

## 摘要

随着气候变化的日益严重,准确预测降水量对于农业、水资源管理和灾害防范至关重要。本文搜集了中国近百年全国范围内的月降水量数据,采用 BP 神经网络和长短期记忆(LSTM)神经网络,对未来月降水量进行了预测。本文以武汉市地区为例,首先,对数据进行预处理和基于时间维度分析,求出武汉市 1901 年至 2022 年的月平均降水量,分析了月降水量在时间维度上的分布特点。其次,先后建立了 BP 神经网络与长短期记忆神经网络模型,对模型进行多次训练,并根据评价指标对模型进行进一步优化。最后,采用遗传算法优化了 LSTM 模型参数,再使用优化后的 GA-LSTM 模型对 2023-2030 年的月降水量进行了预测,并对未来趋势进行了讨论。

**关键词:** BP 神经网络; 长短期记忆神经网络; GA-LSTM 模型; 时间序列模型; 月降水量预测

## Abstract

With climate change on the rise, accurate precipitation forecasting is critical for agriculture, water management and disaster preparedness. This paper collects the monthly precipitation data of China in the last hundred years, and uses BP neural network and long short-term memory (LSTM) neural network to forecast the future monthly precipitation. Taking Wuhan as an example, this paper firstly preprocessed the data and analyzed the monthly average precipitation of Wuhan from 1901 to 2022 based on time dimension, and analyzed the distribution characteristics of monthly precipitation in time dimension. Secondly, BP neural network and long short-term memory neural network models are established successively, and the models are trained many times, and the models are further optimized according to the evaluation indicators. Finally, the parameters of the LSTM model were optimized by genetic algorithm, and then the monthly precipitation from 2023 to 2030 was predicted by the optimized GA-LSTM model, and the future trend was discussed.

**Key words:** BP neural network; Long short-term memory neural network; GA-LSTM model; Time series model; Monthly precipitation forecast

## 目录

摘要 .....	I
表格与插图清单 .....	IV
一、引言 .....	1
(一) 研究背景 .....	1
(二) 研究现状 .....	2
1. 降雨预测的研究现状.....	2
2. BP 神经网络的研究现状 .....	2
3. 长短期记忆神经网络的研究现状.....	2
4. 遗传算法的研究现状.....	3
(三) 研究意义 .....	3
二、数据处理与分析 .....	4
(一) 数据来源 .....	4
(二) 数据预处理 .....	4
(三) 基于时间维度的数据分析 .....	6
三、模型构建 .....	8
(一) 基本思路 .....	8
(二) 模型介绍 .....	8
1. BP 神经网络 .....	8
2. 长短期记忆神经网络.....	10
3. 遗传算法.....	12
(三) 模型构建 .....	13
1. BP 神经网络预测模型 .....	13
2. LSTM 预测模型 .....	13
3. GA-LSTM 预测模型 .....	14
(四) 模型性能评价指标 .....	14
四、结果分析 .....	15

（一）统计描述 .....	15
1. 中国降水量的空间分布.....	15
2. 武汉历史降水量的描述与分析.....	16
（二）预测效果 .....	17
1. BP 神经网络模型预测结果 .....	17
2. 长短期记忆神经网络模型预测结果.....	18
3. GA-LSTM 模型预测结果 .....	20
五、结论与分析 .....	23
六、总结 .....	24
参考文献 .....	26
附录 .....	28

