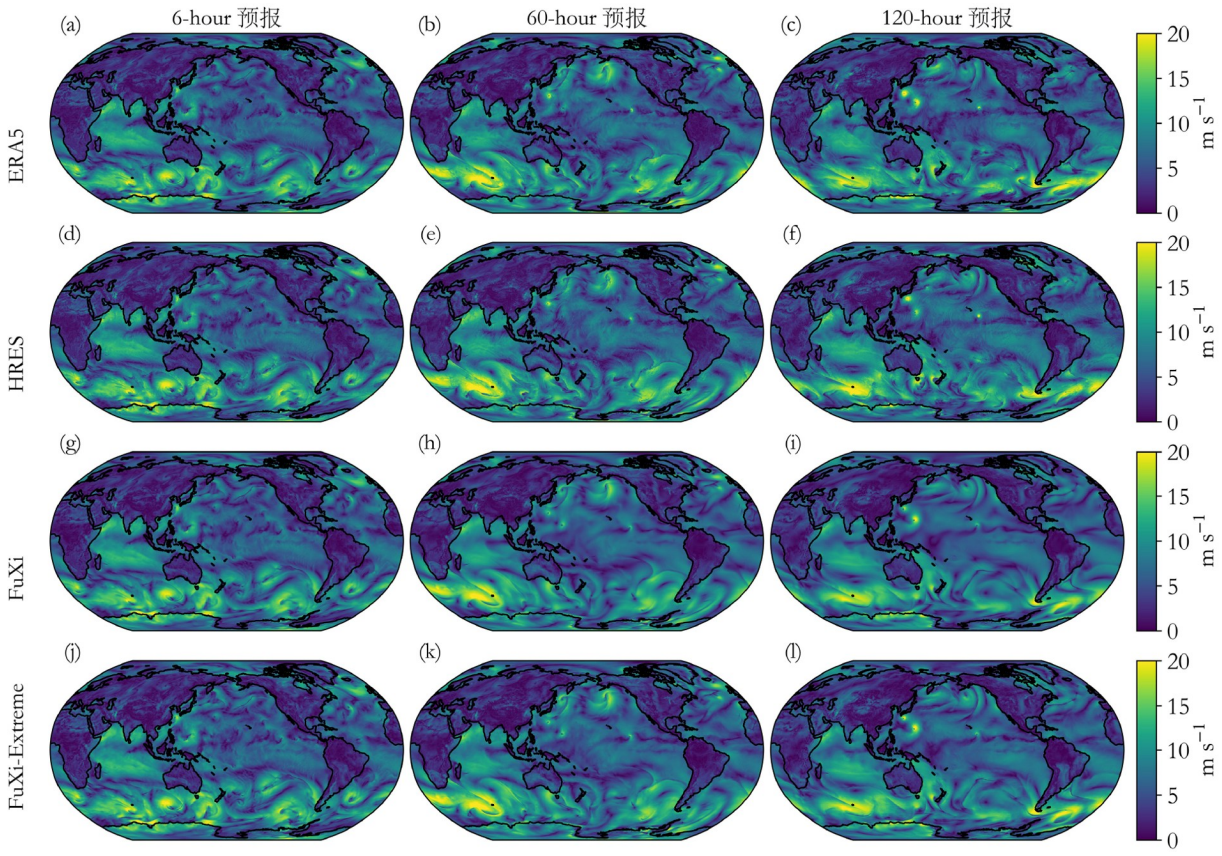


图 7　10m风速(WS10)的快照示例比较

2018年8月15日12时UTC初始化的6小时(第一列)、60小时(第二列)和120小时(第三列)预报中，10m风速(WS10)的快照示例比较. ERA5(真实值, 第一行)、HRES(第二行)、FuXi(第三行)和FuXi-Extreme(第四行)

