

基于领域先验知识的时空神经网络 模型在 MJO 预报中的应用

张 弄^{1,2}, 徐哲文², 刘长征³

(1. 中国工程物理研究院 计算机应用研究所, 四川 绵阳 621900;

2. 吉林大学 计算机科学与技术学院, 长春 130012;

3. 国家气候中心 中国气象局气候预测研究重点开放实验室, 北京 100081)

摘要: 针对目前人工神经网络方法无法准确预报季节性气候现象 Madden-Julian 振荡(MJO)的问题, 提出一种基于领域先验知识的时空神经网络模型。首先, 该方法结合气候环流数据的特性, 融入领域先验知识进行数据预处理; 其次, 采用预训练-微调架构, 利用次季节-季节的模式数据进行模型预训练, 并通过再分析数据(ERA5)完成微调; 最后, 通过时空建模, 选用卷积神经网络和长短期记忆网络结合的框架, 将先验知识嵌入预训练过程并优化预报。实验结果表明, 该模型能实现 23 d 的 MJO 准确预报, 其性能优于其他人工神经网络方法及国内数值预报方法。

关键词: 领域先验知识; 时空神经网络; Madden-Julian 振荡预报; 预训练模型

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2026)03-0617-10

Application of Domain Prior Knowledge-Based Spatio-temporal Neural Network Model in MJO Forecasting

ZHANG Nong^{1,2}, XU Zhewen², LIU Changzheng³

(1. Institute of Computer Application, China Academy of Engineering Physics, Mianyang 621900, Sichuan Province, China;

2. College of Computer Science and Technology, Jilin University, Changchun 130012, China;

3. China Meteorological Administration Key Laboratory for Climate Prediction Studies,
National Climate Center, Beijing 100081, China)

Abstract: Aiming at the problem that current artificial neural network methods could not accurately forecast the Madden-Julian oscillation (MJO), a seasonal climate phenomenon, we proposed a domain prior knowledge-based spatio-temporal neural network model. Firstly, we integrated domain prior knowledge into data preprocessing according to the characteristics of climate circulation data. Secondly, a pretraining-finetuning architecture was adopted, model data from the subseasonal-to-seasonal scale were used for pretraining model, and we completed finetuning by using the ERA5 reanalysis data. Finally, through spatio-temporal modeling, a framework combining convolutional neural networks and long short-term memory networks was selected to embed the prior knowledge into the pretraining process and optimize the forecasting. Experimental results show that the proposed

收稿日期: 2024-12-25.

第一作者简介: 张 弄(1999—), 男, 汉族, 硕士, 从事分布式计算和高性能计算的研究, E-mail: 18844128232@163.com. **通信**

作者简介: 徐哲文(1996—), 男, 汉族, 博士, 从事高性能计算和图神经网络的研究, E-mail: zwxu20@mails.jlu.edu.cn.

基金项目: 吉林省自然科学基金面上项目(批准号: 20230101062JC)和国家自然科学基金(批准号: 42175052).

model can achieve accurate MJO forecasting for 23 d, and its performance is superior to other artificial neural network methods and domestic numerical forecasting methods.

Keywords: domain prior knowledge; spatio-temporal neural network; Madden-Julian oscillation forecasting; pretraining model

0 引 言

季节性气候现象 Madden-Julian 振荡(MJO)^[1-2]是热带地区一种重要的大气季节内振荡现象,其周期约为 30~90 d,主要表现为对流和环流耦合的东传特征,是全球次季节-季节(sub-seasonal to seasonal, S2S)^[3]天气预报的主要来源之一.同时,MJO 对流和环流的异常还会影响全球天气和气候,研究 MJO 可为决策者提供探索极端天气事件风险的变化或优化资源管理的机会.为跟踪 MJO 的强度和位置,研究人员通常使用实时多变量 MJO(real-time multivariate MJO, RMM)的 MJO 指数^[4],如图 1 所示.