## 表格与插图清单

表 1 中国地区空间位置范围.....	5
表 2 武汉地区空间位置范围.....	5
表 3 模型的数据特征.....	5
表 4 变量名称及解释（部分） .....	6
表 5 模型训练集拟合结果性能指标统计表.....	23
表 6 模型测试集拟合结果性能指标统计表.....	23
表 7 武汉市 1-3 月的降水量预测值 .....	23
表 8 武汉市 4-6 月的降水量预测值 .....	23
表 9 武汉市 7-9 月降水量预测值 .....	23
表 10 武汉市 10-12 降水量预测值 .....	24
表 11 变量名称及解释（完整） .....	28
表 12 武汉市 1901-2022 年历史月降水量数据集 .....	28
表 13 武汉市 2023-2030 年预测月降水量数据集 .....	31
图 1 数据集所取的空间范围.....	4
图 2 武汉市 1901 年-2022 年的月降水量时序图 .....	7
图 3 武汉市 1901 年-2022 年的月降水量箱线图 .....	8
图 4 BP 神经网络结构图.....	9
图 5 BP 神经网络算法流程图.....	10
图 6 LSTM 结构图 .....	11
图 7 遗传算法流程图.....	12
图 8 中国地面气候示意图(1981-2010).....	16
图 9 武汉市 1901 年-2022 年的月降水量趋势图 .....	17
图 10 BP 神经网络模型下训练集真实值与预测值拟合折线图.....	17
图 11 BP 神经网络模型下测试集真实值与预测值拟合折线图.....	18
图 12 BP 下训练集预测值和真实值（左）、测试集预测值和真实值（右） .....	18
图 13 LSTM 模型下训练集真实值与预测值拟合折线图.....	19
图 14 LSTM 模型下测试集真实值与测试值拟合折线图.....	19

图 15 LSTM 模型下训练集预测值和真实值（左）、测试集预测值和真实值（右）	
.....	31
图 16 GA-LSTM 模型下训练集真实值与预测值拟合折线图.....	21
图 17 GA-LSTM 模型下测试集真实值与预测值拟合折线图.....	21
图 18 GA-LSTM 训练集预测值和真实值（左）、测试集预测值和真实值（右）	
.....	22

# 基于中国百年月降水量大数据的统计建模与预测模型研究

## Research on Statistical Modeling and Prediction Models Based on Big Data of Century-Long Monthly Precipitation in China

### 一、引言

#### (一) 研究背景

近年来,随着气候变化的加剧和全球环境问题的日益突出,对于气象数据的分析和预测成为了科学研究和社会发展的重要课题。月降水量作为气象数据中的重要指标之一,对于农业、水资源管理、防灾减灾等领域具有重要意义。特别是在中国这样一个气候多样、降水分布不均匀的国家,对月降水量进行统计建模和预测具有重要的实际意义。

中国是世界上人口最多的国家之一,气候对其农业生产和社会发展有着重要影响。因此,准确地预测月降水量对于指导农业生产、合理利用水资源、制定防灾减灾措施等具有重要的实际意义。然而,传统的气象预测模型往往受限于样本数据的局限性、模型复杂度以及区域性的差异等问题,导致预测精度有限。

随着大数据和人工智能技术的不断发展,中国近百年的月降水量数据得到了更为广泛和详尽的收集与整理。这为深入研究中国降水量的统计规律和预测趋势提供了良好的数据基础和技术支撑。通过分析历史降水数据,我们可以发现降水量的周期性、趋势性以及与其他气象因素的相关性,从而构建更加精准的预测模型。利用机器学习算法和深度学习技术,可以帮助我们更深入地探究发现数据之间复杂的关系,从而提升预测结果的准确性。

武汉是湖北省的省会城市,经济发达,地处亚热带季风气候区,在降水方面表现出明显的季节性,夏季降水集中,而冬季则较为干燥。年降水量 1150 毫米-1450 毫米,预测这种季节性降水对该地区未来的农业生产和生态环境有着重要价值。

因此,本论文旨在基于中国百年月降水量的大数据,以武汉市地区为例,通过统计建模和预测模型的研究,提高对月降水量变化的理解和预测能力,为农业生产、水资源管理、生态环境保护以及防灾减灾工作提供科学依据和决策支持。通过深入分析中国百年来的月降水量数据,并结合大数据统计和机器学习等先进

