

黄颖,陆虹,黄小燕,等. 基于 EOF 和 LSTM 的广西月降水量预测模型研究[J]. 气象研究与应用,2023,44(2):20–26.

Huang Ying, Lu Hong, Huang Xiaoyan, et al. Study on monthly precipitation prediction model in Guangxi based on EOF and LSTM[J]. Journal of Meteorological Research and Application, 2023, 44(2): 20–26.

基于 EOF 和 LSTM 的广西月降水量预测模型研究

黄颖¹, 陆虹², 黄小燕¹, 赵华生¹, 吴玉霜³

(1.广西壮族自治区气象科学研究所, 南宁 530022; 2.广西壮族自治区气候中心, 南宁 530022;

3.广西壮族自治区气象台, 南宁 530022)

摘要: 针对夏季降水天气过程具有时间相关性和非线性变化的特点,以及现有预报方法未能充分获取月降水量的本质特征而造成的建模因子处理和预报建模困难等问题,提出了一种以自然正交展开(EOF)与深度学习长短期记忆神经网络(LSTM)相结合的月降水量预测模型。以广西 81 个气象观测站 7 月降水量为预报研究对象,对 81 站 7 月降水量作 EOF 计算,选取累积方差贡献超过 76%的前 7 个主分量作为预报分量,再利用 LSTM 模型建立月降水量的深度学习预测模型,以 1960—2016 年 81 站 7 月降水量为建模样本,2017—2022 年为独立样本进行建模研究。结果表明,在相同的预报建模样本和相同的预报因子条件下,新建立的预测模型比线性逐步回归预报方法有更高的预报能力,显示了对非线性月降水量预测问题的适用性。由于 LSTM 模型隐层里引入了存储单元状态和门结构,使得网络能够保留长期的状态,更适合于处理和预测时间序列中间隔和延迟相对较长的重要问题。

关键词: 气候预测;长短期记忆神经网络;自然正交展开;深度学习

中图分类号: P457.6

文献标识码: A

doi: 10.19849/j.cnki.CN45-1356/P.2023.2.04

引言

广西地处亚热带季风气候区,受低、中高纬度天气系统的交替影响,降雨量时空分布不均,尤其是夏季雨季暴雨频繁,往往导致洪涝、滑坡泥石流等灾害,造成巨大的经济损失及危害人民生命。另一方面,由于广西夏季降水年际变化差异极大,致使广西夏季干旱也频繁发生,严重制约了社会和经济的发展。因此,如果能够提前预测旱涝灾害发生的具体区域和时段,可以有效减轻灾害损失。然而,影响月降水量变化的因素很多,具有显著的非线性不规则变化特征,是目前预报中的重点、难点问题之一。

目前国内外学者开展降水量短期气候预测主要研究方法可分为数理统计预报方法、动力模式方法,以及动力统计预报三种方法^[1-4]。杜良敏等^[5]采用最小二乘回归法建立中国夏季降水预测模型,5a 独立

样本预测检验的平均评分为 77.1。郑然等^[6]采用多元回归法建立中国西南夏季降水年际增量预测模型,后报检验 7a 的降水距平百分率的均方根误差为 16%。覃志年等^[7]研发了广西月尺度动力模式产品解释应用系统,解释应用预测准确率一般在 70 分以上。何慧等^[8]利用动力延伸预报产品制作广西月降水预报。由于短期气候预测所涉及的气候异常变化因子具有多样性和复杂性,相互作用机理的认识还远远不够等现状,困扰了短期气候预测准确性的进一步提高^[9]。近年来,人工智能算法的发展为更好的深度挖掘隐藏在海量数据中的天气气候变化的物理机制,提取更有效的预报模型提供了一种新的方式,尤其在非线性智能计算和机器学习理论的发展中,深度学习的长短期记忆神经网络(Long Short-Term Memory, LSTM)是以时间序列数据建模、数据特征提取和预测为目的,在国内外已成功地应用到水文、海洋工程、气象等领域,并取得到了较好的预报效果^[10-19]。

