

基于逻辑回归 - 径向基神经网络的广西前汛期降水预测

蒙芳秀, 苏健昌*, 覃碧莉, 吴俊皇, 蒋宜蓉

桂林理工大学, 广西 桂林

收稿日期: 2022年11月7日; 录用日期: 2022年12月1日; 发布日期: 2022年12月9日

摘要

基于2000~2020年4~6月广西5个代表站地面气象观测站降水量, 建立基于逻辑回归 - 径向基神经网络(logical regression-radial basis function neural network, 简称LR-RBF)的广西前汛期降水预测模型。结果表明, 基于LR-RBF预测广西前汛期降水效果较好, 实测降水与预测降水的R值均高于0.91, 呈现高度相关。预测精度较径向基神经网络(Radial Basis Function, RBF) MAE值最大减少为45.9%, 而RMSE值最大减少为35.71%, 特别是贺州站; 与逐步回归相比, R值最高增大13.09%, MAE最大减少26.91%, RMSE最多减少23.10%。结果表明, LR-RBF预测能力有显著的提升, 对广西前汛期防洪防控工作具有一定的指导价值。

关键词

逻辑回归, 径向基神经网络, 降水预测, 前汛期

Prediction of Precipitation during the Pre-Rainy Season in Guangxi Province Based on Logistic Logical Regression-Radial Basis Function Neural Network

Fangxiu Meng, Jianchang Su*, Bili Qin, Junhuag Wu, Yirong Jiang

Guilin University of Technology, Guilin Guangxi

Received: Nov. 7th, 2022; accepted: Dec. 1st, 2022; published: Dec. 9th, 2022

*通讯作者。

文章引用: 蒙芳秀, 苏健昌, 覃碧莉, 吴俊皇, 蒋宜蓉. 基于逻辑回归-径向基神经网络的广西前汛期降水预测[J]. 应用数学进展, 2022, 11(12): 8540-8549. DOI: [10.12677/aam.2022.1112901](https://doi.org/10.12677/aam.2022.1112901)

Abstract

Based on the daily precipitation data of 5 representative stations in Guangxi Province from April to June 2000 to 2020, a logistic regression-radial basis function neural network method (LR-RBF) is established to prediction precipitation during the pre-rainy season. From the results observed, the LR-RBF method has a good effect on modeling precipitation during the pre-rainy season in Guangxi Province. For instance, the R value between observed precipitation and predicted precipitation is higher than 0.91, showing a high correlation. Compared with the radial basis function neural network (RBF), from the prediction accuracy, MAE decreased by 45.9%, and the reduction of RMSE is 35.71%, especially at Hezhou station. Compared with the stepwise regression, the increase of R value is 13.09%. Meanwhile, the decrease of MAE is 26.91%, and the decrease of RMSE is 23.10%. The results show that the prediction ability of the LR-RBF method has been significantly improved, which provides a reference for flood control in Guangxi Province during the pre-rainy season.

Keywords

Logistic Regression, Radial Basis Function Neural Network, Precipitation Prediction, Pre-Rainy Season

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

广西位于亚热带季风湿润区，为我国降水量最充沛的地区之一。其中前汛期(4~6月)降水量约为全年总量的40%~50%，暴雨日数也占全年暴雨日数的50%以上。前汛期是广西汛期降水的主要时段，也是洪涝灾害高发时段。因此研究该时段的降水预报，为各级政府防灾减灾提供参考，显得尤为重要[1][2][3]。

目前我国降水预报主要分为三类方法：统计建模、数值模拟和数值统计相结合。数值模拟优势在于能够很好的解释对降水的大气物理机制，但是计算复杂，定量分析精度稍差。而统计建模计算流程相对简单，并且定量模拟效果较好[4]。已有的降水量统计建模方法较多，例如：顾伟宗等[5]和赵中军等[6]分别采用最优回归方法和偏最小二乘回归方法建立降水量预测模型，预测效果远高于模式预测效果。刘婷婷等[7]基于变形的典型相关分析建立我国东部夏季降水集合预报模型，预报技巧较高。陈潇潇等[8]基于指数自回归预测福建平潭降水量，表明外强迫因子有效提高预报能力。这些方法都需要明确降水与预测因子之间明确的函数关系(线性或者非线性)，但是降水影响因素复杂，有很强的时空变异性，上述传统模型拟合效果依然欠佳，需要引入现代统计方法来改进。