方法，我们期望构建出高效准确的月降水量预测模型。这个模型可以为应对气候变化和自然灾害提供重要参考，帮助农业生产、水资源管理和应急响应等方面做出更有针对性的决策。

## （二）研究现状

目前，地区降水量预测已成为气象学、水资源管理和灾害防治等领域的研究热点之一。在过去几十年里，随着科技的进步和数据的积累，人们对地区降水量的研究取得了显著进展。

### 1. 降雨预测的研究现状

降雨预测作为气象学和水文学领域的关键研究课题，对于水资源管理、农业规划、城市排水系统设计等方面具有重要意义。近年来，随着机器学习和深度学习技术的兴起，基于数据驱动的降雨预测方法得到了快速发展。

2006 年，陈明金和欧阳祖熙利用基于遗传算法的 BP 神经网络对孔隙水压力的时间序列进行内插和外延预测。他们得到的预测效果良好，成功地反映了降雨过程与孔隙水压力变化之间的内在物理关系。这项研究在水资源管理和灾害预防等方面具有潜在的应用前景。<sup>[12]</sup>

### 2. BP 神经网络的研究现状

BP 神经网络是一种采用误差反向传播算法进行训练的多层前馈神经网络。

2023 年，曹昕等研究者通过对混凝效果预测算法研究历史进行回顾与分析，总结神经网络模型在混凝效果预测中的研究现状和脉络，深入分析不同数据来源与数据格式的优缺点。<sup>[11]</sup>在硬件加速与并行计算方面，随着大数据和深度学习的发展，BP 神经网络对计算资源的需求日益增加。

2005 年，秦伟刚研究者采用 BP 神经网络和其改进算法建立了车流量预测模型。<sup>[6]</sup>因此，利用 GPU、FPGA 等硬件进行加速计算，以及开发并行算法，成为了研究的热点。在跨学科应用方面，BP 神经网络在图像识别、语音处理、金融分析等多个领域的应用不断拓展，研究者正尝试将其与专业知识结合，解决更多实际问题。

### 3. 长短期记忆神经网络的研究现状

长短期记忆神经网络（Long Short-Term Memory, LSTM）是一种特殊的循环神经网络（Recurrent Neural Network, RNN），由 Hochreiter 和 Schmidhuber 于

1997 年提出。在自然语言处理（NLP）领域，LSTM 被广泛应用于语言模型的构建、机器翻译、文本生成、情感分析等任务，其在处理序列数据方面的优势使其在这些任务中取得了突破性进展。

2023 年，柯杰研究者主要设计在编码器部分采用 VGG19 和 Res Net101 对图片进行特征提取,解码器使用长短期记忆网络,并提出了一种新的改进注意力机制来增强图片与词之间的相关性,最后输出自然语言。<sup>[1]</sup>在生物信息学和医学研究中，LSTM 被用于基因序列分析、疾病预测等，通过学习生物序列数据的模式，有助于发现生物标志物和理解疾病的发生机制。

#### 4. 遗传算法的研究现状

遗传算法（Genetic Algorithm, GA）是一种模拟生物进化过程的启发式搜索和优化算法。研究者们持续关注遗传算法的数学基础、收敛性分析以及算法性能的理论保障。

1997 年，吴新余和孙力娟在遗传算法中利用多种混沌模型构造随机开关，以此来控制遗传算法中的交叉运算，从而可迅速地求解 0-1 规划问题，并进一步在求解通信网的优化问题中得到成功的应用。<sup>[8]</sup>

### （三）研究意义

地区的降水量对整个地区有着非常重要的影响。通过掌握地区降水量的规律并准确预测未来降水情况，可以增强相关部门应对涝灾和旱灾的能力，提前部署相关战略，保障人们的生活。

其次，该研究将为农业生产提供重要参考。降水是农业生产中不可或缺的因素，预测未来月降水量，能够帮助农民做出更合理的种植和灌溉决策，提高农业生产效益，确保农作物稳产高产。

此外，该研究还对水资源管理和生态环境保护具有重要意义。降水量的变化直接影响着地表水和地下水资源的补给与分布，通过建立降水量的统计模型和预测方法，可以为水资源的科学调度和合理利用提供支持，保障水资源的可持续利用。同时，科学的降水量预测也有助于生态环境保护和恢复，为生态系统的可持续发展提供指导。