收稿日期: 2023-03-01

基金项目: 广西自然科学基金项目(2023GXNSFAA026414)、国家自然科学基金项目(42065004)、广西重点研发计划项目(桂科 AB21196041)、广西自然科学基金项目(2018GXNSFAA281229)

作者简介: 黄颖(1983—),女,硕士,高级工程师,主要从事天气预报技术方法研究与业务应用工作。E-mail: yinger2001@126.com

在大气科学领域,孙际钰^[14]建立了基于 LSTM 网络的海表温度预报模型。沈皓俊等^[15]利用 LSTM 网络预测 2014 年和 2015 年中国夏季降水,Ps 评分分别为 74 分、71 分。程鹏宇等^[16]、倪铮等^[17]运用 LSTM 深度神经网络建立温度预报模型。徐楠楠^[18]、张飞鸿^[19]采用 LSTM 模型分别建立日尺度降水预测模型、短临降水预测模型。这些预报模型和预报方法在预报试验中都取得了较好的预报效果。

为了进一步改进和提高月降水量预测准确性,探索新的短期气候预测方法,本文尝试以广西 81 站 7 月降水量为预报研究对象,先采用 EOF 方法对 81 站 7 月降水量作 EOF 计算,选取前几个主分量作为预报分量,再利用更适用于解决非线性问题的 LSTM 模型建立月降水量的深度学习预测模型,进行预报建模试验。对预报分量(主分量)进行预报后,再乘以空间向量便可得到广西 81 站降水量预测值,而不需要对每个站建立预报模型进行预测,同时也极大地减少了计算量和解决单站资料建模时会遇到的随机性和不稳定性问题。

1 LSTM 方法介绍

长短期记忆网络 (Long Short-Term Memory, LSTM) 是在循环神经网络 (Recurrent Neural Network, RNN) 基础上提出的一种新型深度学习神经网络模型^[20],目前已在国内外成功地应用众多学科领域^[10-11,21-24]。

由于 RNN 隐藏层只有一个 h 状态,仅对短期的输入较敏感,却较难捕获相隔较长时刻的历史数据信息,易出现梯度爆炸或消失问题,Hochreiter 和 Schmidhuber^[20]于 1997 年提出了 LSTM 模型,他们在 RNN 的隐藏层里引入了存储单元状态 (Cell State, 也称为细胞状态) 和门结构,使得网络保留长期的状态,有效解决了 RNN 中的梯度爆炸和梯度消失等问题。因此 LSTM 更适合于处理和预测时间序列中间隔和延迟相对较长的重要问题,比如短期气候预测中的月降水量预测。

LSTM 模型主要由一个或多个记忆细胞 c_t (Memory Cell)、一个输入门 i_t (Input Gate)、一个遗忘门 f_t (Forget Gate)、一个输出门 o_t (Output Gate) 组成 (见图 1)。遗忘门决定上一时刻的单元状态 c_{t-1} 有多少保留到当前时刻 c_t , 输入门决定当前时刻网络的输入 x_t 有多少保存到单元状态 c_t , 而 LSTM 的当前输出值 h_t 是由输出门和单元状态 c_t 来决定。图 1

给出了 LSTM 网络隐层结构示意图。由图 1 可以看出,当前时刻网络输入值 x_t , 上一时刻网络隐层输出值 h_{t-1} , 以及上一时刻网络单元状态输出 c_{t-1} 为 LSTM 隐层输入数据, 当前时刻网络隐层输出值 h_t 和当前时刻网络单元状态输出 c_t 为 LSTM 隐层输出数据。LSTM 通过门控单元控制信息是否通过从而实现存储单元信息的添加和删除^[25]。