神经网络因其具有自组织、自学习、非线性映射等优点被广泛应用于天气预报中[9][10]。其中径向基神经网络(radial basis function neural network，简称RBF)因其计算简单、学习速度快、分类能力强、预测精度高等优点，被广泛用于像降水这样复杂的时间序列预测中。沈艳等[11]、季刚等[12]将RBF神经网络模型预测月降水量，并与BP神经网络进行对比，其预测精度显著优于BP神经网络。农吉夫等[13]、李洁[14]改进RBF，建立PCA-RBF预测广西降水量，并与模式资料进行对比，预测精度更高。张飞涟等[15]、蒋林利[16]、罗芳琼[17]建立了改进RBF的集合预报模型，取得了较好的预报效果。以往的研究都

是直接对降水时间序列进行预测，没有区分有无降水。然而降水时间序列中存在很多无降水值，即降水量为 0，极大影响了模型模拟效果。有必要区分有无降水，再进行降水量预测。本文首先基于逻辑回归(logistic regression，简称 LR)发生概率模型，区分无降水日和有降水日，再将有降水日输入 RBF 神经网络，建立逻辑回归 - 径向基神经网络(logistic regression-radial basis function neural network，简称 LR-RBF)广西逐日降水量预报模型。并与单纯的 RBF 神经网络、逐步回归方法进行对比。

2. 研究区域及数据

根据不同地理位置、地形特征、降水类型等特征，本文选取了桂林站、贺州站、来宾站、百色站、北海站 5 个站点作为广西代表站，各站点的基本情况及气候类型见表 1。基于其 2000~2020 年前汛期(4~6 月)逐日降水数据，以及各种气候指数数据，研究 LR-RBF 降水预报模型的适用性。其中选取 2000~2018 年数据用做统计建模，2019~2020 年数据作为评价模型预报。

Table 1. The basic conditions and climate types of the stations in Guangxi Province
表 1. 广西各站点的基本情况和气候类型

区站号	站点	经度(°E)	纬度(°N)	海拔高度(m)	最大降雨量(mm)	平均降雨量(mm)
57957	桂林站	110.18	25.19	164.4	135	11.16989
59065	贺州站	111.3	24.25	148.7	249.4	12.67521
59242	来宾	109.14	23.45	84.9	243.6	17.66798
59211	百色站	106.36	23.54	173.5	123.8	11.78008
59644	北海站	109.08	21.27	12.8	138.1	13.00815

3. 研究方法

3.1. 逻辑回归方法

逻辑回归是一种常用的判别分析模型[18]，被广泛用于各种分类问题尤其二分类问题中。有无降水预报可以看成就是一种二分类问题。建立逻辑回归降水发生模型，其中对降水产生影响的因子视为分类指标，有无降水视为分类预报结果[19]。预报因子对降水为非线性影响，而逻辑回归判别模型对这类问题的预报效果更好[20]。因此可利用逻辑回归建立降水发生模型。

逻辑回归降水发生模型的计算公式为：

$$y = \frac{1}{1 + e^{-z}} \quad (1)$$

其中 y 表示逻辑回归降水发生的概率值， y 大于等于概率阈值 0.5，表示模型判断有降水发生，否则无降水发生。其中 z 是分类指标的线性组合，表示为：

$$z = \theta^T x = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (2)$$

其中， α 为常数， β 为逻辑回归系数， x 为分类指标。

将(2)式带入(1)式可得：

$$y = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}} \quad (3)$$

更具体，将逻辑回归模型任意预测结果 y 的值映射到 [0, 1] 区间内，当 y 值越靠近 1 说明降水发生的

概率越大。因此,当 $y < 0.5$ 时,输出结果为0;当 $y \geq 0.5$ 时,输出结果为1,其中,0表示无降水日,1表示有降水日。可见,利用逻辑回归模型对降水进行预报,可通过计算 y 的概率值预测其发生降水的可能性。

3.2. 径向基神经网络

径向基神经网络是一种非线性局部逼近的神经网络,于1990年代由Moody和Darken提出。RBF神经网络计算量小,学习速度快,比通常的BP神经网络要快 $10^3\sim 10^4$ 倍[11][12][13]。