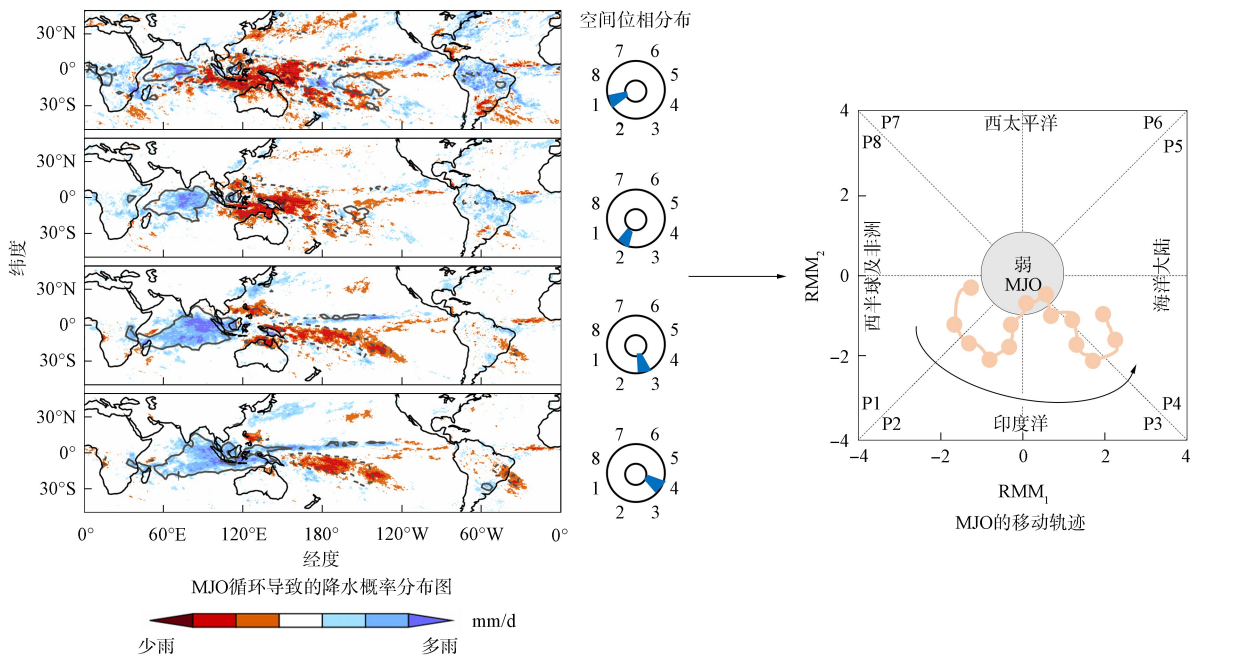


图 1 MJO 向东传播的示例
Fig. 1 Example of MJO spreading eastward

目前的 MJO 预报方法包括求解 Navier-Stokes 方程的数值天气预报(numerical weather pediction, NWP)^[5]和基于深度学习的人工神经网络(artificial neural network, ANN)^[6]的方法.大多数 NWP 方法都依赖于气候系统模型,该模型将地球系统的演化过程转化为计算流体动力学、偏微分方程等,并以数值方式求解.这类方法的 MJO 预报利用 NWP 预测的天气气候预测环流数据计算 RMM 指数.目前,NWP 方法提供了较准确的 MJO 预报,如中国气象局(China meteorological administration, CMA)^[7]的 MJO 预报达到 21 d.因为需要计算大量的方程,因此 NWP 方法需消耗海量的计算资源.由于 ANN 的泛化性和实用性主要从大数据中提取,一次大规模的训练后便可部署到普通机器上进行快速推理,相比数值方法更高效,因此,ANN 在天气气候预测方面具有重要作用.人工智能应用于 ENSO(EI Niño-southern oscillation)的预测^[8]展示了 ANN 方法在理解复杂气候模式中的重要作用.然而,与 NWP 方法相比,ANN 方法在 MJO 预测上准确性仍不足,例如,全连接神经网络(FNN)^[9]对 MJO 的预报在最高冬季达 17 d,最低是夏季,只有 10 d.近年来,以复旦大学“伏羲”大模型^[10]为代表的人工智能气象预测系统也取得了显著进展.该模型在 S2S 尺度预测中表现出色,但其作为综合性大模型,参数规模大、依赖全球多变量数据,对算力资源相对有限或仅聚焦 MJO 特定预测任务的场

景,存在部署门槛高、计算开销大的问题.

目前,ANN预测MJO存在诸多阻碍,其中较显著的原因列举如下:

1) MJO事件样本不足,即训练数据不足.通常,利用第5代欧洲中期天气预报中心(European centre for medium-range weather forecast, ECMWF)的再分析数据(ERA5)^[11]进行ANN方法的训练,但ERA5是从1950年开始记录的,且每年仅有60~90 d(12月至次年2月)出现明显的MJO事件,表明只有约6 000个MJO事件的训练样本.这对于依靠数据驱动的ANN方法显著不足,因此ANN方法无法有效捕捉信息,导致预测精度持续低于NWP方法.

2) 由于天气数据存在大量气候要素成分的叠加,以及这类混沌理论数据分布各不相同,因此捕捉MJO复杂的时空信息具有挑战性.上述样本的缺乏也加剧了解决该问题的难度.图1中的RMM相图显示了MJO随时间变化的轨迹,它在空间和时间上都是非线性分布的.同时,气候环流数据包含大量与MJO预报无关的信息,这些信息无法被ANN主动去除,导致对MJO信息的捕捉更困难.

为解决上述问题,本文提出一种基于领域先验知识的时空神经网络模型(domain prior knowledge-guided spatio-temporal neural network, DPK-STNN),它以传统的时空神经网络^[12],即卷积神经网络(convolutional neural network, CNN)与长短期记忆网络(long short-term memory, LSTM)结合的框架为基础,在预训练-微调的训练架构中嵌入了有关气候的领域先验知识,以增强对小样本环境下MJO关键特征的捕捉能力.相较于“伏羲”等大型综合预测模型,DPK-STNN更注重轻量化设计和小样本适应性,在保持较高预测精度的同时降低计算成本.实验结果表明,DPK-STNN在MJO预测性能上有较大进步,在预测效果上超过了CMA和FNN的方法.

1 MJO问题建模

本文旨在创建一个端到端的多变量时间序列回归模型进行较准确的MJO预报,因此对MJO预报问题进行了形式化建模.其目标是根据跨时间的相关气候环流变量预测未来一段时间的RMM指数.与其相对应的问题建模需使用时间序列的观测数据作为输入.然后,通过一个函数推断输出,即未来一段时间的RMM指数.本文先对输入和输出进行以下定义.

输入:首先,如图2所示,将全球热带地区 $g(0^\circ\sim 360^\circ, 15^\circ\text{N}\sim 15^\circ\text{S})$ 根据经度和纬度划分为 $l\times w(2.5^\circ\times 2.5^\circ)$ 的网格图像.每个网格 $r_{i,j}$ 被描述为位于第 i 行和第 j 列的区域.其次,将该区域在时间 t 内影响MJO的相关气候变量定义为输入数据图像,即矩阵 $\mathbf{X}_t\in\mathbb{R}^{l\times w\times 4}$,其中4表示4个大气和海洋的环流变量,包括地球长波辐射(outgoing longwave radiation, OLR)、海平面温度(sea surface temperature, SST)、200 hPa纬向风(U200)和850 hPa纬向风(U850),经过咨询气象专家和查阅相关文献^[13],可知这些变量是影响MJO的主要因素.最后,在给定最新时间点 t 和历史时间段 k 的条件下,将输入数据图像时间序列表示为 $\mathbf{X}=(\mathbf{X}_{t-k+1},\dots,\mathbf{X}_{t-1},\mathbf{X}_t)\in\mathbb{R}^{k\times l\times w\times 4}$,其中 $\mathbf{X}(t,i,j,*)$ 表示时间 t 内 $r_{i,j}$ 区域的任一变量(OLR/U200/U850/SST).