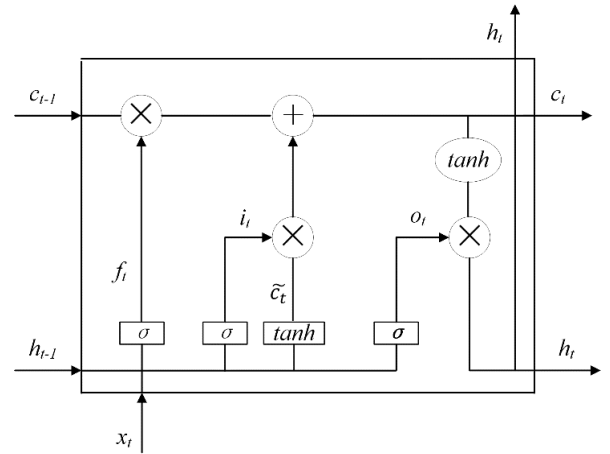


图 1 LSTM 网络隐层结构示意图

LSTM 模型主要分为 3 个阶段^[26]:

(1) 遗忘门: 遗忘门决定上一时刻的细胞状态 c_{t-1} 有多少信息保留到当前时刻。根据上一时刻网络隐层输出值 h_{t-1} 和当前网络输入值 x_t , 由激活 sigmoid 函数生成遗忘门 f_t , 产生一个 0~1 之间的数字, 来筛选上一个细胞状态 c_{t-1} , 其中 0 表示完全舍弃上一状态信息, 1 表示完全保留上一状态信息, 数字在 0 到 1 之间表示保留部分信息。遗忘门 f_t 计算公式如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

(2) 输入门: 输入门决定取多少新信息用于更新细胞状态 c_t 。根据上一个时刻的隐藏状态 h_{t-1} 和当前状态的输入 x_t , 分别由激活函数 sigmoid 生成输入门 i_t (见式(2)), 由激活函数 tanh 产生候选记忆值 (\tilde{c}_t) (见式(3))。 \tilde{c}_t 经过输入门 i_t 后得到的信息就是新加入的信息, 将上一时刻的细胞状态 c_{t-1} 经过遗忘门 f_t 得到的信息加上新加入的信息, 就得到当前细胞状态 c_t (见式(4))。计算公式如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (4)$$

由式(4)可知,LSTM 能保持长期记忆的关键在于单元状态 c_t 。因为新单元状态 c_t 的记忆与前一时刻 $t-1$ 的记忆相连接^[27],在很长一段时间内保持信息的不变性,有效防止了梯度消失或爆炸问题。

(3)输出门:输出门确定当前细胞状态 c_t 的输出。与遗忘门相似,根据上一时刻隐藏层输出值 h_{t-1} 和当前网络输入值 x_t ,由激活函数 sigmoid 生成输出门 o_t ,去筛选当前细胞状态 c_t 经过激活函数 tanh 层的结果 h_t 。输出门 o_t 和隐层的输出结果 h_t 的计算公式如下:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (6)$$

将 h_t 传送到输出层,计算预测结果 y_t ,计算公式如下

$$y_t = W_d \cdot h_t + b_d \quad (7)$$

式中, W_f, W_i, W_c, W_o, W_d 是权重矩阵, b_f, b_i, b_c, b_o, b_d 是偏置项, $[h_{t-1}, x_t]$ 表示将两个矩阵连接成一个更长的矩阵, σ 为 sigmoid 激活函数; \tanh 为双曲线正切激活函数。

LSTM 主要是求解权重矩阵和偏置项,定义交叉熵损失函数为损失函数,采用梯度下降的误差反向传播算法 (Back-Propagation Through Time, BPTT) 求出权重和偏置参数,具体推导和计算过程参考相关文献^[28]。

2 LSTM 模型的预报量和模型输入计算处理方法

2.1 预报量的预处理

本文是以广西 81 个站 7 月降水量作为预报对象,预报量资料年限是从 1960—2016 年共 57a 为预报建模样本,2017—2022 年共 6a 为独立预报样本,进行基于 LSTM 模型的月降水预测试验。考虑到如果 81 个站逐站建立预报方程,将要建立 81 个预报