RBF神经网络是一种具有三层网络结构的前馈型网络结构,其结构主要分为输入层、隐含层和输出层(图1)。

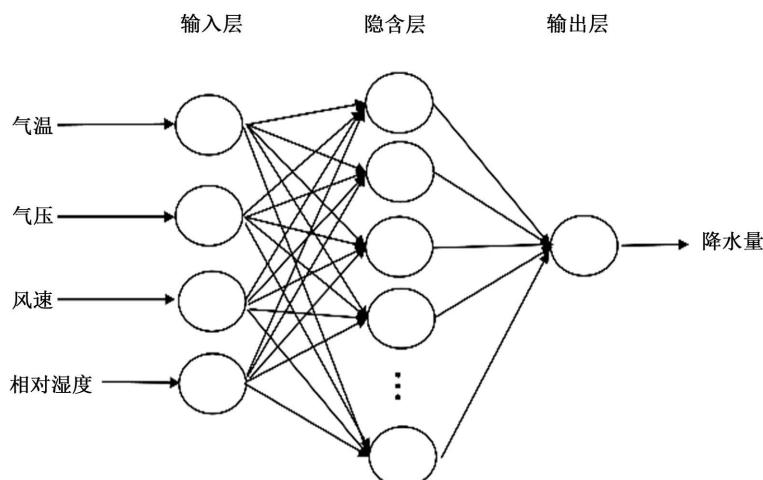


Figure 1. RBF neural network topology structure
图 1. RBF 神经网络拓扑结构

输入层将地面常规因子:气温、气压、风速、相对湿度输入,通过隐含层到达输出层,输出降水量值。输入层到隐含层非线性变化,从隐含层到输出层线性变化,建立RBF神经网络降水预报训练模型。其中隐含层由激活函数构成,激活函数形式很多,通常采用高斯函数[21]。高斯函数主要公式如下:

$$R_i(x) = \exp\left[-\frac{\|X - C_j\|^2}{2\sigma_j^2}\right], j = 1, 2, 3, \dots, k \quad (4)$$

其中, X 为输入向量;气温、气压、风速、相对湿度; C_j 为第 j 个基函数的中心; $\|X - C_j\|$ 为范数,表示输入向量 X 与基函数中心 C_j 的距离; σ_j 为基函数中心点宽度的标准化常数; K 是隐含层节点数,为输入样本个数。

3.3. LR-RBF 预报模型

通过将逻辑回归二分类与RBF神经网络相结合,将降水数据划分为有降水日数据和无降水日数据,并将有降水数据代入RBF神经网络模型进行降水量预测,所建立LR-RBF预报模型[22]的步骤如下:

- 1) 以地面常规气温、气压、风速、相对湿度作为分类指标,建立逻辑回归降水发生模型,将降水数据集分为无降水和有降水,并将有降水日数据输入RBF神经网络降水预测模型;
- 2) 将气温、气压、风速、相对湿度、有降水数据进行归一化处理;
- 3) 选取高斯函数作为激活函数,基于K均值聚类选择激活函数的中心,隐含层至输出层之间的神经元的连接权值用最小二乘法直接计算;

4) 直到达到预定误差要求或者设定的最大隐含层神经元数为止, 输出预测值。再将其反归一化还原, 即为所预测降水量。

3.4. 逐步回归

逐步回归基本思想是利用相关显著性检验系数依次剔除无显著意义的自变量[23], 即逐次引入自变量, 每次引入对 y 影响显著的自变量, 并对方程中的已存在的变量逐个进行检验, 把不显著的变量逐个从方程中剔除, 重新建立不包含该因子的多元回归方程。通过反复的检验, 筛选出有显著影响的因子作为自变量, 并建立“最优”回归方程。预报模型为:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (5)$$