输出:利用RMM作为MJO指数跟踪其强度和位置.每个时间点 t (本文表示1 d)有两个RMM指数,即 RMM_1 和 RMM_2 ,其根据文献[4]计算得出.因此,用 $\mathbf{Y}_t=[\text{RMM}_1,\text{RMM}_2]$ 表示时间 t 内输出的MJO指数.最后,在给定下一个时间点 $t+1$ 和未来时间段 n 的条件下,用 $\mathbf{Y}=(\mathbf{Y}_{t+1},\mathbf{Y}_{t+2},\dots,\mathbf{Y}_{t+n})\in\mathbb{R}^{n\times 2}$ 表示输出的RMM指数时间序列.

如图3所示,使用函数 f 连接输入和输出,其中 f 是一个时空网络模型,通常表示为 $f(\mathbf{X},\theta)$, $\mathbf{X}$ 为输入数据, θ 表示可学习的超参数.时空网络模型通常有多个层,每层都包含大量的可学习参数,这些参数初始化为白噪声,并通过反向传播网络模型中的预测误差进行优化,常见的有梯度下降算法.通过反复的训练和迭代优化,可获得一个最优函数,使预期值和真实值之间的误差最小,可表示为

$$\min_f \sum_{i=t+1}^{t+n} \text{error}(\mathbf{Y}_i, \hat{\mathbf{Y}}_i), \quad \hat{\mathbf{Y}} = f(\mathbf{X}, \theta), \quad (1)$$

其中 $\hat{\mathbf{Y}}=(\hat{\mathbf{Y}}_{t+1},\hat{\mathbf{Y}}_{t+2},\dots,\hat{\mathbf{Y}}_{t+n})$ 为模型的预测结果, $\mathbf{Y}=(\mathbf{Y}_{t+1},\mathbf{Y}_{t+2},\dots,\mathbf{Y}_{t+n})$ 为观测的真实值,error可根据平均绝对误差(mean absolute error, MAE)或均方根误差(root mean square error, RMSE)^[14]进

行计算.

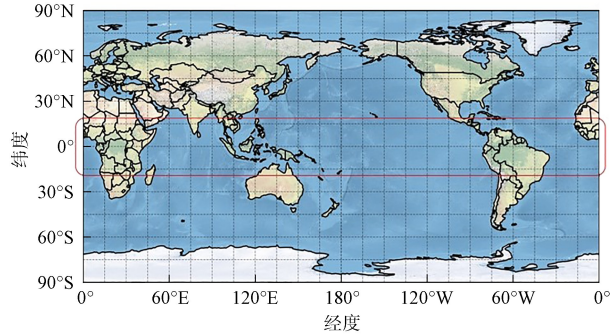


图 2 全球热带区域网格图

Fig. 2 Global grid map of tropical regions

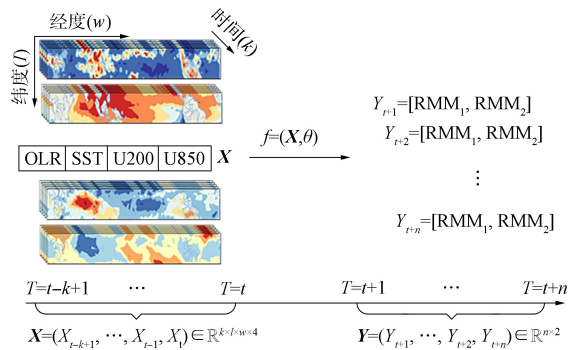


图 3 MJO 预报问题建模

Fig. 3 MJO forecasting problem modeling

2 预测模型

针对上述 MJO 预报的问题建模, 本文设计了对应的时空神经网络模型, 即 DPK-STNN. 参考传统的时空网络模型, 分别设计了空间提取模块和时间提取模块, 先按顺序提取对应的空间和时间特征, 再进行输出. 同时, 为解决简单时空神经网络模型预测 MJO 的准确率较低问题, 本文选择预训练-微调架构, 并参考气候专业知识, 在模型中嵌入了领域先验知识模块, 用于缓解模型训练与提取时空特征^[15-16]的压力. 如图 4 所示, 模型的框架结构概述图主要分为 4 个模块, 包含预训练与微调、领域先验知识、空间信息提取以及时间信息提取模块.

