

文章编号: 2096-1618(2023)03-0264-07

基于 CMA-REPS 格点预报数据的深度学习风速订正方法

毛波¹, 杨昊^{1,2}, 周世杰², 杨康权^{3,4}, 陈敏¹

(1. 成都信息工程大学计算机学院, 四川 成都 610225; 2. 电子科技大学信息与软件工程学院, 四川 成都 610054; 3. 四川省气象台, 四川 成都 610072; 4. 四川省气象局重点实验室, 四川 成都 610072)

摘要: 准确的风速预测对风能资源的充分利用和风电场的经济效益提升具有显著的意义。为提高集合数值预报的风速预报能力, 弥补现有深度学习集合预报订正模型对格点预报数据时间特征提取的不足, 引入 ConvLSTM 深度学习模型, 对 CMA-REPS(中国气象局区域集合预报模式) 预测的华北地区近地面 10 m 风速格点数据进行偏差订正实验, 以均方根误差 (RMSE) 作为评分标准将订正结果与 CMA-REPS 原始预报数据和 Unet 深度学习模型方法得到的订正结果进行对比。结果表明, ConvLSTM 模型的订正效果相比 Unet 模型有进一步的提升, 经 ConvLSTM 模型订正后的近地面 10 m 风速预报数据整体上更趋近于实况数据。

关键词: CMA-REPS; 集合预报; 偏差订正; 深度学习; 风速

中图分类号: TP183

文献标志码: A

doi: 10.16836/j.cnki.jcui.2023.03.003

0 引言

风能作为一种高效清洁的新能源, 其巨大的发展潜力引起了国内外广泛的重视。实现准确的风速预测能够有效地提高风能资源的利用率, 降低风功率波动对电网稳定性的影响, 从而提升风电场的经济效益^[1]。集合数值预报模式由于提供了关于天气预报的不确定性信息, 相比于单一的确定性数值预报模式, 往往能够得到更准确的风速预报结果, 现已在全球多个国家的天气业务预报中心实现常规运行^[2]。但在实际应用中, 由于集合预报模式设计之初的各种不完善以及大气系统的不确定性, 集合预报模式一直存在系统性误差, 从而造成风速预报的结果出现系统性偏差。为了满足更加精确和可靠的风速预报需求, 需要对集合数值预报模式的结果进行偏差订正, 以提升模式对风速的预报能力。

目前, 受限于风场本身所具有的变异性特征, 国内对风场集合预报结果偏差订正的方法还是多以传统的统计订正方法为主^[3], 如集合模式输出统计法^[4]和贝叶斯模型平均法^[5]。这类订正方法均通过建立线性统计模型来订正风场集合预报的误差, 对于风场变化的随机特征, 传统的线性统计模型往往不能很好地捕捉其中的非线性关系。随着近年来计算机软硬件设备的不断提升以及人工智能技术的蓬勃发展, 国内外气象领域的研究人员开始逐步结合机器学习方法或深度学习

方法对数值预报模式结果进行偏差订正, 实现风速的精细化预报。Lahouar A 等^[6]使用随机森林的机器学习方法, 建立了提前一小时的风速预测模型。与其他大多数机器学习方法相比, 随机森林方法不需要进行调整或优化, 结果表明, 使用该模型的风速预报精度有显著的提高。孙全德等^[7]使用常见的几种机器学习算法 (LASSO 回归、随机森林等) 对数值天气预报模式 ECMWF 预测的华北地区近地面 10 m 风速进行订正, 将订正结果与传统订正方法得到的订正结果进行对比。结果表明, 基于机器学习算法的风速订正方法效果均好于传统订正方法, 显示了机器学习方法提升数值预报风速预报精度的潜力。但上述机器学习方法仅限于对确定性数值预报的数据进行偏差订正, 尚未对集合数值预报模式进行尝试。且机器学习较依赖于复杂的特征工程, 将机器学习应用于模式的偏差订正时, 需要研究人员具备专业的气象知识和大量的时间, 使训练过程成本偏高且十分具有挑战性。

深度学习作为机器学习的一种特定形式, 由于不需要进行特征工程, 因此在对海量数据进行非线性建模时具有更大的优势。过去十年中, 深度学习已在计算机视觉、语音识别及自然语言处理等多个领域取得了丰富的成果, 为人类社会的进步和发展带来了极大的帮助。对于大气系统所具有的特殊的高度非线性特征, 深度学习现已成为人工智能领域与大气科学领域交叉应用的研究热点和主流发展方向。Stephan Rasp 等^[8]率先使用深度神经网络在分布回归框架中对集合预报预测的 2 m 温度进行偏差订正实验, 将连续分级概率评分 (continuous ranked probability score, CRPS)