其中, y 表示所要预报的降水量, x_n 为预报因子: 气温、气压、风速、相对湿度, β_1 为常数项系数, β_n 为回归系数。

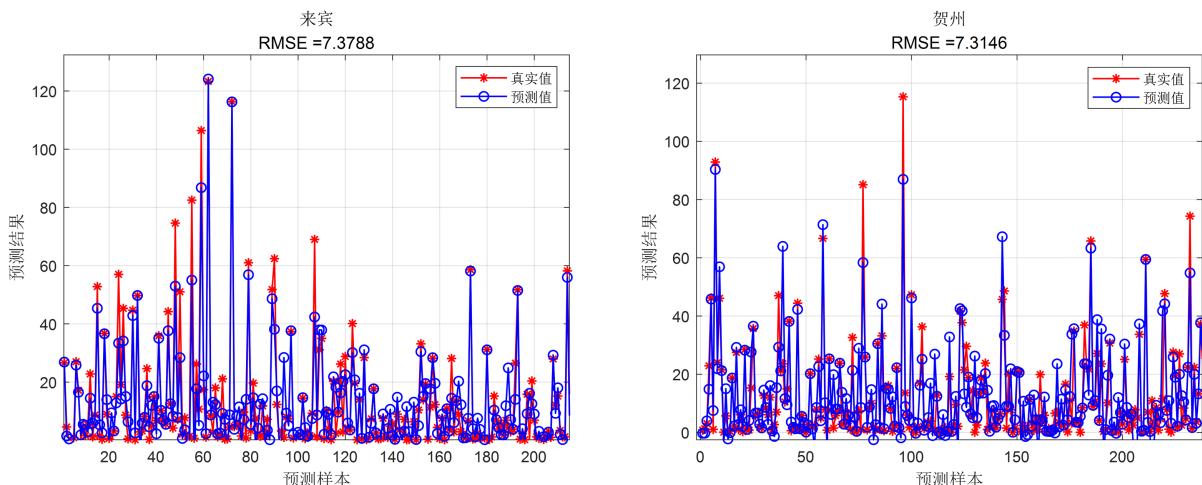
3.5. 评价指标

选取相关系数 R、平均绝对误差(mean absolutely error, MAE)、均方根误差(root mean square error, RMSE)作为模型评价指标[3] [24] [25]。其中 MAE 和 RMSE 越接近于 0, 表明预测值与实测值的误差越小; R 越接近于 1, 表明预测值和实测值的相关性越强, 模型预测能力越好。

4. 结果分析

4.1. RBF 神经网络降水预测结果

利用单纯的 RBF 神经网络建立预测模型, 对桂林、贺州、来宾、百色、北海等五个代表站点的前汛期降水进行预报, 取 2000~2018 年降水数据作为训练集, 2019~2020 年降水数据作为测试集, 所得五个站点的拟合结果如图 2 所示, 五个站点的模型评价指标如表 2 所示。结果表 2 与图 2 可知, 桂林站降水预测值和观测值相关性最强, R 值为 0.86, 最低为百色站降水预测值和观测值相关性最弱, R 值为 0.81。MAE 和 RMSE 都衡量预测值与实测值之间的误差, 百色站降水预测值和观测值之间的误差 MAE 最小, 为 3.93, 贺州站降水预测值和观测值之间的误差 RMSE 最小, 为 7.31。而北海站降水预测值和观测值之间的误差 MAE, RMSE 都是最大, 这与北海受台风等因素影响有关, 仅考虑地面常规因子的影响是不够的。