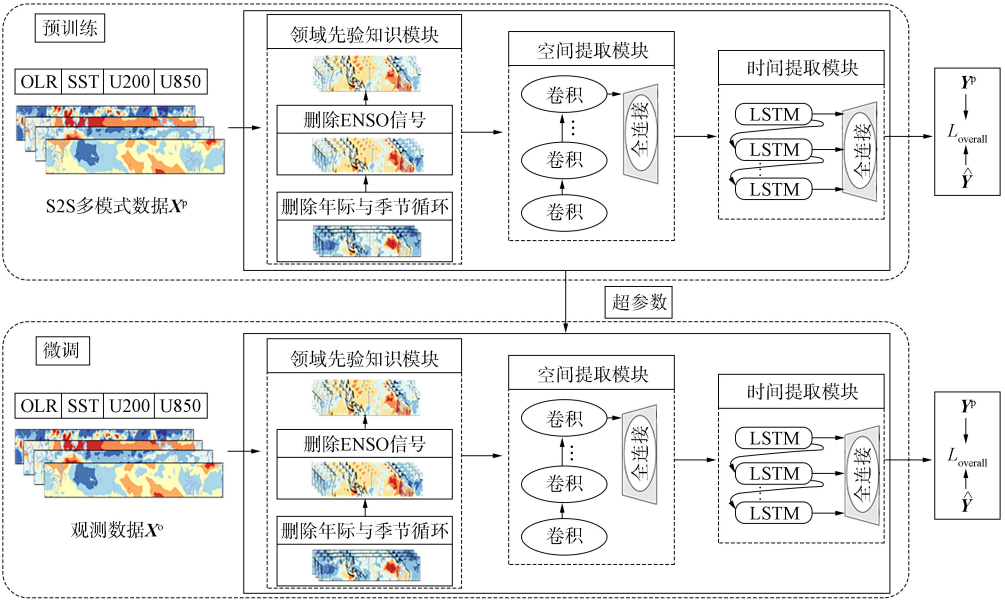


图 4 DPK-STNN 模型概述

Fig. 4 Overview of DPK-STNN model

2.1 预训练-微调架构

基于上述分析可知, 在有限的气候观测数据下, 深度学习模型的训练样本不足会严重影响其性能. 为应对这一挑战, 本文参考并借鉴了预训练-微调方法的思想, 设计出适用于本文研究的新颖架构. 预训练-微调方法的基本原理是通过两个阶段的训练提高深度学习模型的性能. 在预训练阶段, 模型通过大规模数据集(通常是模拟数据或大量无标签数据)进行训练, 学习捕捉数据集中的一般性特征和结构. 这个预训练的模型在第二阶段, 即微调阶段, 会被继续训练, 但这次使用特定任务的有标签数据, 以调整模型适应任务特定的要求. 预训练-微调方法广泛应用于自然语言处理领域, 如 BERT (bidirectional encoder representations from transformers)^[17]. 在 BERT 的预训练阶段, 模型通过大规

模的语言模型预训练,学习了深层次的语言表示. 然后通过微调,使用特定的任务数据,如情感分析或文本分类, BERT 能适应特定任务的语境,使性能显著提升.

基于此,本文设计了类似的两阶段策略,以更好地利用现有的气象信息. 首先是预训练阶段. 在此阶段,使用数值预报方法生成历史回测数据,即模式数据^[7],这是一种传统而可靠的数据源,用于模拟气象系统的演变. 这样的预训练过程使模型能从中学到丰富的气象特征和系统动力学规律,为后续微调提供了基础. 这里,本文主要依靠国家气候中心提供的 S2S 产品数据库进行预训练,该数据库集成了 11 个国家气候中心最先进的模式数据,并能进行多模式耦合. 本文选取中国气象局(CMA)、欧洲中期天气预报中心(ECMWF)以及美国国家环境预报中心(NCEP)的模式数据进行预训练. 数据集 S2S 中部分数据信息列于表 1.

表 1 数据集 S2S 中部分数据信息
Table 1 Partial data information in S2S dataset

模式	分辨率	预报时段	频率	预报时间范围/d
CMA	1°×1°	1994 年—2014 年	每天	0~60
ECMWF	0.25°×0.25°	过去 20 年	每周 2 次	0~46
NCEP	1°×1°	1999 年—2010 年	每天	0~44

其次是微调阶段,本文利用真实的气象观测数据,即 ERA5,以更好地调整模型适应实际气象情境. 通过在微调过程中对模型超参数进行调整,最小化模型输出与真实观测数据之间的差异,以提高模型的准确性和适用性.

2.2 领域先验知识模块

在气候数据中, MJO 相关的信息与其他信息并存,甚至 MJO 信息只占了很小一部分,因此 MJO 信息更像是气候数据中的“噪音”,使本文的深度网络模型难以捕捉. 这也是阻碍 MJO 准确预报的第二个原因. 本文分析了气候环流数据的主要组成,在咨询了相关的气候专家,并查阅了相关文献^[4]后,发现气候数据中主要包含年际循环、季节循环和 ENSO 信号等. 因此,本文设计了领域先验知识模块,作为数据预处理,分步删除相关的信号特征,使 MJO 信号作为主要信息. 同时,将其嵌入到时空网络模型中参与训练.

如图 4 所示,先删除年际与季节循环. 参考文献[18]的 Kalman 滤波方法,先对每个数据 $\mathbf{X}(t, i, j, *)$ 的气候态进行谐波分析,按

$$\mathbf{X}(t) = a_0 + \sum_{n=1}^{\infty} [a_n \cos(n\omega t) + b_n \sin(n\omega t)] \tag{2}$$

展开,再按

$$\bar{\mathbf{X}}^3(t) = a_0 + \sum_{n=1}^3 [a_n \cos(n\omega t) + b_n \sin(n\omega t)] \tag{3}$$

计算其逐日气候态的 0~3 波,即年际与季节循环对应的数据. 其中: t 表示时间变量,本文选择单位 d; n 为谐波数; ω 为圆频率,单位为每天(d^{-1}); a_0 为常数,表示时间序列 $\mathbf{X}(t)$ 的平均值; a_n, b_n 分别为第 n 波的余弦、正弦函数系数. 最后,删除其中的 0~3 波,表示为

$$\mathbf{X}'(t) = \mathbf{X}(t) - \bar{\mathbf{X}}^3(t). \tag{4}$$

对于 ENSO 信号,其主要存在于环流数据的前 120 d 平均中^[19],因此,按

$$\mathbf{X}''(t) = \mathbf{X}'(t) - \left(\sum_{i=1}^{120} \mathbf{X}'(t-i) \right) / 120 \tag{5}$$

再减去前 120 d 的平均值.