收稿日期: 2022-10-21

基金项目: 国家重点研发计划资助项目 (2021YFC3000902); 四川省科技计划重点研发专项资助项目 (2022YFS0542)

作为其模型的损失函数。该实验的局限在于只考虑单一站点位置以及单一天气变量的偏差订正,对目前的数值预报模式生成的二维网格数据不具有适用性。对于二维网格数据所具有的空间特性,需考虑使用诸如卷积神经网络^[9]等方法来提取其空间特征。受此启发,Grönquist Peter等^[10]基于Unet模型搭建了集合预报的偏差订正模型,结果表明其模型在CRPS评分方面相比于传统订正方法取得了超过14%的相对改进。L Han等^[11]基于Unet提出了一个CU-net(Correction U-net)模型,将格点预报数据的偏差订正问题转化为深度学习中的图像识别问题,对ECMWF-IFS模式的格点预报数据进行订正,也取得了不错的订正效果。张延彪等^[12]为进一步加强CU-net模型对复杂的气象格点数据的偏差订正能力,引入稠密卷积模块^[13](Denseblock)对CU-net进行改进,构建了Dense-CU-net模型。上述研究虽然都对格点预报数据的偏差订正有较好的效果,但考虑到格点预报数据的空间特征,仍存在一些不足,主要体现在没有考虑模型对格点预报数据的时间维度特征的提取融合。

集合数值预报模式预报的格点数据具有复杂的空间特征和时间特征,属于典型的时空序列数据。对这类数据进行偏差订正的深度学习方法,本文引入时空序列预测问题常用到的ConvLSTM模型^[14]对CMA-REPS预报的近地面(10 m)风速格点数据进行偏差订正,订正的预报时间为24 h、48 h和72 h。将订正结果与CMA-REPS原始预报数据和Unet模型得到的结果进行对比,结果表明该模型方法能进一步提升集合数值预报模式风场格点数据的预报精度。

1 模式与方法

1.1 CMA-REPS

CMA-REPS是中国气象局自主研发的新一代区域集合预报系统,其前身是GRAPES-REPS(global and regional assimilation and prediction enhanced system-regional ensemble prediction system)区域集合预报系统,基于中国自主研究并建立的新一代多尺度通用资料同化与数值天气预报系统-GRAPES的区域模式发展而来^[15-16]。集合预报扰动方法对集合预报的可靠性提升极为重要,当前CMA-REPS初值扰动采用6 h循环计算方案的GRAPES模式面集合变换卡尔曼滤波^[17];模式扰动采用单一物理过程参数化方案与随机物理过程倾向项(stochastic perturbed parameterization tendencies, SPPT)^[18]组合;边界扰动通过提取GRAPES-

GEPS(global and regional assimilation and prediction enhanced system-global ensemble prediction system)全球集合预报系统扰动成员相比于自身控制预报的扰动测边界,并叠加在CMA-REPS的控制预报测边界上得到^[19]。CMA-REPS参数配置如表1所示,其水平分辨率为0.1°,垂直分辨率为50层,集合预报成员数15个,预报区域为中国区域,预报时效84 h(00时,12时(协调世界时)),6 h(06时,18时(协调世界时))。CMA-REPS中的控制预报初值和侧边界来源于NCEP-GFS(national centers for environmental prediction-global forecast system)全球模式的预报场;并且CMA-REPS加入了云分析同化技术与条件性台风涡旋重定位技术^[20],以期提高短临降水和台风预报能力。CMA-REPS具体参数配置见表1。

表1 CMA-REPS具体参数配置

参数名称	参数配置
控制预报	CMA-MESO
分辨率	0.1°/L50
控制预报初值和侧边界	NCEP-GFS
同化分析技术	云分析
初值不确定性	ETKF(6 h循环)
模式不确定性	单一物理+SPPT
边界不确定性	CMA-GEPS
台风不确定性	条件性台风涡旋重定位技术
集合成员数	15
预报区域	15°N~65°N,70°E~140°E(2022年6月7日之前),10°N~60°N,70°E~145°E(2022年6月7日之后)
预报时效	84 h(0000、1200UTC),6 h(0600、1800UTC)
模式输出间隔	1 h