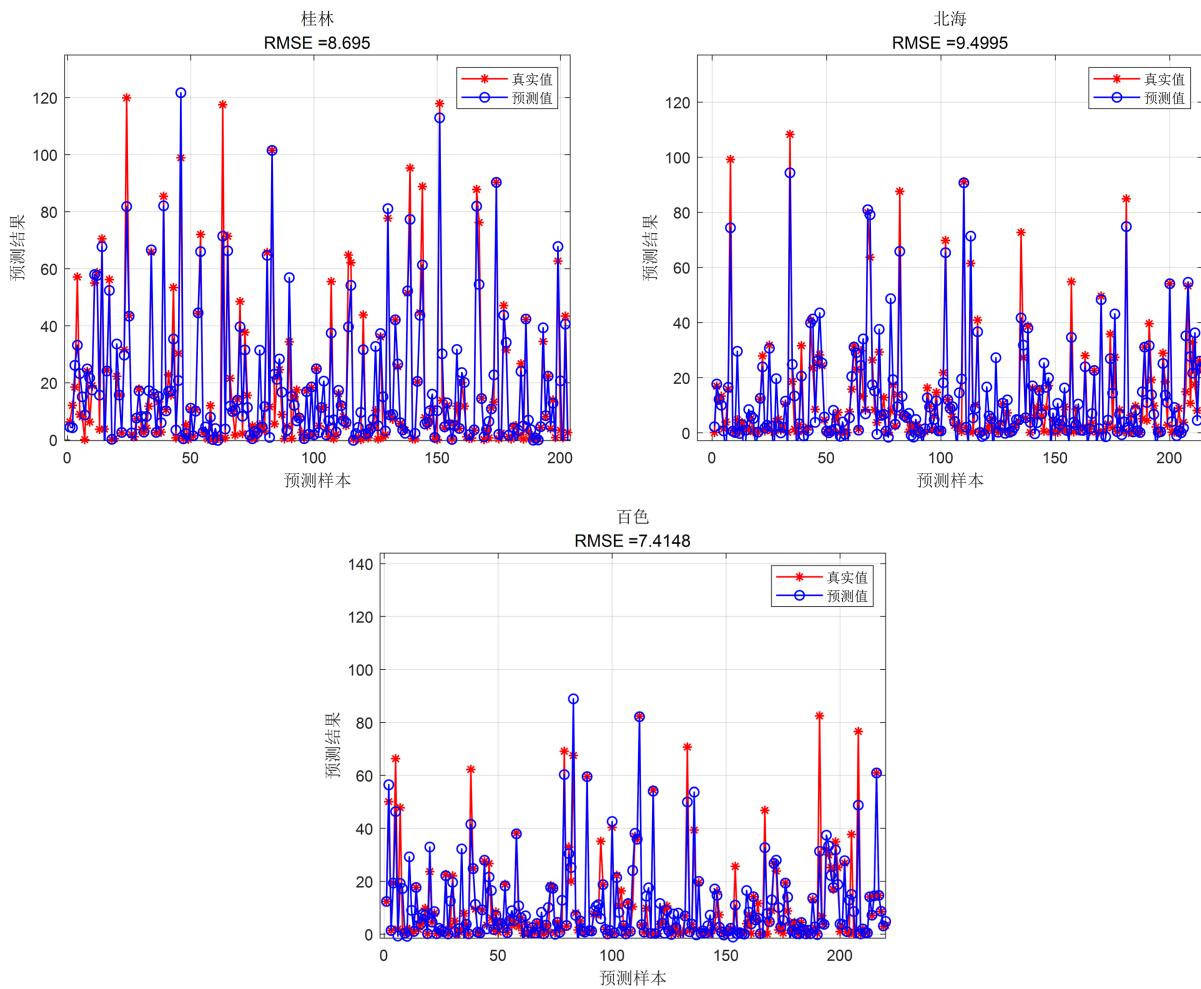


Figure 2. Comparison between the predicted value and the measured value of RBF Neural Network
图 2. RBF 神经网络预测值与实测值对比

Table 2. The precipitation forecast evaluations of RBF, LR-RBF, stepwise regression during the pre-rainy season in Guangxi Province

表 2. RBF 神经网络、LR-RBF 神经网络、逐步回归广西前汛期降水预测评估

站点	LR-RBF			RBF 神经网络			逐步回归		
	R	MAE	RMSE	R	MAE	RMSE	R	MAE	RMSE
来宾站	0.95	2.77	5.61	0.82	4.38	7.38	0.84	3.79	7.11
贺州站	0.93	1.92	4.80	0.82	4.20	7.31	0.89	2.30	6.31
桂林站	0.95	2.37	6.15	0.86	5.03	8.70	0.90	3.16	7.19
北海站	0.91	4.25	9.01	0.82	5.57	9.50	0.86	4.72	9.24
百色站	0.92	2.40	5.44	0.81	3.93	7.41	0.87	3.17	6.74

4.2. LR-RBF 神经网络降水预测结果

利用 LR-RBF 预测广西五个代表站前汛期降水，所得五个站点的拟合结果如图 3 所示，五个站点的模型评价指标如表 2 所示。结合表 2、图 3 可知，对于 R 值，来宾站与桂林站 R 值都为 0.95，降水预测

值和观测值呈现高度相关。从模拟精度而言，贺州站降水预测值和观测值的误差最小，MAE 和 RMSE 都是最低，分别为 1.92 和 4.80。北海站降水预测值和观测值的三个指数都是最小的，拟合效果为五个站点中最差的，这与北海的地理位置有关，考虑影响因素应该更全面，是下一步改进的方向。