模型,而对于月降水量这样的短期气候预测问题,以单站作为预报量去计算分析前期的环流场预报因子,会增加预报因子的随机性和不稳定性。为此我们考虑首先采用自然正交展开方法 (Empirical Orthogonal Function, EOF)^[29,30],先计算降水量距平率,再对 7 月 81 站降水量距平率作 EOF 计算,得出 7 月 81 站降水空间分布特征向量以及代表平均降水量年际变化的时间系数(主分量)。某时段降水量距平率计算公式如下:

$$P_a = \frac{P - \bar{P}}{\bar{P}} \times 100\% \quad (8)$$

式中, P 为某时段降水量 (单位: mm); \bar{P} 为计算时段同期气候平均降水量。本文 \bar{P} 取 1991—2020 年共 30a 平均降水量。

对 7 月 81 站降水量作自然正交展开计算时,可以分解成时间系数 Z 和空间向量 V 两部分:

$$X = VZ \quad (9)$$

由于对平均降水量作 EOF 计算分析后得到的空间向量随时间变化很小^[29],因此可以将 7 月降水量展开的时间系数(主分量)作为预报量,进行预报建模。表 1 给出广西 81 个站 7 月降水量作自然正交展开后,前 7 个主分量累积方差贡献为 76.81%。为此,本文将前 7 个主分量(时间系数)作为预报量建立预报模型,并对时间系数(主分量, Principal component, PC)进行预报后,再乘以空间向量便可得到该区域的 7 月 81 站降水量预测距平值^[29],再结合式(8),即可得到 7 月 81 站降水量预测值。这样的计算方式,使得在对广西 81 个站 7 月降水量作预报时,不需要对每个站(共 81 站)建立预报模型进行预测,只要建立 7 个预报方程,有效减小了需要建立的预报模型个数,并避免了单站资料的随机性和不稳定性。

表 1 7 月 81 站降水量前 7 个主分量展开的方差和累积方差贡献 (%)

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7
方差贡献	48.22	11.80	6.35	3.37	2.62	2.30	2.14
累积方差贡献	48.22	60.03	66.38	69.75	72.37	74.67	76.81

2.2 预报因子的计算分析

为了建立 7 月 81 站降水量 7 个预报分量的预报模型,本文以 NCEP 再分析资料的 500hPa 月高度

场作为基本的预报因子场,分别计算分析了 7 个预报分量与前期各月基本预报因子场 (上一年 1 月至 12 月,当年 1 月至 5 月)格点的相关关系。从各预报

分量与前期预报因子场成片相连的相关格点内, 挑选出相邻成片的相关系数绝对值 ≥ 0.34 (达到 0.01 相关显著性水平) 的相关格点作为一个预报因子的计算格点区。这样的计算规定, 可以保证每一个前期预报因子是由高相关 ($|r| \geq 0.34$) 的格点区域计算得出, 保证了预报因子的稳定性和高相关性。表 2 统计给出了广西 81 站 7 月降水量 7 个预报分量 (预报分量样本长度为 57) 与前期各月 500hPa 高度场相关普查计算得出的高相关 (相关系数绝对值 ≥ 0.34) 预报因子区的格点数。

表 2 广西 81 站 7 月降水量场第 1—7 主分量相关系数绝对值大于 0.34 的初选预报因子数

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
因子数	52	50	48	43	43	45	41

3 短期气候预测试验

3.1 LSTM 方法建模

以 2.2 节计算得到的广西 7 月降水量 7 个预报分量的前期初选预报因子 (表 2) 为基础, 再利用第 1 节介绍的 LSTM 建模方法, 对广西 7 月 81 个站的月降水量进行预测试验。根据预报分量的初选预报因子预报出各预报分量未来时刻的时间系数, 再乘以相应的空间向量, 并结合式 (8), 得到各个站点的月平均降水量预测值。