1.2 基于ConvLSTM的风速订正模型

集合数值预报模式的格点预报数据是在时间上连续分布的空间数据^[21],具有典型的时空特征。仅靠单一的基于CNN的深度学习模型如Unet,虽然可以有效地提取格点预报数据复杂的空间分布特征,但无法捕捉数据的时间序列信息。随着深度学习的发展,LSTM(long short-term memory)对于气象要素站点数据的序列预测能力得到了验证。因此,考虑将时间成分纳入深度学习模型(例如通过使用LSTM),可以进一步提高深度学习模型对集合数值预报模式格点预报数据的偏差订正能力。但是,传统的长短期记忆单元无法实现对空间特征的提取。基于此,引入ConvLSTM模型,

尝试对 CMA-REPS 预报的近地面 (10 m) 风速格点数据进行偏差订正。ConvLSTM 模型是 Shi 等^[14]为解决时空序列预测问题,结合卷积神经网络对数据空间特征的提取与编码能力及长短期记忆网络的序列特征编码特性而提出的。

ConvLSTM 的单元结构图如图 1 所示。由图 1 可以看到,ConvLSTM 是 LSTM 结构的一种变体,通过遗忘门(f_t)、输入门(i_t)和输出门(o_t)控制数据在细胞内部的交流。各个门控制数据参与到细胞状态的更新,通过门来选择性地保留或舍弃信息。ConvLSTM 与 LSTM 的区别在于 ConvLSTM 在“输入到状态”和“状态到状态”两部分都采用卷积运算来代替矩阵乘法,并且其所有的输入 X_1, \dots, X_t , 细胞输出 C_1, \dots, C_t , 隐藏状态 H_1, \dots, H_t , 和遗忘门(f_t)、输入门(i_t)和输出门(o_t)都是 3 维张量,这样做的好处是在提取序列的空间信息同时,可以去除大量空间冗余特征并且解决数据的时间依赖问题。

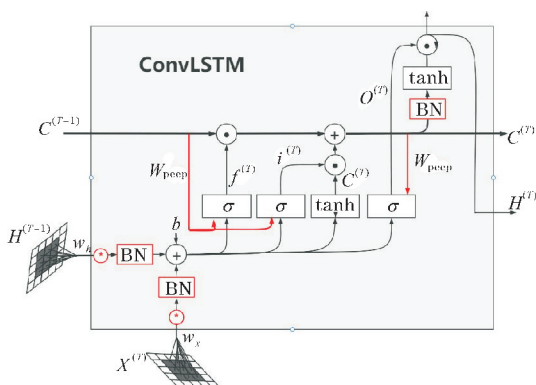


图1 ConvLSTM 单元结构图

ConvLSTM 门之间的传递关系如下:

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\ H_t &= o_t \circ \tanh(C_t) \\ \sigma(x) &= \frac{1}{1 + e^{-x}} \end{aligned}$$

式中 i_t 为输入门, f_t 为遗忘门, C_t 为细胞状态, o_t 为输出, H_t 为隐藏层输出, $*$ 表示卷积运算符, \circ 表示 Hadamard 乘积, σ 为 sigmoid 激活函数。遗忘门可以控制信息的遗忘, 丢弃被认为冗余的信息, 保留有用的信息并将其向后传递; 继续传递的信息进到输入门, 通过 sigmoid 层确定需要更新的信息, 并通过 tanh 层得到新的细胞信息对细胞进行更新。最后通过输出门中的 sigmoid 的信息乘以通过 tanh 的记忆细胞的信息得到模型的最终输出。

对于本文的集合数值预报模式风场格点预报数据的偏差订正问题, 通过堆叠多个 ConvLSTM 网络层搭建偏差订正模型, 模型结构如图 2 所示。模型输入由 Y_t 和 $P_{t+\Delta t}$ 组成, 其中 Y_t 是 $2t, 10u, 10v$ 和 $10w$ 这 4 个要素在 t 时刻的实况数据, $P_{t+\Delta t}$ 是 $2t, 10u, 10v$ 及 $10w$ 在 t 时刻的 Δ_t 小时集合平均预报数据, Δ_t 可取的值有 24, 48 和 72。将实况数据 Y_t 与预报数据 $P_{t+\Delta t}$ 拼接在一起, 得到的模型输入数据大小为 $101 \times 101 \times 8$ 。输入数据经过最小最大值归一化后进入两个 ConvLSTM2D 网络层, 最后经过一层全连接层输出偏差订正后的 CMA-REPS 的 Δ_t 预报数据。两层 ConvLSTM-2D 网络层卷积核数目均为 10, 考虑数据本身的特征分布和网格分辨率大小, 选用 3×3 大小的卷积核以避免丢失局部特征, 并且卷积过程中使用填充设置使中间过程生成的特征图大小一致。输出层 (全连接层) 的神经元个数为 1, 其激活函数为 ReLu, 用以提高模型的非线性表达能力。在每个 ConvLSTM-2D 网络层后加入批标准化层, 通过对每层的输出进行批标准化处理, 防止模型出现训练过拟合的同时加速模型训练的收敛过程。