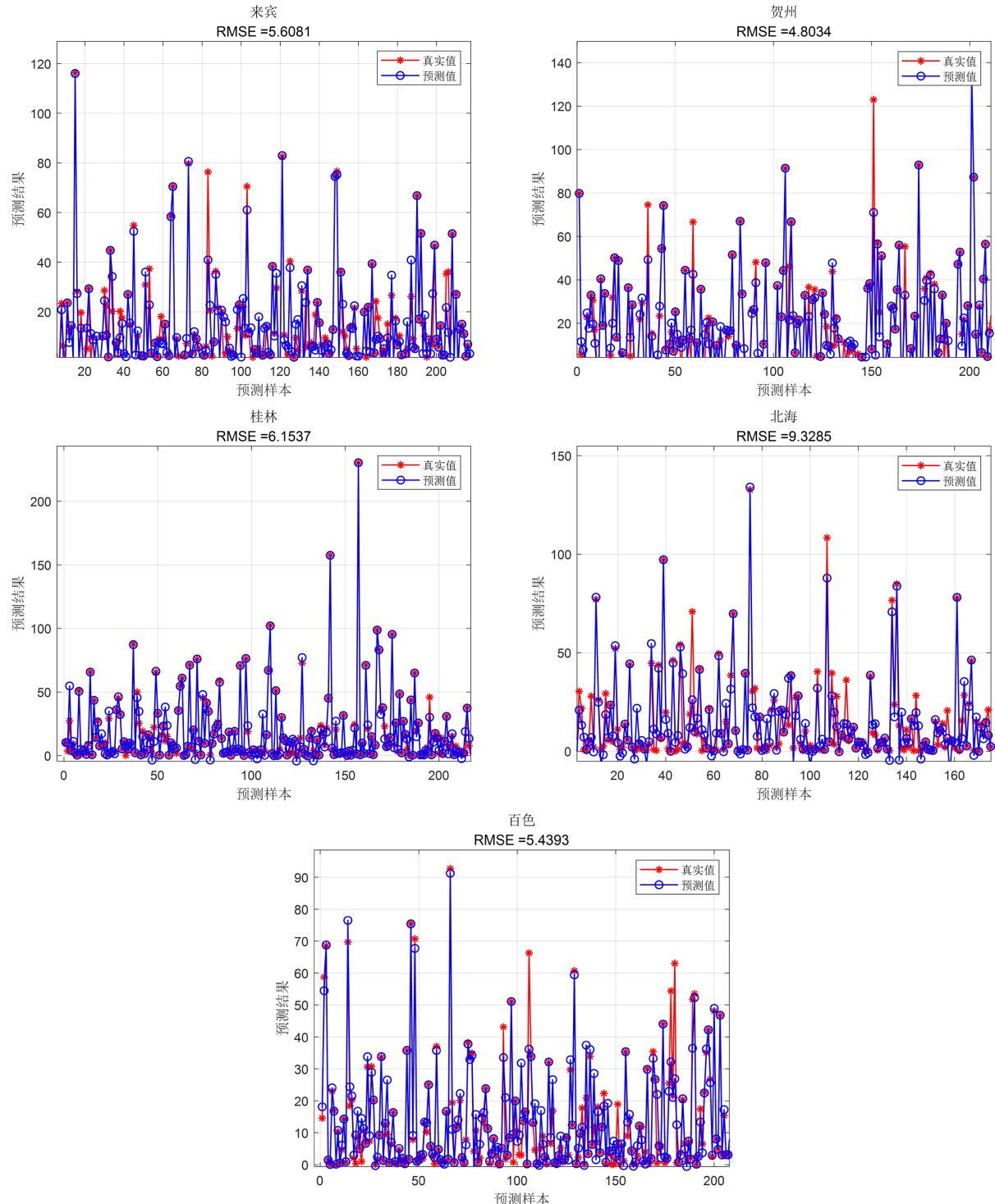


Figure 3. Comparison between the predicted value and the measured value of LR-RBF method
图 3. LR-RBF 预测值与实测值对比

4.3. 模型预测精度对比分析

结合图 3 和图 4 可知, 两种神经网络的预测值与实测值重合度较高, 表明 RBF 神经网络对降水预测效果较好。具体而言, 基于 LR-RBF 降水预测模型较单纯的径向基神经网络模拟效果更好, 特别是对贺州站, 其 RMSE 最低为 4.80, 最大减少为 34.34%。而两种模型对北海站降水预测效果最差, 这与北海的地理位置有关, 北海降水常受台风等因素影响, 仅考虑地面常规气象要素是不够的。

进一步, 将两种神经网络方法与传统的统计建模方法 - 逐步回归做比较。分别计算 RBF 神经网络模型、LR-RBF 降水预测模型、逐步回归预测能力评价指标, 结果如表 2。并画出 RBF 神经网络、LR-RBF、逐步回归各站点降水预测值与真实值的平均相对误差如图 4。结合表 2 和图 4 可知, LR-RBF 降水预测模型预测能力较 RBF 神经网络模型有显著地提升, 具体表现为其预测精度较 RBF 神经网络 MAE 值最大减少为 54.29%, 而 RMSE 值最大减少为 34.34%, 特别是贺州站。相对误差也有显著减少, 其中减少最多为来宾站降水预测, 减少了 10.26%。而实测降水与预测降水的 R 值均高于 0.91, 呈现高度相关, 特别是桂林站和来宾站降水预测值和实测值 R 值达到 0.95。与逐步回归相比, LR-RBF 的 R 值最高增大 13.09%, MAE 最多减少 26.91%, RMSE 最多减少 23.10%。