在广西 7 月降水量 LSTM 预测模型的实际预测建模计算过程中, 需要分别对第 1—7 个主分量 (预报分量) 进行逐年的预报, 以第 1 个主分量 (预报分量) 为例进行说明。预报分量的建模样本长度为 1960—2016 年, 共 57a, 2017—2022 年共 6a 作为独立预报样本, 进行预测试验。由表 2 可以看到第 1 个主分量 (预报分量) 有 52 个预报因子, 面对众多的初选预报因子, 首先采用逐步回归方法, 取 $F=1$, 从 52 个预报因子中, 自动筛选出 15 个左右的预报因子作为 LSTM 预报模型的输入因子进行预报建模。

为了保证预报建模的客观性, 在建立 LSTM 预报模型时, 各项参数统一设定为输入节点为模型输入的因子个数, 输出节点为 1, 网络层数为 1, 隐节点数为 30, 训练次数为 300, 总体收敛误差为 10^{-6} 。图 1 给出了采用 LSTM 预测模型, 对 7 月降水量 2017 年第 1 主分量预测计算时, LSTM 对历史样本的拟合结果和实况值的对比结果。由图 1 可以看

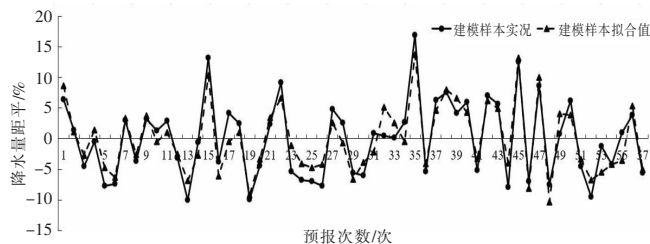


图 1 7 月 81 站第 1 主分量历史样本拟合图

出, LSTM 模型对历史样本的拟合情况还是较为接近的。

利用该预报模型可以对 7 月 81 站降水量 2017 年第 1 主分量进行预报计算, 得出第 58 个样本的主分量预测值。依据上述相同的计算步骤, 并保持 LSTM 网络模型的各项参数全部不变, 将 2017 年的实况值加入建模样本, 再对广西 81 站 7 月降水量 2018 年的第 1 主分量进行预测计算, 以此类推, 可以继续逐次计算 81 站 2019—2022 年第 1 主分量的预测值。再采用上述计算步骤, 对广西 81 站 7 月降水量的第 2—7 主分量进行了 2017—2022 年逐年的 LSTM 预测计算。根据降水量预测场自然正交展开后, 空间向量随时间变化较小的特点, 将 LSTM 预测模型逐年预测得到的 7 个主分量 2017—2022 年的预测值与相应的空间向量合成, 得到广西 81 个站 2017—2022 年每年各站 7 月降水量的预测值 (见表 3)。

为了定量评价 LSTM 预报模型对 2017—2022 年广西 81 个站 7 月降水量的趋势预报准确率, 选取的评价指标有平均绝对误差 (Mean Absolute Error, MAE) 和 Ps 评分。计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$Ps = \frac{a \times N_0 + b \times N_1 + c \times N_2}{(N - N_0) + a \times N_0 + b \times N_1 + c \times N_2 + M} \times 100\% \quad (11)$$

式中, y_i 为降水量实况值, \hat{y}_i 为降水量预测值, N_0 为趋势预测正确站数, N_1 为 1 级异常预测正确站数, N_2 为 2 级异常预测正确站数, M 为 2 级异常漏报站数, N 为参加考核总站数, a, b, c 为权重系数, 分别取 $a=2, b=2, c=4$ 。由表 3 的统计结果得出, 广西 81 个站 2017—2022 年 7 月降水量预报结果的 Ps 平均得分为 75.95, 降水量平均绝对误差为 68.77, 6 年中, 历史降水样本评分达到 70 分以上的站数平均百分比达到 56.61%, 距平率平均绝对误差为 23.06, 平均 Ps 评分为 74.56。