尽管降雨量预测研究取得了一定的进展，但仍然存在不确定性，特别是在极端事件和复杂地形区域的预测。因为机器学习模型在降雨量预测中表现出色，故



本文针对武汉市月降水量的历史发展规律,采用 BP 神经网络、长短期记忆神经网络方法,对武汉市过去 122 年 1 月至 12 月的降水量数据进行分析,利用预处理后武汉市月降水量数据,结合多种算法进行预测比较,从而获得一个较好的长期时序预测模型,用于对未来 8 年月降水量进行预测,为武汉市月降水量的预测提供一个积极的指导作用。

## 二、数据处理与分析

### (一) 数据来源

本文使用分析的数据来源于“国家青藏高原科学数据中心”(http://data.tpdc.ac.cn)。该数据集为中国逐月降水量数据,其空间分辨率为  $0.0083333^{\circ}$  (约 1km),时间涵盖 1901.1 至 2022.12。数据呈现 NETCDF 格式,即.nc 格式。该数据是依据 CRU 所发布的全球  $0.5^{\circ}$  气候数据集以及 WorldClim 所发布的全球高分辨率气候数据集,通过 Delta 空间降尺度方案在中国区域降尺度而生成。此外,还使用了 496 个独立气象观测点的数据进行验证,验证结果具有可信度。本数据集所包含的地理空间范围是全国主要陆地(包括港澳台地区),不包括南海岛礁等区域。<sup>[13]</sup>如图 1 所示。



图 1 数据集所取的空间范围

### (二) 数据预处理

本文利用全国近 100 年的月降水量数据集,采用 BP 神经网络和 LSTM 神经网络

络模型来预测未来 8 年的月降水量。为了确保我们的模型能够取得良好的预测效果，我们需要进行一系列合理的数据预处理步骤，将原始数据转换为适合模型训练的形式。

第一，将.nc 文件通过 Matlab 中 ncread 函数读取全国范围内 1902-2022 年月降水量，并将收集到的数据按照年份划分，对表 1 中国地区空间位置范围经纬度范围内的所有观测到的数据求和并除以观测点总数，得到每年全国范围内的月降水量数据。

表 1 中国地区空间位置范围

方位	经纬度（单位：度）
East	136.196045022846
West	72.1960450229045
South	16.2504610568287
North	55.5587943901263

第二，由于以 1901 年至 2022 年全国主要陆地（包含港澳台地区），不含南海岛礁等区域（空间位置见下表 1）月降水量数据为研究对象，在考虑到我国幅员辽阔，降水量受地理空间分布差异较大，不同地区气候差异明显，无法很好的体现降水量的时序规律，为提高预测精度，以武汉市地区（空间位置见下表 2）为例作为主要研究对象。

表 2 武汉地区空间位置范围

方位	经纬度（单位：度）
East	115.07704
West	113.696596
South	29.971819
North	31.362913

第三，由于数据的长期性和来源的多样性，存在少量缺失值和异常值，这会影影响模型的训练和预测。用填充缺失值、删除异常值或进行插值处理对数据集进行初步处理，以确保数据的完整性和准确性。

第四，本文目标是预测未来的月降水量，还需要将数据划分为训练集和测试集。本研究基于连续 122 年共计 1464 条月降水量数据建立模型，选取整理后前 1200 组数据作为训练集样本，后 264 组数据作为测试集样本，用以评价模型预测的准确性。

表 3 模型的数据特征

数据集	数据起止时间	样本数
-----	--------	-----

训练数据	1901 年 1 月-2000 年 12 月	1200
测试数据	2001 年 1 月-2022 年 12 月	264

第五，对数据进行归一化处理，以确保不同特征的尺度统一，避免模型受到特征尺度的影响，以收敛数据范围，加快模型的训练速度。归一化后数据的范围为[0,1]。

$$x_{norm} = \frac{x - \min}{\max - \min} \quad (1)$$

通过以上数据预处理步骤，我们将为模型提供高质量的数据基础，为接下来的建模和预测任务奠定坚实的基础。

表 4 变量名称及解释（部分）

变量	解释
Time 降水时间	月份（公历）
Latitude 纬度	[-90.0, 90.0] 北纬为正，南纬为负
Longitude 经度	[-180.0, 180.0] 东经为正，西经为负
Precipitation 降水量	月降水量 单位：0.1mm