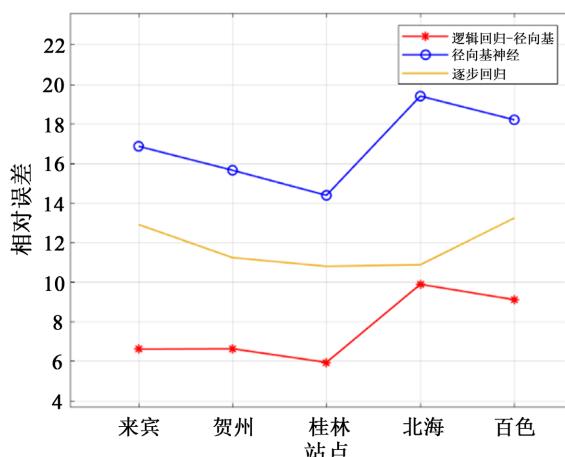


Figure 4. The relative errors of prediction of precipitation using RBF, LR-RBF, Stepwise regression
图 4. RBF 神经网络、LR-RBF、逐步回归降水预测的相对误差

4.4. 模型在降水预测中适用性对比分析

上面展示了 RBF 神经网络、LR-RBF 神经网络、逐步回归在降水预测中的应用, 以下总结一下三种方法在降水预测中的技术特点如表 3。三种方法在降水预测中都得到广泛的应用, 但是未分类的降水量数据中 0 值较多时, 对 RBF 神经网络的预测精度及对逐步回归的参数估计等都产生较大影响。而 LR-RBF 神经网络先通过 LR 区分降水量, 将有降水量输入 RBF 网络进行降水预测, 较好提升预测能力。

Table 3. The applicability analysis of RBF, LR-RBF, stepwise regression in precipitation prediction

表 3. RBF 神经网络、LR-RBF、逐步回归在降水预测的适用性对比

方法	优点	缺点
RBF 神经网络	结构简单、收敛速度快、预测精度高	降水数据 0 值较多, 影响预测精度和运行速度
LR-RBF 神经网络	消除降水 0 值的影响、运行速度更快、预测精度更高	将降水量数据进行分类, 分类模型的精度会影响预测模型的精度, 需要寻找一个分类效果更好的分类器

Continued

逐步回归	消除多重变量共线的变量的影响	降水数据 0 值较多, 影响参数估计及预测精度且当两个变量对因变量的影响相近时, 易受噪声影响, 效果不稳定
------	----------------	--

5. 结论与讨论

以地面常规气温、气压、风速、相对湿度作为预测因子, 建立基于 LR-RBF 的广西前汛期降水预测模型。并与单纯的 RBF 神经网络、逐步回归预报结果作对比, 验证组合模型的优越性。

对 2019~2020 年 4~6 月广西代表站点降水进行预测, 结果表明基于 LR-RBF 的广西前汛期降水预测精度显著优于 RBF 神经网络和逐步回归预报结果。其中, 桂林站和来宾站降水预测值和实测值呈现高度相关, R 值相关性最大达到 0.95。特别地, 相对于 RBF 神经网络预测结果, 对贺州站降水预测精度提升最为明显, MAE 值最大减少为 54.29%, 而 RMSE 值最大减少为 34.34%。而对来宾站降水预测相对误差提升最显著, 减少了 10.26%。相对于逐步回归预测结果, 对来宾站降水预测精度提升最为明显, MAE 最大减少 26.91%, RMSE 最大减少 23.10%。

本文利用 LR-RBF 预测降水, 并与单一的 RBF 神经网络和传统的统计模型作对比, 验证组合模型预测效果较好。本模型首先利用逻辑回归分类有无降水, 然后将有降水日带入 RBF 神经网络模型进行降水预测, 避免太多 0 值对模型预测的干扰。同时 RBF 神经网络模型有较好的非线性局部逼近、自学习能力, 对比传统的线性统计方法预测降水更适合。该模型主要是考虑地面常规气象因子对降水预测的影响, 没有考虑更多大气环流预报因子, 预测结果具有一定的局限性。在今后的研究中, 可以考虑加入其他可能的影响因子做进一步的探讨。

致 谢

本论文的顺利完成离不开大家共同的努力, 蒋宜蓉老师提供基金支持并给出审稿建议, 苏健昌负责径基神经网络模型的建立, 覃碧莉负责逻辑回归模型的建立, 吴俊皇负责气象数据的下载和处理, 向四位同仁表示感谢。