表 3 LSTM 模型对广西 81 个站 7 月降水预报拟合样本、预报样本的平均绝对误差和 Ps 评分统计

年份	拟合结果			预测结果	
	距平率平均绝对误差	Ps 评分	Ps≥70 分百分比/%	降水量平均绝对误差	Ps 评分
2017	22.81	75.33	57.89	59.28	68.67
2018	22.69	75.04	60.34	97.03	41.74
2019	23.31	73.67	55.93	68.36	72.83
2020	23.17	73.90	51.67	50.41	86.69
2021	22.78	75.20	57.38	84.04	84.56
2022	23.57	74.22	56.45	53.47	78.50
平均	23.06	74.56	56.61	68.77	75.95

从表 3 可以看出,总体上广西 7 月 81 个站的降水量预测值与实况总体趋势上是相近的,其中 5a 的预报 Ps 评分在 68 分以上,最高可以达到 86.69 分,显示出 LSTM 模型的预测稳定性。

3.2 逐步回归方法建模

由于深度学习的 LSTM 模型在数学建模方法进行了新的尝试,因此,有必要对该预报方法与传统的一些预报方法进行预报性能的对比分析。为了能使预报效果有较好的客观的对比分析,利用与 3.1 节 7

月降水 7 个预报分量 LSTM 预报模型完全相同的预报因子建立了 7 个相应的逐步回归预报分量预报模型(见表 1 和表 2)。并利用这 7 个预报分量的逐步回归预测模型进行了 2017—2022 年逐年的主分量预测,再将主分量预测值与相应的空间向量合成,得到广西 7 月 81 个站 2017—2022 年各站的月降水量平均预测值。同样也利用式(10)~式(11)计算了 81 个站月降水量预测的 Ps 评分,结果见表 4。

表 4 逐步回归模型对广西 81 个站 7 月降水预报拟合样本、预报样本的平均绝对误差及 Ps 评分统计

年份	拟合结果			预测结果	
	距平率平均绝对误差	Ps 评分	Ps≥70 分百分比/%	降水量平均绝对误差	Ps 评分
2017	23.20	75.63	56.14	60.21	55.38
2018	23.26	75.47	58.62	89.49	34.62
2019	23.62	74.51	55.93	71.45	69.41
2020	23.38	74.89	55.00	57.06	81.70
2021	23.26	75.34	59.02	91.50	87.11
2022	23.70	74.31	50.00	54.86	73.47
平均	23.40	75.03	55.79	70.76	72.37

将表 3 与表 4 的预报 Ps 评分结果进行对比分析可以清楚看到,虽然 LSTM 预报模型与逐步回归方程中的预报因子是相同的,但是采用 LSTM 网络的预报方法,预报精度明显高于传统的回归预报方程。在 81 个站 6a 独立样本预报中,LSTM 预报模型的 Ps 评分基本高于逐步回归方法,就平均而言 LSTM 预报模型 6a 的 Ps 评分平均为 75.95 分,高于逐步回归方法的 72.37 分;LSTM 模型的降水量平均绝对误差为 68.77,低于逐步回归方法的 70.76。由于两种方法依据的预报因子是完全相同的,因此导致预报效果的好坏显然是与预报建模的数学建模方法

有关。由于 LSTM 方法激励函数是一种非线性函数,可能比线性回归方法更适用于月降水的短期气候预测问题。

4 结论

本文利用 EOF 展开和 LSTM 网络相结合的方法构建了一种新的非线性智能计算短期气候预测预报模型,该预报建模方法的主要特点和结论如下:

(1)深度学习的 LSTM 模型的隐层里引入了存储单元状态和门结构,使得网络保留长期的状态,有效解决了 RNN 中的梯度爆炸和梯度消失等问题,也

能更深入挖掘深层次数据特征信息。因此 LSTM 更适合于处理和预测时间序列中间隔和延迟相对较长的重要问题, 适用于其它中短期天气的非线性统计预报和数值预报产品释预报等客观业务预报和研究工作。