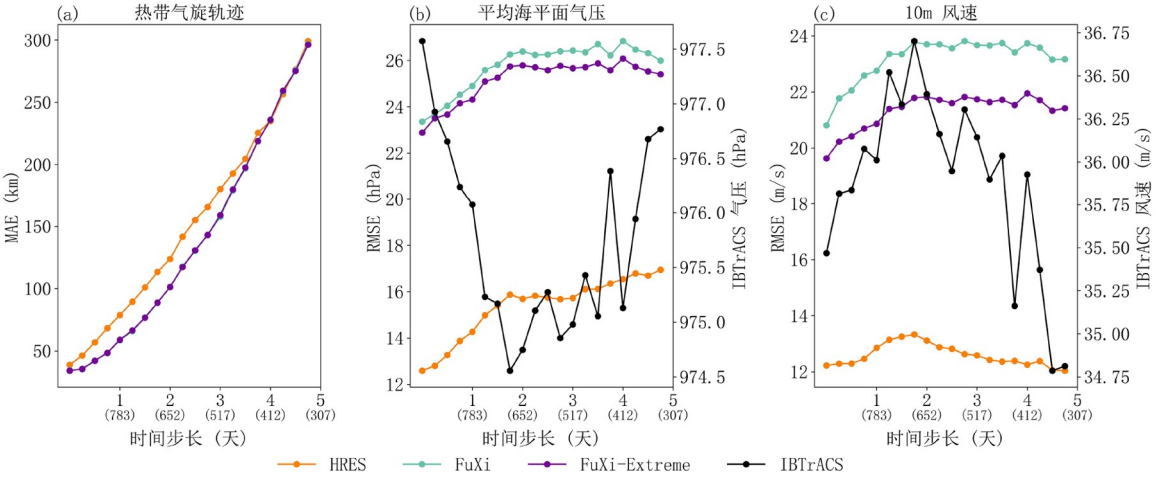


图 8　以IBTrACS数据为真值的热带气旋(TC)的路径和误差强度对比

三个模型的TC路径预报的平均绝对误差(MAE，第一列)和TC强度预报中MSL(第二列)和WS10(第三列)的均方根误差(RMSE)比较，包括HRES(橙色线)、FuXi(深绿线)和FuXi-Extreme(浅紫线)，作为预报时长的函数. 评估覆盖了90个TC的预报，并与IBTrACS数据集进行比较. MSL和WS10预报比较为双*Y*轴图, 次级*Y*轴显示IBTrACS数据(黑线)