基金项目

国家自然科学基金项目(12161028)、广西科技基地和人才专项(AD20159017)。

参考文献

- [1] 周林. 华南前汛期降水年代际异常特征及机理研究[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2016.
- [2] 廖胜石, 卓健, 罗建英, 等. 广西汛期极端短时强降水特征分析[J]. 暴雨灾害, 2022, 41(3): 308-314.
- [3] 覃卫坚, 何莉阳, 蔡悦幸. 基于两种机器学习方法的广西后汛期降水预测模型[J]. 气象研究与应用, 2022, 43(1): 8-13.
- [4] 吴香华, 蒙芳秀, 熊萍萍, 等. 基于自组织映射神经网络的吉林省春夏期降水统计模拟研究[J]. 大气科学学报, 2018, 41(6): 829-837.
- [5] 顾伟宗, 张莉, 伯忠凯, 等. 不同时间尺度上黄淮地区夏季降水异常成因及预测研究[J]. 大气科学学报, 2018, 41(3): 388-394.
- [6] 赵中军, 刘善亮, 游大鸣, 等. 偏最小二乘回归模型在辽宁汛期降水预测中的应用[J]. 干旱气象, 2015, 33(6): 1038-1044.
- [7] 刘婷婷, 陈海山, 蒋薇, 等. 基于土壤湿度和年际增量方法的我国夏季降水预测试验[J]. 大气科学, 2016, 40(3): 591-603.
- [8] 陈潇潇, 石银, 张楠楠, 等. 包含外强迫因子的平潭旬降水量预测[J]. 气象与环境科学, 2016, 39(2): 99-103.

- [9] 王建华, 安海钰, 张敬. 基于灰色预测与 BP 神经网络的极端降水事件预测[J]. 应用数学进展, 2022, 11(4): 1847-1856.
- [10] 马婷婷. WRF-CMAQ-BP 神经网络空气质量预报模型研究[J]. 应用数学进展, 2022, 11(2): 641-650.
- [11] 沈艳, 杨春雷, 张庆国, 等. 基于 RBF 神经网络的池州市降水序列预测[J]. 安徽农业大学学报, 2012, 39(3): 451-455.
- [12] 季刚, 姚艳, 江双五. 基于径向基神经网络的月降水量预测模型研究[J]. 计算机技术与发展, 2013(12): 186-189.
- [13] 农吉夫, 金龙. 基于 MATLAB 的主成分 RBF 神经网络降水预报模型[J]. 热带气象学报, 2008, 24(6): 713-717.
- [14] 李洁. 基于 MATLAB 工具箱的 PCA-RBF 广西降雨量预测模型研究[J]. 现代计算机(专业版), 2015(18): 7-10.
- [15] 张飞涟, 刘严萍. 经验模态分解与神经网络方法在降水预测领域的应用研究[C]//中国系统工程学会学术年会·中国系统工程学会第十八届学术年会论文集-A01 系统工程. 2014: 2.
- [16] 蒋林利. 基于混合优化的 RBF 神经网络集成的降水预报模型[J]. 柳州师专学报, 2012, 27(2): 113-119.
- [17] 罗芳琼, 吴春梅. 基于 PCA-LLE 组合的优化 RBF 神经网络降水预测模型[J]. 柳州师专学报, 2013, 28(2): 126-130.
- [18] 夏侯杰, 郑婧. 基于 Logistic 判别模型的江西省降水分等级预报研究[C]//中国气象学会. 第 35 届中国气象学会年会 S1 灾害天气监测、分析与预报. 2018: 1791-1803.
- [19] 张芳华, 曹勇, 徐珺, 等. Logistic 判别模型在强降水预报中的应用[J]. 气象, 2016, 42(4): 398-405.
- [20] 李天才. 基于 Logistic 回归模型的降雨型泥石流临界雨量定量分析[J]. 吉林水利, 2014(8): 48-49.
<https://doi.org/10.3969/j.issn.1009-2846.2014.08.015>
- [21] 赵晗, 孙德山. 基于灰色关联与 RBF 神经网络的港口吞吐量预测[J]. 应用数学进展, 2020, 9(10): 1751-1756.
- [22] 蒙芳秀, 覃碧莉, 苏健昌, 吴俊皇, 蒋宜蓉. 一种基于神经网络的降水预测方法、装置和计算机设备[P]. 中国, CN114169502A. 2022-03-11.
- [23] 田东霞, 曹久才. 基于逐步回归法和 BP 神经网络模型的苹果产量预测[J]. 现代农业科技, 2022(14): 131-133, 142. <https://doi.org/10.3969/j.issn.1007-5739.2022.14.036>
- [24] 万夫敬, 袁慧玲, 宋金杰, 等. 南京地区降水预报研究[C]//第 30 届中国气象学会年会论文集. 2013: 1-12.
- [25] 高昂, 吴时强, 吴修锋, 等. BP 神经网络在浅水湖泊风浪波高预测中的应用[J]. 水电能源科学, 2022, 40(8): 41-44.