2.3 空间提取模块

常见的时空神经网络模型都是时间网络模型和空间网络模型的变换组合,分步提取时间和空间的相关特征. 本文参考常见的方法,先利用空间网络模型,即 CNN^[20]提取空间特征. 由于本文数据在进行领域先验知识模块的预处理后,其大小不变,仍是 $\mathbf{X}'' \in \mathbb{R}^{k \times l \times w \times 4}$,属于时空数据. 因此先不考虑时间序列,得到 k 个数据 $\mathbf{X}^* \in \mathbb{R}^{l \times w \times 4}$. 针对每个 $\mathbf{X}^*(i, j, c)$, 它的 i, j 和 c 分别对应卷积层中图片的长

度、宽度和通道数。
经过多层卷积：

$$\mathbf{Z} = \text{conv}(\text{conv}(\cdots \text{conv}(\mathbf{X}^*))) , \tag{6}$$

可得到新的 k 个数据, 再拼接, 得到 $\mathbf{Z} \in \mathbb{R}^{k \times L \times W \times C}$, 其中 $L \times W$ 为平铺后的二维矩阵的维度, C 为输出的特征维度. 最后, 为匹配时间网络模型, 加入一个全连接层, 以组合所有的空间特征. 在数据上, 得到了 $\mathbf{Z}^* \in \mathbb{R}^{k \times \hat{C}}$, 其中 $\hat{C} = L \times W \times C$.

2.4 时间提取模块

常见的时间神经网络模型包括循环神经网络(recurrent neural network, RNN)、门控循环单元神经网络(gated recurrent unit neural network, GRU)以及长短期记忆神经网络 LSTM^[21-22]等. 其中 GRU 和 LSTM 均由 RNN 发展而来, 为解决时间序列预测中存在的长存依赖问题. 如图 4 所示, 本文选择 LSTM 作为时间网络模型, 因为 LSTM 比 GRU 能存储更多的相关信息. 图 5 为 LSTM 的网络结构, 其中 f_t, i_t, o_t 分别表示遗忘门、输入门和输出门, W_* 和 b_* 分别表示每个门的权重和偏差. 与 GRU 相比, LSTM 多了单元 c_t , 能用来存储更多的隐藏信息, 对于复杂非线性的气候序列数据极其重要.

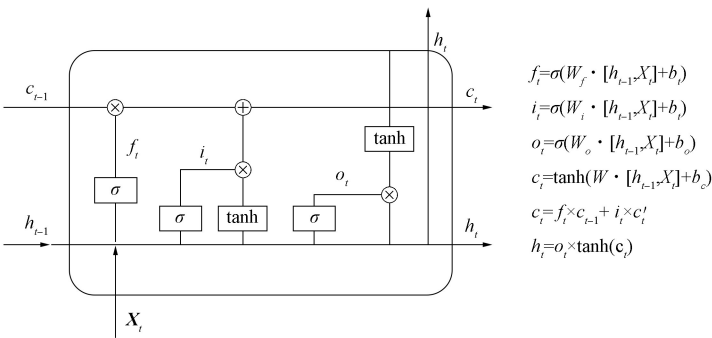


图 5 LSTM 的网络结构
Fig. 5 Network structure of LSTM

通过空间卷积模块, 得到了数据 $\mathbf{Z}^* \in \mathbb{R}^{k \times \hat{C}}$, 结合图 3, 需利用前 k 个时间段(d)预测未来 n 个时间段(d), 因此在 DPK-STNN 的时间提取模型(见图 4)中串行了 n 个 LSTM. 对于第 1 个 LSTM, 输入数据 $\mathbf{Z}^*$, 得到下一个时间段时间特征, 即 h_k 和 c_k . 其中 h_k 不仅作为输出的一部分, 保留在集合 H 中, 还与 c_k 一起作为下一个 LSTM 的输入. 此外, 下一个 LSTM 的输入还需对 $\mathbf{Z}^*$ 的数据滑窗, 利用后 $(k-1)$ 个数据与 h_k 合并, 得到新的 $\mathbf{Z}^*$ 进行输入. 最后, 再串连了 n 个 LSTM 后, 得到 $\mathbf{H} = (h_1, h_2, \cdots, h_n) \in \mathbb{R}^{n \times D}$. 由于需要得到时间序列为 n 的 RMM₁ 和 RMM₂ 指数序列, 所以再加入了一个线性全连接层, 将 $\mathbf{H} \in \mathbb{R}^{n \times D}$ 输出为 $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times 2}$.

2.5 自适应损失函数

损失函数作为深度学习(主要是监督学习)最重要的组成部分, 反映了神经网络模型预测输出与真实样本之间的差异程度. 常见的损失函数有误差的均方误差损失(mean squared error loss, MSELoss)^[23]和交叉熵损失(cross entropy loss, CELoss)^[24]等. MSELoss 适用于回归任务, 能有效评估连续输出的误差. 但一般的损失函数很难适用于所有情形. 在本文的时空序列预测任务中, 需要预测结果包含 RMM₁ 和 RMM₂ 两个指数序列. 由于两个指数的动态范围和重要性不同, 因此传统 MSELoss 无法达到最佳效果. 为解决该问题, 本文设计一个自适应的损失函数^[25]:

$$L_{\text{overall}} = \beta L_1 + \gamma L_2 = \beta \left(\frac{1}{n} \sum_{i=1}^n \hat{Y}(i, 1) - Y(i, 1) \right)^2 + \gamma \left(\frac{1}{n} \sum_{i=1}^n \hat{Y}(i, 2) - Y(i, 2) \right)^2 . \tag{7}$$

该偏执归纳可帮助损失函数自适应地为 RMM₁ 和 RMM₂ 设置权重比例, 使模型能关注更重要的指数. 本文引入两个权重系数 β 和 γ , 其中 β 与 RMM₁ 预测结果的误差相关, γ 与 RMM₂ 预测结果的误差相关, 而 RMM₁ 和 RMM₂ 分别利用 MSELoss 函数计算. 权重系数 β 和 γ 可通过梯度下降算法动

态调整, 以最小化验证集上的损失函数. 在后续实验中, 发现该方法明显提升了 RMM_1 和 RMM_2 的预测精度, 证明了其有效性.