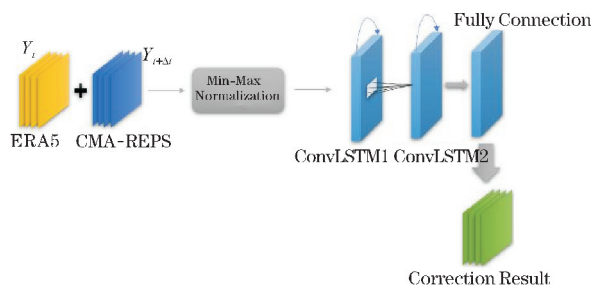


图2 基于 ConvLSTM 的风速订正模型结构图

在模型训练过程中, 使用 Adam^[22] 作为模型的优化器, 其学习率设为 0.001, 一阶矩估计的指数衰减率设为 0.9, 二阶矩估计的指数衰减率设为 0.999。Adam 算法相比于传统的随机梯度下降算法, 可以通过计算梯度的阶矩为不同的参数设计独立的自适应性学习率, 并且拥有更高效的计算效率。训练的每一批数据大小 batchsize 设为 10, 并选取均方误差 (MSE) 作为模型的损失函数:

$$MSE = \frac{1}{n} \sum_{i=0}^n (y_i - \tilde{y}_i)^2$$

式中 y_i 和 \tilde{y}_i 分别代表训练集中的观测值和经模型训练后的订正值, n 代表训练集样本数量。

2 数据

2.1 数据选取

选取 CMA-REPS 预报的格点数据作为预报数据,

采用第五代 ECMWF 全球再分析产品 (ERA5) 数据作为格点“实况”数据。其中, CMA-REPS 的格点预报数据的空间分辨率为 $0.1^{\circ} \times 0.1^{\circ}$, 网格数为 751×501 , 空间范围为 $10^{\circ} \text{N} \sim 60^{\circ} \text{N}$ 、 $70^{\circ} \text{E} \sim 145^{\circ} \text{E}$ (2022 年 6 月 7 日之前的空间范围为 $15^{\circ} \text{N} \sim 65^{\circ} \text{N}$ 、 $70^{\circ} \text{E} \sim 145^{\circ} \text{E}$), 起报时间为 00UTC 和 12UTC, 预报时效为 0–84 h (逐小时), 成员数为 15。ERA5 数据的空间分辨率为 $0.1^{\circ} \times 0.1^{\circ}$, 网格数为 101×101 , 空间范围为 $32^{\circ} \text{N} \sim 42^{\circ} \text{N}$ 、 $110^{\circ} \text{E} \sim 120^{\circ} \text{E}$ 。对 CMA-REPS 预报的华北地区近地面 10 m 风速格点数据进行偏差订正, 因此需要将 CMA-REPS 的格点预报数据裁剪成网格数为 101×101 , 空间范围为华北地区 ($32^{\circ} \text{N} \sim 42^{\circ} \text{N}$ 、 $110^{\circ} \text{E} \sim 120^{\circ} \text{E}$), 并选取预报时效为 24 h、48 h 以及 72 h 的 2 m 气温 ($2t$)、10 m 风速 ($10u$: 东西风; $10v$: 南北风) 的预报数据。两套数据的时间范围均为 2019 年 10 月 20 日至 2022 年 5 月 1 日。

2.2 数据预处理及划分

首先将 10 m 纬向风分量 (u) 和 10 m 经向风分量 (v) 合成 10 m 风速 (w), 合成公式:

$$w = \sqrt{u^2 + v^2}$$

现在选取的 CMA-REPS 预报数据和 ERA5 数据包含的气象要素有 $2t$ 、 $10u$ 、 $10v$ 和 $10w$, 然后对两套数据进行质量控制, 检测数据中是否存在缺失值或异常值。经检测, ERA5 4 个要素的数据均存在缺失值。这是由于 ECMWF 数值模式预报资料数量大, 存储空间占比高, 导致资料在存储过程中出现数据缺失。针对格点预报数据空间上存在连续性的特点, ERA5 缺失值采用相邻非缺失值线性插值方法进行补缺。