### （三）基于时间维度的数据分析

通过绘制武汉市 1901 年至 2022 年的降水量的三维视图（见图 2），我们可以很直观地看到降水量在时间和空间上的变化情况。在这个三维视图中，横轴代表年份，纵轴代表月份，而立体图形的高度则代表降水量的大小。通过观察这个三维图，我们可以清晰地看到降水量在不同年份和月份之间呈现出明显的变化趋势。我们可以看到在每年的六、七和八月降水量明显增加，通常在 2000 到 4000 毫米之间，在七月份降水量达到峰值。这三个月通常是武汉市的雨季，降水量较大，对应着夏季和初秋季节。在这个季节里，降水量较多，对应着气候的潮湿和多雨特点。然后可以发现不同年份间，这三个月降水量的差异最为明显，呈现出明显的不稳定性。有些年份的降水量较高，有些年份则较低，反映出降水量在年度间的波动性和不确定性。这可能受到气候变化、大气环流和其他气象因素的影响，导致不同年份的降水量存在着较大的差异。

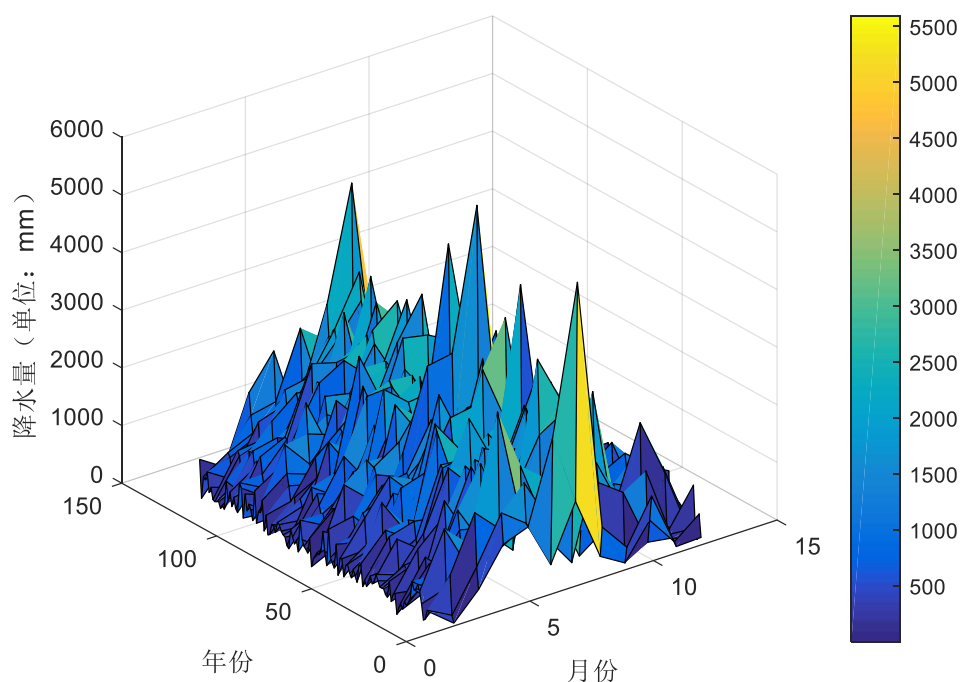


图 2 武汉市 1901 年-2022 年的月降水量时序图

箱线图能够清晰地展示数据的分布情况和统计特征，根据武汉市 1901 年-2022 年的月降水量的箱线图（见图 3），可以更加直观观察到降水量的均值、中位数和标准差等指标在不同月份之间存在着明显的差异。降水量在夏季和秋季通常较高，在冬季较低，呈现出明显的季节性变化；在七月份异常值最多，且跨度较大，二、十和十二月虽然异常值也较多，但是相对于比较集中，说明降水量在七月可能受到其他因素影响较大，如海洋风等。此外月降水量在前七个月逐渐增加后五个月逐渐减少到趋势明显。