3 实验及结果分析

3.1 实验设置

3.1.1 数据集

本文基于模式数据和再分析数据构建相关的数据集. 首先, 从对应数据来源中选择 4 种环流气候变量, 即 OLR, U200, U850 和 SST; 其次, 选择它们的全球热带区域($0^{\circ}\sim 360^{\circ}$, $15^{\circ}\text{N}\sim 15^{\circ}\text{S}$)并按分辨率 2.5° 进行划分; 再次, 分别选择从 1950, 2001, 2006 及 2004—2020/2024 年 4 月 30 日的逐日数据, 同时, 为进一步扩充数据量, 还对不同数据进行不同程度的滑窗; 最后, 对数据进行划分, 具体信息列于表 2. 本文采用预训练-微调架构, 预训练阶段, 选择 S2S 模式数据中的 CMA, ECMWF 和 NCEP 的全部数据进行预训练, 即到 2020 年 12 月 31 日. 微调阶段, 仅使用 ERA5 再分析数据, 其中 2020 年 12 月 31 日之前的数据作为训练集进行训练, 2021 年 1 月 1 日至 2021 年 12 月 31 日的数据作为验证集进行验证. 剩余的 ERA5 数据作为测试集进行性能评估.

表 2 数据集及时间划分
Table 2 Dataset and its temporal division

数据来源	时间范围	训练集	验证集	测试集
ERA5	1950-01-01—2024-04-30	1950-01-01—2020-12-31	2021-01-01—2021-12-31	2022-01-01—2024-04-30
CMA	2001-01-01—2020-12-31	2001-01-01—2020-12-31	—	—
ECMWF	2006-01-01—2020-12-31	2006-01-01—2020-12-31	—	—
NCEP	2004-01-01—2020-12-31	2004-01-01—2020-12-31	—	—

3.1.2 评价标准

参考相关的气象研究资料, 本文采用双变量相关系数(bivariate correlation coefficient, COR)和均方根误差(root mean square error, RMSE)两个指标评估 MJO 的预测能力, 公式如下:

$$\text{COR}(\tau) = \frac{\sum_{t=1}^N [Y_0(t) \times \hat{Y}_0(t, \tau) + Y_1(t) \times \hat{Y}_1(t, \tau)]}{\sqrt{\sum_{t=1}^N [Y_0^2(t) + Y_1^2(t)]} \sqrt{\sum_{t=1}^N [\hat{Y}_0^2(t, \tau) + \hat{Y}_1^2(t, \tau)]}}, \tag{8}$$

$$\text{RMSE}(\tau) = \sqrt{\frac{\sum_{t=1}^N [|Y_0(t) - \hat{Y}_0(t, \tau)|^2 + |Y_1(t) - \hat{Y}_1(t, \tau)|^2]}{N}}, \tag{9}$$

其中 $Y_0(t)$ 和 $Y_1(t)$ 分别表示在时间 t 处观测到的 RMM_1 和 RMM_2 , $\hat{Y}_0(t, \tau)$ 和 $\hat{Y}_1(t, \tau)$ 对应在时间 t 处预测未来 τ 天的 RMM_1 和 RMM_2 值, N 为样本总数. COR 表示预测值和观测值之间的共现强度, 即表示为 MJO 预测的位置; 而 RMSE 主要测量二者之间的误差, 即 MJO 预测的精度. 这两者的 MJO 技巧阈值通常设为 $\text{COR}=0.5$ 和 $\text{RMSE}=1.4$, 用于评估预测技巧. 特别地, 当 $\text{COR}(\tau)<0.5$ 且 $\text{RMSE}(\tau)>1.4$ 时, 其预测时间为第 τ 天的 MJO 预测技巧被认为不准确.

3.1.3 超参数及实验环境设置

对于空间卷积模块, 本文将卷积核分别设为 7×7 和 3×3 ; 对于时间模块, LSTM 需设置其隐藏特征大小, 通过实验选择设置为 512; 而对于 L_{overall} , 初始化两个权重系数 β 和 γ 为 0.5, 并随迭代训练改变. 在模型训练过程涉及的迭代测试、丢失率及学习率, 分别设为 30, 0.001 和 $1\times e^{-4}$. 选择 Adam 作为神经网络模型训练的优化器, 因为它可以自适应地改变学习率. DPK-STNN 模型基于 Pytorch 的深度学习框架, 在 GTX3090 24 GiB GPU 上进行所有训练.

3.1.4 基准模型

将 DPK-STNN 模型与以下方法进行比较:

- 1) 传统 NWP 方法, 选择中国气象中心的模式数据 CMA 计算对应测试集时间段的 RMM_1 和

RMM₂ 指数, 并计算对应的评价指标;

2) ANN 方法, 包括 FNN 方法及常见的时空网络模型 CNN+LSTM^[26] 和 ConvLSTM^[27], 其中, CNN+LSTM 模型也是本文模型的基础, 即空间提取模块+时间提取模块, 对它与 DPK-STNN 模型进行消融实验.