为了防止模型训练出现过拟合的现象, 提升模型的泛化能力, 还需要将所选数据分为训练集、验证集和测试集。由于实验订正的预报时间为 24 h、48 h 和 72 h, 因此将数据分为 3 组, 每组选取 2019 年 10 月 20 日至 2021 年 4 月 30 日的对应预报时效的数据为训练集, 2021 年 5 月 1 日至 2021 年 10 月 31 日对应预报时效的数据为验证集, 2021 年 11 月 1 日至 2022 年 5 月 1 日对应预报时效的数据为测试集。在单个训练数据集中, 以订正 24 h 预报数据为例, 模型输入数据为起报时间 t 的 ERA5 实况格点数据和 CMA-REPS 的 24 h 预报数据, 训练标签为 $t+24$ h 的 ERA5 实况格点数据。

3 实验和结果

3.1 实验评价指标

采用 ConvLSTM 深度学习模型对 CMA-REPS 模式

预报的华北地区近地面 10 m 风场数据 ($10u$ 、 $10v$ 和 $10w$) 进行偏差订正, 订正的预报时效为 24 h、48 h 和 72 h, 并将订正结果分别与 CMA-REPS 原始预报结果、EMOS 方法订正结果和 Unet 模型订正结果进行比较, 同时结合 ERA5 再分析场对模型的订正能力进行客观检验。对于模型订正性能的衡量, 使用均方根误差 (RMSE) 来评估模型的订正效果。RMSE 的计算公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2}$$

式中 n 代表测试集样本数量, y_i 和 \hat{y}_i 分别代表测试集的观测值和模型订正值。

3.2 10u 订正结果

图 3 显示了 CMA-REPS、Unet 模型和 ConvLSTM 模型对测试集 24 h、48 h 及 72 h 预报的 $10u$ 数据进行偏差订正的 RMSE 对比结果。由图 3 可以看出, 使用 ConvLSTM 模型订正的预报结果相比 CMA-REPS 原始预报数据和 Unet 模型的订正结果要更贴近于实况数据。尤其对 24 h 预报的 $10u$ 数据的订正效果最好, ConvLSTM 模型的 RMSE 比 CMA-REPS 降低 36.57%, 比 Unet 模型降低 15.73%。订正效果其次的是对 72 h 预报的 $10u$ 数据的订正, RMSE 比 CMA-REPS 和 Unet 模型分别降低了 14.83% 和 4.91%。对于 48 h 预报的 $10u$ 数据的订正, ConvLSTM 模型相比于 Unet 模型提升效果微乎其微, RMSE 仅降低 0.54%, 但相比 CMA-REPS 原始预报数据, RMSE 降低了 6.46%, 依旧取得不错的偏差订正效果。

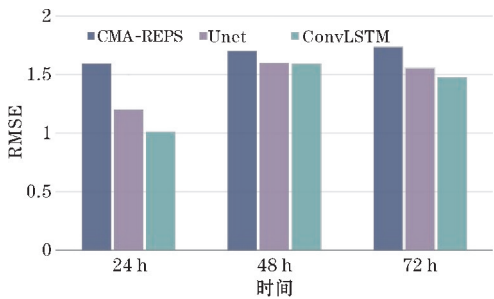


图3 24 h、48 h 及 72 h 预报 10u 订正前后 RMSE 对比图

图 4 显示了 2022 年 4 月 4 日 00 时 24 h 预报的 $10u$ 订正结果, 可见 CMA-REPS (图 4b) 原始预报数据整体误差较大, 经过 Unet 模型 (图 4c) 和 ConvLSTM 模型 (图 4d) 订正后误差有明显的改善, 并且 ConvLSTM 模型订正的结果从整体上更趋近于 ERA5 (图 4a), 着重体现在研究区域的东部沿海和北部地区。