(2)在预报量的计算处理上,将多站点的月降水量预报对象,通过采用 EOF 方法提取前几个主分量作为预报量,可以有效减少需要建立的预报模型个数,以及避免以单站降水作为预报对象容易造成预报不稳定,随机性增大的可能。同时,在预报因子的计算处理上,对前期月高度场的大量成片高相关区提取的预报因子,进行逐步回归方法的因子筛选的计算处理,减小神经网络的网络结构规模和预报因子间的复共线性影响,有利于提高预报因子的预报能力。

(3)通过对 LSTM 预报模型与传统回归分析方法的预报试验结果表明,这种新的非线性智能计算预报方法,对 2017—2022 年广西 7 月 81 个站月降水预报结果表明,在预报因子完全相同的情况下,预报准确率和预报稳定性都明显好于传统的线性回归分析方法,显示了很好的应用推广前景。

参考文献:

- [1] 丁一汇. 我国短期气候预测业务系统[J]. 气象, 2004, 30(12): 11-16.
- [2] 李维京, 张培群, 李清泉, 等. 动力气候模式预测系统业务化及其应用[J]. 应用气象学报, 2005(增刊 1): 1-11.
- [3] SCHEPEN A, WANG Q J, ROBERTSON D E. Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall [J]. Journal of Geophysical Research: Atmospheres, 2012, 117(D20).
- [4] WEI F Y, XIE Y, MANN M E. Probabilistic trend of anomalous summer rainfall in Beijing: role of interdecadal variability [J]. Journal of Geophysical Research: Atmospheres, 2008, 113(D20).
- [5] 杜良敏, 柯宗建, 刘长征, 等. 基于聚类分区的中国夏季降水预测模型[J]. 气象, 2016, 42(1): 89-96.
- [6] 郑然, 刘嘉慧敏, 马振峰. 年际增量方法在西南夏季降水预测中的应用[J]. 气象学报, 2019, 77(3): 489-496.
- [7] 覃志年, 陈丽娟, 唐红玉, 等. 月尺度动力模式产品解释应用系统及预测技巧[J]. 应用气象学报, 2010, 21(5): 614-620.
- [8] 何慧, 金龙, 覃志年, 等. 动力延伸预报产品在广西月降水预报中的应用[J]. 应用气象学报, 2007(5): 727-731.
- [9] 何慧, 陆虹, 覃卫坚, 等. 人工神经网络在月降水量预测业务中的研究和应用综述[J]. 气象研究与应用, 2021, 42(1): 1-6.
- [10] KARIMA N, BOUDHAR A, OUATIKI H, et al. Deep learning approach with LSTM for daily streamflow prediction in a semi-arid area: a case study of Oum Er-Rbia River Basin, Morocco [J]. Water, 2023, 15(2): 262.
- [11] FAN G J, YU P Y, WANG Q, et al. Short-term motion prediction of a semi-submersible by combining LSTM neural network and different signal decomposition methods [J]. Ocean Engineering, 2023.
- [12] LIU H J, LEI D X, YUAN J, et al. Ionospheric TEC prediction in China based on the multiple-attention LSTM model [J]. Atmosphere, 2022, 13(11): 1939.
- [13] THANONGSAK X, Lee H, Lee G. Air pollution prediction using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models [J]. Sustainability, 2020, 12(6): 2570.
- [14] 孙际钰. 基于 LSTM 网络的海表温度预报技术研究[D]. 大连: 大连海洋大学, 2022.
- [15] 沈皓俊, 罗勇, 赵宗慈, 等. 基于 LSTM 网络的中国夏季降水预测研究[J]. 气候变化研究进展, 2020, 16(3): 263-275.
- [16] 程鹏宇, 赵嘉, 韩龙哲, 等. 双向多尺度 LSTM 的短时温度预测[J]. 江西师范大学学报(自然科学版), 2022, 46(2): 134-139.
- [17] 倪铮, 梁萍. 基于 LSTM 深度神经网络的精细化气温预报初探[J]. 计算机应用与软件, 2018, 35(11): 233-236.
- [18] 徐楠楠. 基于 LSTM 的中国大陆地区 ERA5 日尺度降水预测方法研究[D]. 南京: 南京邮电大学, 2021.
- [19] 张飞鸿. 基于对抗型 LSTM 和时序残差网络的短临降水预测研究[D]. 南京: 南京信息工程大学, 2020.
- [20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [21] 陈剑飞, 李勇, 刘俊江, 等. 耦合陆面水文模型和机器学习方法的水库径流量预报及应用[J]. 气象研究与应用, 2022, 43(1): 1-7.
- [22] 黄启桥, 麦雄发, 李玲, 等. 基于 ConvLSTM 的广西短临降水预报[J]. 气象研究与应用, 2021, 42(4): 44-49.
- [23] 梁振清, 陈生. 基于深度学习和雷达观测的华南短临预报精度评估[J]. 气象研究与应用, 2020, 41(1): 41-47.
- [24] 张明, 杜裕, 洪国平. LSTM 在宜昌短期电力负荷预测中的试验应用[J]. 气象研究与应用, 2019, 40(3): 72-77.
- [25] FELLX A G, JURGEN S, FRED C. Learning to forget: continual prediction with LSTM [C]. International Conference on Artificial Neural Networks, Edinburgh, UK, 1999(2): 850-855.