3.2 准确性

表 3 列出了不同方法的性能比较结果. 由表 3 可见, NWP 方法仍保持了强大的预测效果, 其对应的 CMA 方法, 预测技巧达 21 d; 相比之下, ANN 方法准确性较低, FNN 方法只能预测 16 d, 这与其简单的网络架构有关, 其模型结构主要是线性全连接层, 无法有效提取相关的 MJO 信息特征. 而常见的一些时空网络模型, 其复杂的网络结构能较好地提取相关的 MJO 特征信息, 包括 CNN+LSTM 和 ConvLSTM, 但其准确度还是低于 CMA 方法, CNN+LSTM 达 17 d, ConvLSTM 达 18 d. 与上述方法相比, 本文方法 DPK-STNN 模型在准确性上明显更优, 达 23 d. 同时, DPK-STNN 模型的 RMSE, 即 MJO 预测精度方面也远高于其他方法.

表 3 不同方法的性能比较结果

Table 3 Performance comparison results of different methods

方法	输入数据长度/d	评价指标/d		预测技巧/d
		COR≥0.50	RMSE≤1.40	(COR≥0.50)∩(RMSE≤1.40)
CMA	—	21	34	21
FNN	1	16	37	16
CNN+LSTM	7	17	30	17
ConvLSTM	7	18	32	18
DPK-STNN	7	23	38	23

为进一步显示 DPK-STNN 模型的预测效果, 选择对应测试集的时间段, 对 CMA, FNN 和 DPK-STNN 进行再次预测, 结果如图 6 所示. 由图 6(A)可见, 本文方法在 MJO 预报位置上远优于 FNN 方法, 同时也超过了传统的 NWP 方法; 由图 6(B)可见, 本文方法的 RMSE 较低, 表示其预测指数与真实值之间的误差较小. 实验结果表明, 本文方法在前 25 d 内, 其 RMSE 都低于 1.0; 随着预测时间的增加, RMSE 增大, CMA 方法在 34 d 左右达 1.4, 即 MJO 预报的最低阈值, 而此时本文方法仍低于 1.4. 表明本文方法不仅在预测位置上具有较高的准确性, 且在同一时间点, 其预测精度更高, 即 DPK-STNN 模型的预测更接近观测结果.

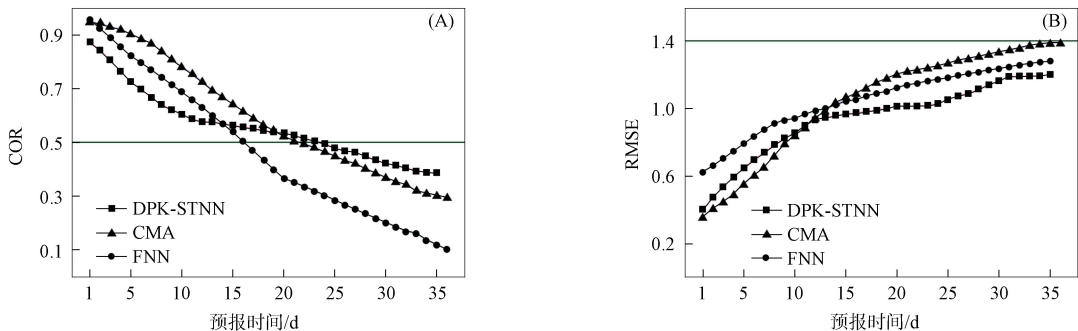


图 6 不同方法的准确性实验结果

Fig. 6 Accuracy experimental results of different methods

3.3 消融实验

本文方法是基于 CNN+LSTM 的模型框架, 加入了预训练-微调架构以及领域先验知识模块得到的. 因此, 设计消融实验验证相关模块的有效性. 表 4 列出了消融实验对比结果. 由表 4 可见, 基础的 CNN+LSTM 框架达到 17 d 的 MJO 准确预报, 虽然其结构能在一定程度上提取相关的 MJO 特征信息, 但由于训练样本不足和 MJO 信息复杂的原因, 其准确度仍低于 CMA 方法. 预训练-微调架构, 先利用模式数据进行预训练, 再利用观测的再分析数据进行模型微调, 在一定程度上解决了 MJO 训练

样本不足的问题, 其准确度达 19 d; 领域先验知识模块参考了气候领域的专业知识, 对数据进行了一系列预处理, 删除了气候数据中的许多噪声, 使 MJO 信息更容易提取; 同时, 也在一定程度上缓解了模型训练的压力, 对解决 MJO 训练样本不足有一定帮助, 其 MJO 预测技巧也提高了 3 d, 达到 20 d. 当上述两种方法都加入到 CNN+LSTM 框架, 即 DPK-STNN 方法时, 发现其预测技巧并不是累积增加 5 d, 而是增加了 6 d, 达到 23 d 的效果. 表明这两种方法对彼此具有相互增强的效果.

表 4 消融实验对比结果

Table 4 Comparison results of ablation experiments

方法	评价指标/d		预测技巧/d
	$COR \geq 0.50$	$RMSE \leq 1.40$	$(COR \geq 0.50) \cap (RMSE \leq 1.40)$
CNN+LSTM	17	30	17
只加入预训练-微调架构	19	33	19
只加入领域先验知识模块	20	35	20
DPK-STNN	23	38	23

综上所述, 本文提出了一种基于领域先验知识的时空神经网络(DPK-STNN)模型, 用于提高 MJO 的预报精度. 在数据处理和模型构建中融合了领域先验知识, 采用预训练-微调的架构, 先利用模式数据预训练模型, 再用再分析数据进行微调. 该方法通过 CNN+LSTM 模型框架嵌入领域先验知识, 实现了 MJO 的 23 d 长期精准预报. 实验结果表明, 本文方法显著超过其他神经网络方法和国内数值模式的性能, 证明了该方法的有效性.

参 考 文 献

[1] ZHANG C D. Madden-Julian Oscillation: Bridging Weather and Climate [J]. Bulletin of the American Meteorological Society, 2013, 94(12): 1849-1870.