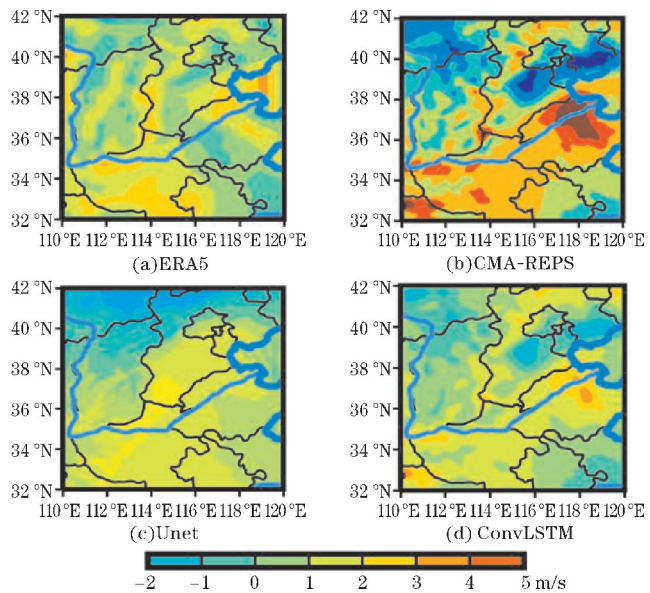


图4 2022年4月4日00时24h预报10u订正结果

3.3 10v 订正结果

图5显示了CMA-REPS、Unet模型和ConvLSTM模型对测试集24 h、48 h及72 h预报的10v数据进行偏差订正的RMSE对比结果。可以看出,ConvLSTM模型得到的预报数据效果更好。其中效果最好的是48 h预报的10v数据,与CMA-REPS和Unet模型相比,ConvLSTM的RMSE分别下降了38.45%和20.56%。其次是24 h预报的10v数据, RMSE比CMA-REPS和Unet模型分别降低25.16%和12.75%。对于72 h预报的10v偏差订正,Unet模型和ConvLSTM模型订正的效果提升幅度不大,但总体上偏差订正的结果更趋近于ERA5。

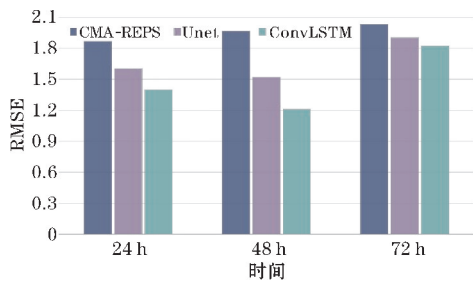


图5 24 h、48 h及72 h预报10v订正前后RMSE对比图

从2022年3月25日00时24h预报的10v订正结果(图6)可以看出,在CMA-REPS原始预报数据整体偏小的情况下,Unet模型对研究区域北部地区的偏差订正效果欠佳,而ConvLSTM模型对其进行偏差订正的结果整体上都具有明显的改善,着重表现在研究区域的中部和北部地区,在东部沿海地区也取得了不错的订正效果。

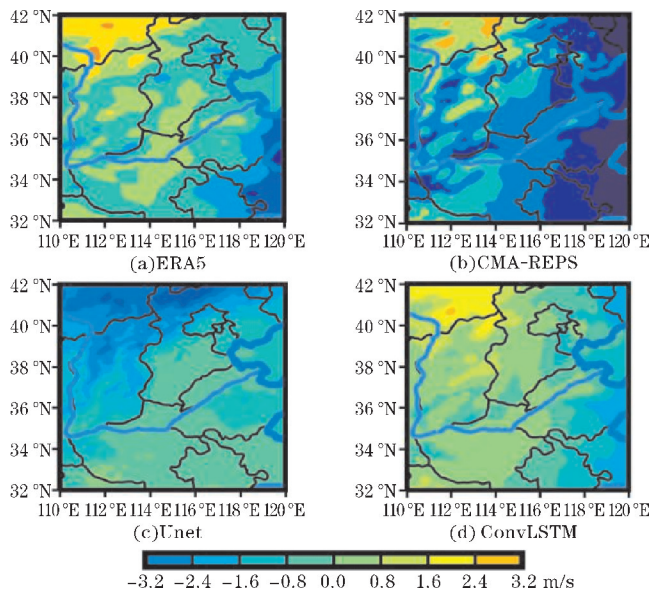


图6 2022年3月25日00时24h预报10w订正结果

3.4 10w 订正结果

图7显示了CMA-REPS、Unet模型和ConvLSTM模型对测试集24 h、48 h及72 h预报的10w数据进行偏差订正的RMSE对比结果。从图7可以看出,ConvLSTM模型对不同预报时效的10 m风速的订正均有明显的提升,尤其是对48 h预报的10w订正,ConvLSTM的RMSE相比CMA-REPS原始预报数据和Unet模型降低了28.37%与11.21%。24 h预报的10w经ConvLSTM订正后,相比CMA-REPS和Unet模型RMSE分别降低了19.42%和6.01%。对于72 h预报的10w,经ConvLSTM订正后的RMSE比CMA-REPS降低18.52%,比Unet模型降低8.19%,整体改善效果与24 h预报的10w相比相差不大。