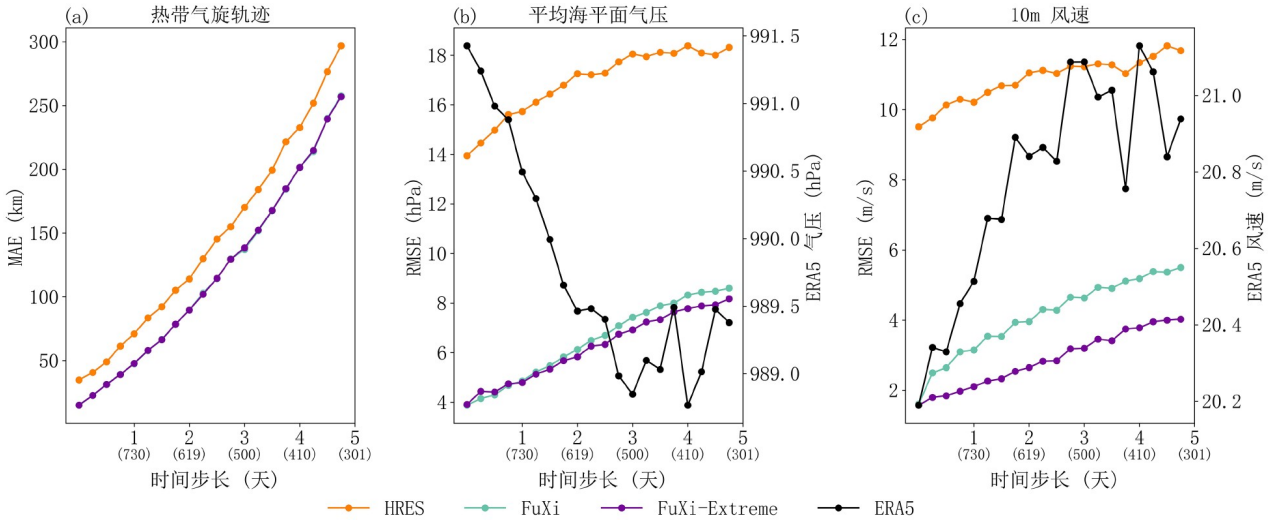


图 9　以ERA5数据为真值的热带气旋(TC)的路径和误差强度对比

三个模型的TC路径预报的平均绝对误差(MAE, 第一列)和TC强度预报中MSL(第二列)和WS10(第三列)的均方根误差(RMSE)比较. 包括HRES(橙色线)、FuXi(深绿线)和FuXi-Extreme(浅紫线), 作为预报时长的函数. 评估覆盖了90个热带气旋的预报, 并与ERA5再分析数据集进行比较. MSL和WS10预报比较为双*Y*轴图, 次级*Y*轴显示ERA5数据(黑线)

和强度预报的评估显示, FuXi和FuXi-Extreme在预报TC路径方面优于HRES. 然而, 使用IBTrACS数据集评估时, 它们在TC强度预报的准确性低于HRES, 而使用ERA5数据集评估时, 它们优于HRES.

尽管FuXi-Extreme在极端事件预报方面显示出良好的改进效果, 但它目前仅关注近地面变量. 未来, 我们将致力于在完成FuXi V2.0的开发后, 增强对高空气象变量极值的预报能力. FuXi V2.0, 作为FuXi的第二版, 将涵盖更多的气象要素. 另一个需要改进的方面是风速预报. 目前, 风速是从U和V风分量计算得出的, 并未在模型训练过程中直接优化. 在FuXi V2.0中, 我们将直接建模并优化风速. 此外, 视觉比较显示, 作为训练ML天气预报模型的真值常用的ERA5数据集相比HRES更为平滑. 因此, 我们需要使用分辨率更高、细节更多的数据集作为真值, 从而提升ML天气预报模型的性能.

此外, 包括FuXi和FuXi-Extreme在内的ML天气预报模型完全依赖数据驱动, 缺乏对所预报的物理系统的先验知识. 因此, 它们经常被称为"黑箱". 虽然这些ML模型能够生成较为准确的预报结果, 但它们在预报过程中的透明度不足, 削弱了人们对其可靠性的信心. 因此, 解释这些ML模型并验证其推理是否符合对天气系统的物理理解是至关重要的. 可解释机器学习(Explainable ML, XML)(McGovern等, 2019; Mamalakis等, 2020; Molnar等, 2020; Toms等, 2021)方法的出现为大气研究开辟了新的机遇. 在未来的工作中, 我们计划利用XML方法, 例如逐层相关传播(Layer-Wise Relevance Propagation, LPR)(Bach等, 2015), 深入了解哪些气象要素和输入对预报结果有较大影响, 以及为什么FuXi-Extreme的预报在极端事件方面相比FuXi有所改进.

致谢　ERA5数据集的一个子集从Copernicus Climate Data(CDS)官网(https://cds.climate.copernicus.eu/)下载. ECMWF HRES TC路径从TIGGE档案中以可下载的XML文件形式获取(https://confluence.ecmwf.int/display/TIGGE/Tools). TC真实路径数据以及2018年在FuXi和FuXi-Extreme数据中检测到的所有TC数据来源于国际热带气旋最佳路径资料集(IBTrACS) (https://www.ncei.noaa.gov/products/international-best-track-archive). 本研究中用于训练和运行FuXi模型的源代码来源于https://doi.org/10.5281/zenodo.8100201(Chen等, 2023c). DDPM型号可在lu2022dpmsolver上购买. 感谢ECMWF的研究人员提供ERA5再分析数据和HRES. 感谢美国国家海洋和大气管理局国家环境信息中心(NOAA)提供IBTrACS数据. 感谢复旦大学未来计算(Computing for the Future at Fudan, CFFF)提供高性能计算平台.