[2] ZHANG C D. Madden-Julian Oscillation [J]. Reviews of Geophysics, 2005, 43(2): 1-36.

[3] HE S J, LI X Y, DELSOLE T, et al. Sub-seasonal Climate Forecasting via Machine Learning: Challenges, Analysis, and Advances [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(1): 169-177.

[4] WHEELER M C, HENDON H H. An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction [J]. Monthly Weather Review, 2004, 132(8): 1917-1932.

[5] LORENC A C. Analysis Methods for Numerical Weather Prediction [J]. Quarterly Journal of the Royal Meteorological Society, 1986, 112: 1177-1194.

[6] SCHULTZ M G, BETANCOURT C, GONG B, et al. Can Deep Learning Beat Numerical Weather Prediction? [J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2021, 379: 0097-1-0097-22.

[7] VITART F, ROBERTSON A W. The Sub-seasonal to Seasonal Prediction Project (S2S) and the Prediction of Extreme Events [J]. NPJ Climate and Atmospheric Science, 2018, 1: 3-1-3-7.

[8] HAM Y G, KIM J H, LUO J J. Deep Learning for Multi-year ENSO Forecasts [J]. Nature, 2019, 573(7775): 568-572.

[9] MARTIN Z, BARNES E, MALONEY E. Predicting the MJO Using Interpretable Machine-Learning Models [EB/OL]. (2022-11-23)[2024-10-15]. <https://doi.org/10.1002/essoar.10506356.1>.

[10] CHEN L, ZHONG X H, 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): 190-1-190-11.

[11] HERSBACH H, BELL B, BERRISFORD P, et al. The ERA5 Global Reanalysis [J]. Quarterly Journal of the Royal Meteorological Society, 2020, 146(730): 1999-2049.

[12] 谭奕鑫, 詹永照, 刘洪麟. 基于显著特征和时空图网络的视频异常事件检测 [J]. 江苏大学学报(自然科学版), 2025, 46(2): 179-188. (TAN Y X, ZHAN Y Z, LIU H L. Video Abnormal Event Detection Based on Salient Features and Space-Time Fusion Graph Network [J]. Journal of Jiangsu University (Natural Science Edition), 2025, 46(2): 179-188.)

- [13] LI H L, XU Z W, FANG Z L, et al. MJOFomer: An Adaptive Land-Ocean Spatio-temporal Transformer for Madden Julian Oscillation Forecasting [J]. *Computers & Geosciences*, 2025; 106097-1-106097-11.
- [14] CHAI T, DRAXLER R R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? —Arguments against Avoiding RMSE in the Literature [J]. *Geoscientific Model Development*, 2014, 7(3): 1247-1250.
- [15] 王绮, 黄国民, 卢恩琪, 等. 基于地理探测器的旅游生态安全时空特征及影响因素研究 [J]. *东北师大学报(自然科学版)*, 2024, 56(4): 133-143. (WANG Q, HUANG G M, LU E Q, et al. Study on the Spatial-Temporal Characteristics and Influencing Factors of Tourism Ecological Security Based on Geodetector [J]. *Journal of Northeast Normal University (Natural Science Edition)*, 2024, 56(4): 133-143.)
- [16] 贺金鑫, 张涵雅, 周俊宏, 等. 基于知识图谱的火星季节性斜坡纹线时空特征分析 [J]. *吉林大学学报(地球科学版)*, 2026, 56(1): 252-265. (HE J X, ZHANG H Y, ZHOU J H, et al. Analyzing Spatial and Temporal Features of Martian Recurring Slope Lineae Based on Knowledge Graph [J]. *Journal of Jilin University (Earth Science Edition)*, 2026, 56(1): 252-265.)
- [17] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.]: ACL, 2019: 4171-4186.
- [18] ROY S K, NICOLSON A, PALIWAL K K. A Deep Learning-Based Kalman Filter for Speech Enhancement [C]//Interspeech. [S.l.]: ISCA, 2020: 2692-2696.
- [19] TANG Y M, YU B. MJO and Its Relationship to ENSO [J]. *Journal of Geophysical Research: Atmospheres*, 2008, 113(14): 2007JD009230-1-2007JD009230-40.
- [20] LI Z W, LIU F, YANG W J, et al. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(12): 6999-7019.
- [21] DEY R, SALEM F M. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks [C]//2017 IEEE 60th International Midwest Symposium on Circuits and Systems. Piscataway, NJ: IEEE, 2017: 1597-1600.
- [22] SHERSTINSKY A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network [J]. *Physica D: Nonlinear Phenomena*, 2020, 404: 132306-1-132306-28.
- [23] ZHOU J X, LI X, DING T Y, et al. On the Optimization Landscape of Neural Collapse under MSE Loss: Global Optimality with Unconstrained Features [C]//International Conference on Machine Learning. [S.l.]: PMLR, 2022: 27179-27202.
- [24] HO Y, WOOKEY S. The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling [J]. *IEEE Access*, 2020, 8: 4806-4813.
- [25] PEZZANO G, RIPOLL V R, RADEVA P. CoLe-CNN: Context-Learning Convolutional Neural Network with Adaptive Loss Function for Lung Nodule Segmentation [J]. *Computer Methods and Programs in Biomedicine*, 2021, 198: 105792-1-105792-13.
- [26] FAN M, IMRAN O, SINGH A, et al. Using CNN-LSTM Model for Weather Forecasting [C]//2022 IEEE International Conference on Big Data (Big Data). Piscataway, NJ: IEEE, 2022: 4120-4125.
- [27] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting [J]. *Advances in Neural Information Processing Systems*, 2015, 28: 1-9.

(责任编辑: 韩 啸)