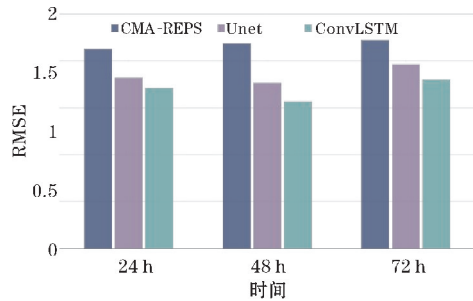


图7 24 h、48 h及72 h预报的10w订正前后RMSE对比

图8显示了2022年2月7日00时的48 h预报的10w订正结果。可以看出,该时刻的CMA-REPS原始预报数据整体上数值偏大,经Unet模型偏差订正得到的预报结果虽有所改善,但整体数值上相比于ERA5依旧稍微偏大。融合了时间序列特征编码特性和空间特征提取功能的ConvLSTM模型偏差订正的结果在整体上与ERA5更加一致,说明本文提出的ConvLSTM

模型可以有效改善CMA-REPS对10 m风速预测值偏高的情况。

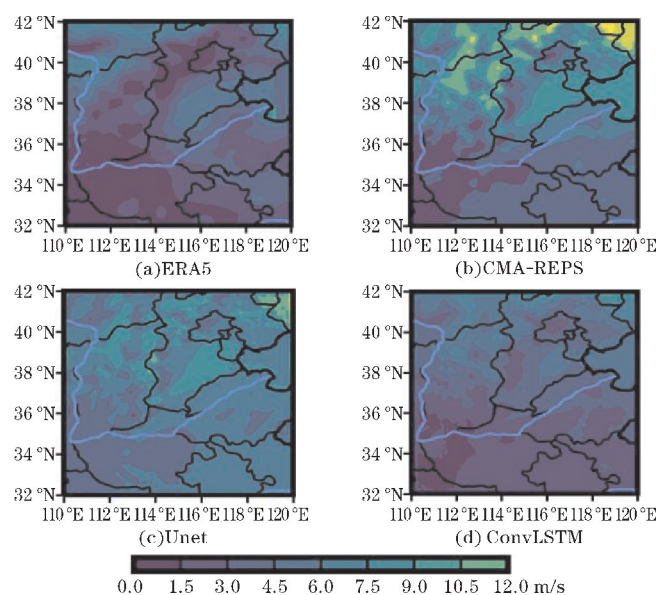


图8 2022年2月7日00时48h预报10m订正结果

4 结论

基于深度学习方法对中国自主研制的CMA-REPS预测的近地面10 m风场格点数据进行了偏差订正。针对格点预报数据具有的空间和时间双重特征,构建了一种融合时间序列特征编码特性和空间特征提取功能的ConvLSTM订正模型,并将其订正结果与仅考虑空间特征提取的Unet模型的订正结果和CMA-REPS原始预报数据进行对比。实验结果表明,ConvLSTM对CMA-REPS 3个预报时效(24 h、48 h和72 h)的近地面10 m风场预报数据均有明显的正向订正作用,且订正效果均要优于Unet模型的订正效果,这说明融合格点预报数据的时间序列特性可以帮助深度学习模型进一步提升对格点预报数据的偏差订正能力。另外从3个个例分析可以看出,在CMA-REPS原始预报数据整体较离散的情况下,ConvLSTM可以减小预报数据与实况数据间的误差,有效提升CMA-REPS对近地面10 m风场数据的预报能力。

随着深度学习技术的不断进步,注意力机制^[20]已被证明可以有效提升模型对序列数据的特征提取能力,考虑将注意力机制与ConvLSTM模型结合可以进一步提升模型的偏差订正效果。另外,将传统集合数值预报模式订正方法如集合模式输出统计(EMOS)与深度学习模型方法融合,也是未来集合数值预报模式偏差订正值得尝试的研究方向。

参考文献:

- [1] Deng Y, Wang B, Lu Z. A hybrid model based on data preprocessing strategy and error correction system for wind speed forecasting[J]. Energy Conversion and Management, 2020, 212(8): 112779.
- [2] 代刊,朱跃建,毕宝贵. 集合模式定量降水预报的统计后处理技术研究综述[J]. 气象学报, 2018, 76(4): 493-510.
- [3] 任晨辰,段明铿. ECMWF全风速场集合预报结果的偏差订正与预报不一致性分析[J]. 大气科学, 2019, 42(3): 360-369.
- [4] Gneiting T, Raftery A E, Westveld A H, et al. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation[J]. Monthly Weather Review, 2005, 133(5): 1098-1118.
- [5] Bao L, Gneiting T, Gruit E P, et al. Bias Correction and Bayesian Model Averaging for Ensemble Forecasts of Surface Wind Direction[J]. Mon. Wea. Rev. 2010, 138(5): 1811-1821.
- [6] Lahouar A, Slama J B H. Hour-ahead wind power forecast based on random forests[J]. Renewable Energy, 2017, 109: 529-541.
- [7] 孙全德,焦瑞莉,夏江江,等. 基于机器学习的数值天气预报风速订正研究[J]. 气象, 2019, 45(3): 426-436.
- [8] Rasp S, Lerch S. Neural networks for postprocessing ensemble weather forecasts[J]. Monthly Weather Review, 2018, 146(11): 3885-3900.
- [9] Shelhamer Evan, Long Jonathan, Darrell Trevor. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(4).
- [10] Grnquist P, Yao C, Ben-Nun T, et al. Deep Learning for Post-Processing Ensemble Weather Forecasts[J]. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200092.
- [11] Han L, Chen M, Chen K, et al. A Deep Learning Method for Bias Correction of ECMWF 24-240 h Forecasts[J]. Advances in Atmospheric Sciences, 2021, 38(9): 1444-1459.
- [12] 张延彪,陈明轩,韩雷,等. 数值天气预报多要素深度学习融合订正方法[J]. 气象学报,

- 2022,80(1):153–167.
- [13] Huang G, Liu Z, Laurens V, et al. Densely Connected Convolutional Networks[J]. CoRR,2016.
- [14] Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[J]. Advances in neural information processing systems,2015,28.
- [15] 陈静,李晓莉. GRAPES 全球/区域集合预报系统10年发展回顾及展望[J]. 气象科技进展,2020,10(2):9–18.
- [16] 沈学顺,王建捷,李泽椿,等. 中国数值天气预报的自主创新发展[J]. 气象学报,2020,78(3):451–47.
- [17] 王婧卓,陈静,庄照荣,等. GRAPES 区域集合预报模式的初值扰动增长特征[J]. 大气科学,2018,42(2):367–382.
- [18] 袁月,李晓莉,陈静,等. GRAPES 区域集合预报系统模式不确定性的随机扰动技术研究[J]. 气象,2016,42(10):1161–1175.
- [19] 王婧卓,陈法敬,陈静,等. GRAPES 区域集合预报对2019年中国汛期降水预报评估[J]. 大气科学,2021,45(3):664–682.
- [20] 吴政秋,张进,陈静,等. GRAPES 区域集合预报条件性台风涡旋重定位方法研究[J]. 气象学报,2020,78(2):163–176.
- [21] 谭晓光,罗兵. 天气预报分析型数据模型及生成[J]. 应用气象学报,2014,25(1):120–128.
- [22] Diederik P Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization[J]. CoRR,2014.
- [23] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems,2017:30.

A Deep Learning Method for Wind Speed Grid Point Forecasting Data Correction based on CMA-REPS

MAO Bo¹, YANG Hao^{1,2}, ZHOU Shijie², YANG Kangquan^{3,4}, CHEN Min¹

(1. Department of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China; 2. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China; 3. Sichuan Meteorological Observatory, Chengdu 610072, China; 4. Key Laboratory of Sichuan Meteorological Bureau, Chengdu 610072, China)

Abstract: Accurate wind speed prediction is of great significance for the full utilization of wind energy resources and the improvement of the economic benefits of wind farms. To improve the wind speed forecasting capability of ensemble numerical forecasting, this paper introduces the ConvLSTM deep learning model to perform a bias correction test on the grid point data of near-surface 10 m wind speed predicted by CMA-REPS (China Meteorological Administration-Regional Ensemble Prediction System), and the root mean square error (RMSE) was used as the scoring criterion to compare the correction results with the original prediction data of CMA-REPS and the correction results obtained by the Unet deep learning model method. The results show that the correction effect of the ConvLSTM model can be further improved than that of the Unet model, and the prediction data of 10 m wind speed near the surface after the modification of the ConvLSTM model is closer to the real data.

Keywords: CMA-REPS; ensemble forecast; bias correction; deep learning; wind speed