参考文献

Bach S, Binder A, Montavon G, Klauschen F, Müller K R, Samek W. 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One, 619: e0130140

Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q. 2023. Accurate medium-range global weather forecasting with 3D neural networks. Nature, 619: 533–538

Bougeault P, Toth Z, Bishop C, Brown B, Burridge D, Chen D H, Ebert B, Fuentes M, Hamill T M, Mylne K, Nicolau J, Paccagnella T, Park Y Y, Parsons D, Raoult B, Schuster D, Dias P S, Swinbank R, Takeuchi Y, Tennant W, Wilson L, Worley S. 2010. The THORPEX interactive grand global ensemble. Bull Amer Meteor Soc, 91: 1059–1072

Chen L, Du F, Hu Y, Wang Z, Wang F. 2023a. Swinrdm: Integrate swinrnn with diffusion model towards high-resolution and high-quality weather forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. 322–330

Chen L, Zhong X, Zhang F, Cheng Y, Xu Y, Qi Y, Li H. 2023b. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. NPJ Clim Atmos Sci, 6: 190

Chen L, Zhong X, Zhang F, Cheng Y, Xu Y, Qi Y, Li H. 2023c. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast (Version 1.0) [Dataset][Software]. NPJ Clim Atmos Sci, 6: 1–11

Croitoru F A, Hondru V, Ionescu R T, Shah M. 2023. Diffusion models in vision: A survey. IEEE Trans Pattern Anal Mach Intell, 45: 10850–10869

Dhariwal P, Nichol A. 2021. Diffusion models beat gans on image synthesis. Adv Neural Inf Process Syst, 34: 8780–8794

Ebi K L, Vanos J, Baldwin J W, Bell J E, Hondula D M, Errett N A, Hayes K, Reid C E, Saha S, Spector J, Berry P. 2021. Extreme weather and climate change: Population health and health system implications. Annu Rev Public Health, 42: 293–315

Ferro C A T, Stephenson D B. 2011. Extremal dependence indices: Improved Verification measures for deterministic forecasts of rare binary events. Weather Forecast, 26: 699–713

der Grijn V. 2002. Tropical cyclone forecasting at ECMWF: New products and validation. ECMWF Tech Memo, 386: 1

Haiden T, Janousek M, Vitart F, Ben-Bouallegue Z, Ferranti L, Prates F, Richardson D. 2021. Evaluation of ECMWF forecasts, including the 2021 upgrade, ECMWF technical memoranda 884. Reading, UK Eur Cent Mediu, Range Weather Forecast

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J. 2020. The ERA5 global reanalysis, quarterly journal of the royal meteorological society

Hess P, Boers N. 2022. Deep learning for improving numerical weather prediction of heavy rainfall. J Adv Model Earth Syst, 14: e2021MS002765

Ho J, Jain A, Abbeel P. 2020. Denoising diffusion probabilistic models. Adv Neural Inf Process Syst, 33: 6840–6851

Ho J, Saharia C, Chan W, Fleet D J, Norouzi M, Salimans T. 2022. Cascaded diffusion models for high fidelity image generation. J Mach Learn Res, 23: 1–33

Kingma D P, Welling M. 2013. Auto-encoding variational bayes. arXiv Prepr. arXiv1312.6114

Kingma D P, Ba J. 2014. Adam: A method for stochastic optimization. arXiv Prepr. arXiv1412.6980

Knapp K R, Kruk M C, Levinson D H, Diamond H J, Neumann C J. 2010. The international best track archive for climate stewardship (IBTrACS). Bull Amer Meteor Soc, 91: 363–376

Knapp K R, Diamond H J, Kossin J P, Kruk M C, Schreck C J. 2018. International best track archive for climate stewardship (IBTrACS) project, version 4. NOAA National Centers for Environmental Information. doi. org/10.25921/82ty-9e16 (accessed 13 Febr. 2020)

Kotz M, Levermann A, Wenz L. 2022. The effect of rainfall changes on economic production. Nature, 601: 223–227

Lam R, Sanchez-Gonzalez A, Willson M, Wirnsberger P, Fortunato M, Alet F, Ravuri S, Ewalds T, Eaton-Rosen Z, Hu W, Merose A, Hoyer S, Holland G, Vinyals O, Stott J, Pritzel A, Mohamed S, Battaglia P. 2023. Learning skillful medium-range global weather forecasting. Science, 382: 1416–1421

Larsson G, Maire M, Shakhnarovich G. 2016. Fractalnet: Ultra-deep neural networks without residuals. arXiv Prepr. arXiv1605.07648

Liu L, He G, Wu M, Liu G, Zhang H, Chen Y, Shen J, Li S. 2023. Climate change impacts on planned supply–demand match in global wind and solar energy systems. Nat Energy, 8: 870–880

Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, Ning J, Cao Y, Zhang Z, Dong L. 2022. Swin transformer v2: Scaling up capacity and resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 12009–12019

Loshchilov I, Hutter F. 2017. Decoupled weight decay regularization. arXiv Prepr. arXiv1711.05101

Magnusson L, Haiden T, Richardson D. 2014. Verification of extreme weather events: Discrete predictands. European Centre for Medium-Range Weather Forecasts Reading, UK

Magnusson L, Majumdar S, Emerton R, Richardson D, Alonso-Balmaseda M, Baugh C, Bechtold P, Bidlot J, Bonanni A, Bonavita M. 2021. Tropical cyclone activities at ECMWF

Mamalakis A, Ebert-Uphoff I, Barnes E A. 2020. Explainable artificial intelligence in meteorology and climate science: Model fine-tuning, calibrating trust and learning new science. In: International Work-

shop on Extending Explainable AI Beyond Deep Models and Classifiers. Springer. 315–339

McGovern A, Lagerquist R, John Gagne D, Jergensen G E, Elmore K L, Homeyer C R, Smith T. 2019. Making the black box more transparent: Understanding the physical implications of machine learning. Bull Am Meteorol Soc, 100: 2175–2199

Molnar C, Casalicchio G, Bischl B. 2020. Interpretable machine learning—A brief history, state-of-the-art and challenges. In: Joint European conference on machine learning and knowledge discovery in databases. Springer. 417–431

Nichol A, Dhariwal P, Ramesh A, Shyam P, Mishkin P, McGrew B, Sutskever I, Chen M. 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv Prepr. arXiv2112.10741

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A. 2017. Automatic differentiation in pytorch

Pathak J, Subramanian S, Harrington P, Raja S, Chattopadhyay A, Mardani M, Kurth T, Hall D, Li Z, Azizzadenesheli K. 2022. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv Prepr. arXiv2202.11214

Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. 2022. Hierarchical text-conditional image generation with clip latents. arXiv Prepr. arXiv2204.06125, 1: 3

Rasp S, Hoyer S, Merose A, Langmore I, Battaglia P, Russell T, Sanchez-Gonzalez A, Yang V, Carver R, Agrawal S, Chantry M, Ben Bouallegue Z, Dueben P, Bromberg C, Sisk J, Barrington L, Bell A, Sha F. 2024. WeatherBench 2: A benchmark for the next generation of data-driven global weather models. J Adv Model Earth Syst, 16: e2023MS004019

Saharia C, Ho J, Chan W, Salimans T, Fleet D J, Norouzi M. 2022. Image super-resolution via iterative refinement. IEEE Trans Pattern Anal Mach Intell, 45: 4713–4726

Schaefer J T. 1990. The critical success index as an indicator of warning skill. Wea Forecast, 5: 570–575

Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning. PMLR. 2256–2265

Stephenson D B, Casati B, Ferro C A T, Wilson C A. 2008. The extreme dependency score: A non-vanishing measure for forecasts of rare events. Meteorol Appl, 15: 41–50

Swinbank R, Kyouda M, Buchanan P, Froude L, Hamill T M, Hewson T D, Keller J H, Matsueda M, Methven J, Pappenberger F, Scheuerer M, Titley H A, Wilson L, Yamaguchi M. 2016. The TIGGE project and its achievements. Bull Am Meteorol Soc, 97: 49–67

T. Parmesan C, Morecroft M D & Y. 2022. Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the IPCC Sixth Assessment Report of the Intergovernmental Panel on Climate Change

Toms B A, Barnes E A, Hurrell J W. 2021. Assessing decadal predictability in an earth-system model using explainable neural networks. Geophys Res Lett, 48: e2021GL093842

Whang J, Delbracio M, Talebi H, Saharia C, Dimakis A G, Milanfar P. 2022. Deblurring via stochastic refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 16293–16303

Wilks D S. 2011. Statistical Methods in the Atmospheric Sciences. Academic Press

Zhong X, Du F, Chen L, Wang Z, Li H. 2024. Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts. Quart J R Meteoro Soc, 150: 275–289

(责任编委: 雷荔傈)