- [26] 张进, 牛志伟, 齐慧君, 等. K-means++和 LSTM 模型在大坝监测资料分析中的应用[J]. 三峡大学学报(自然科学版), 2021, 43(4): 25–30.
- [27] GERS F A, SCHMIDHUBER E. LSTM recurrent networks learn simple context-free and context-sensitive languages [J]. IEEE Transactions on Neural Networks, 2001, 12(6): 1333–1340.
- [28] GRAVES A. Supervised sequence labelling with recurrent neural networks [C]. Springer-Verlag Berlin Heidelberg, 2012.
- [29] 吴洪宝, 吴蕾. 气候变率诊断和预测方法[M]. 北京: 气象出版社, 2005.
- [30] 覃卫坚, 苗春生, 陈丽, 等. EOF 在广西春季降水分析中的应用[J]. 气象研究与应用, 2007(2): 42–45.

Study on monthly precipitation prediction model in Guangxi based on EOF and LSTM

Huang Ying¹, Lu Hong², Huang Xiaoyan¹, Zhao Huasheng¹, Wu Yushang³

(1. Guangxi Institute of Meteorological Sciences, Nanning 530022, China;

2. Guangxi Climate Center, Nanning 530022, China;

3. Guangxi Meteorological Observatory, Nanning 530022, China)

Abstract: In view of the characteristics of time dependence and nonlinear changes of precipitation weather process in summer, and the essential characteristics of monthly precipitation that are failed to fully obtain from the existing forecasting methods, it is difficult to deal with modeling factors and forecast modeling. To solve this problem, a monthly precipitation prediction model based on the Empirical Orthogonal Function (EOF) and short-term memory neural network (LSTM) of deep learning has been proposed in this paper. Taking the July precipitation of 81 stations in Guangxi as the forecast object, EOF is calculated for the July precipitation of 81 stations, and the first seven principal components with cumulative variance contribution of more than 76% are selected as the forecast components. Then a deep learning prediction model has been established using LSTM model for monthly precipitation, with the July precipitation at 81 stations from 1960 to 2016 as the modeling samples and 2017–2022 as independent samples. Results show that under the same modeling samples and factors, the newly established forecast model has a higher forecast ability than the linear stepwise regression forecast method, demonstrating its applicability to the nonlinear monthly precipitation forecast problem. Further analysis shows that the LSTM model introduces the storage cell state and gate structure in the hidden layer, which allows the network to retain long-term states, making it more suitable for processing and predicting the important problems with relatively long interval and delay in the time series.

Key words: Monthly precipitation prediction; long and short-term memory neural network; empirical orthogonal function; deep learning