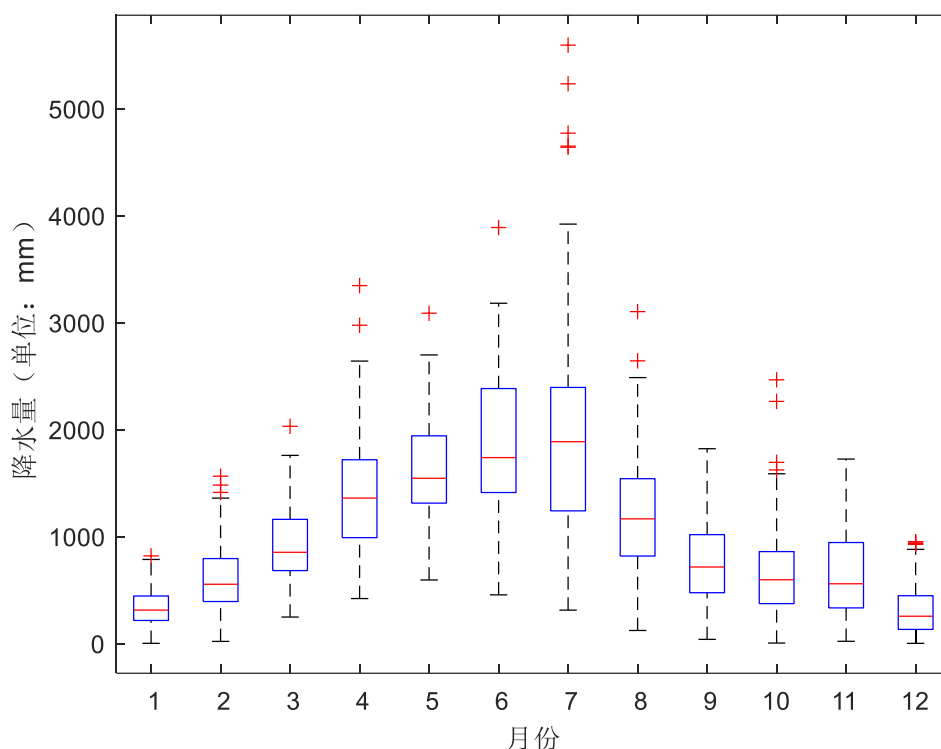


图 3 武汉市 1901 年-2022 年的月降水量箱线图

### 三、模型构建

#### (一) 基本思路

基于上述提到的研究目的和目前针对此类数据的研究现状,我们提出如下研究思路。通过使用 1901 年至 2022 年武汉市月降水量数据,先后采用了 BP 神经网络和长短期记忆神经网络 (LSTM) 对历史月降水量进行长期时序模型训练,并使用训练好的模型预测了 2023 年至 2030 年的月降水量变化趋势,最后使用遗传算法改进长短期记忆神经网络,建立 GA-LSTM 模型,通过遗传算法计算长短期时间记忆网络的模型参数,从而使模型训练效果更好。

#### (二) 模型介绍

##### 1. BP 神经网络

BP 神经网络是一种经典的人工神经网络,用于解决分类、回归等问题。它通过反向传播算法来训练神经网络,利用梯度下降法不断调整网络参数,使得网络的输出逼近真实标签,从而最小化损失函数。BP 神经网络由输入层、隐藏层和输出层组成,每个层都包含多个神经元,神经元之间的连接由权重参数决定,这些

权重相互独立。

神经元是网络的基本组成单位，每个神经元都有输入权重  $W_i$ 、加权和、激活函数  $\sigma$  和输出  $y_i$  这四个重要部分，神经元输出为：

$$y_i = \sigma(\sum W_i x_i + b) \quad (2)$$

整个 BP 神经网络结构的设计使得它能够适应各种复杂的任务，并通过大量数据的训练来不断优化性能。

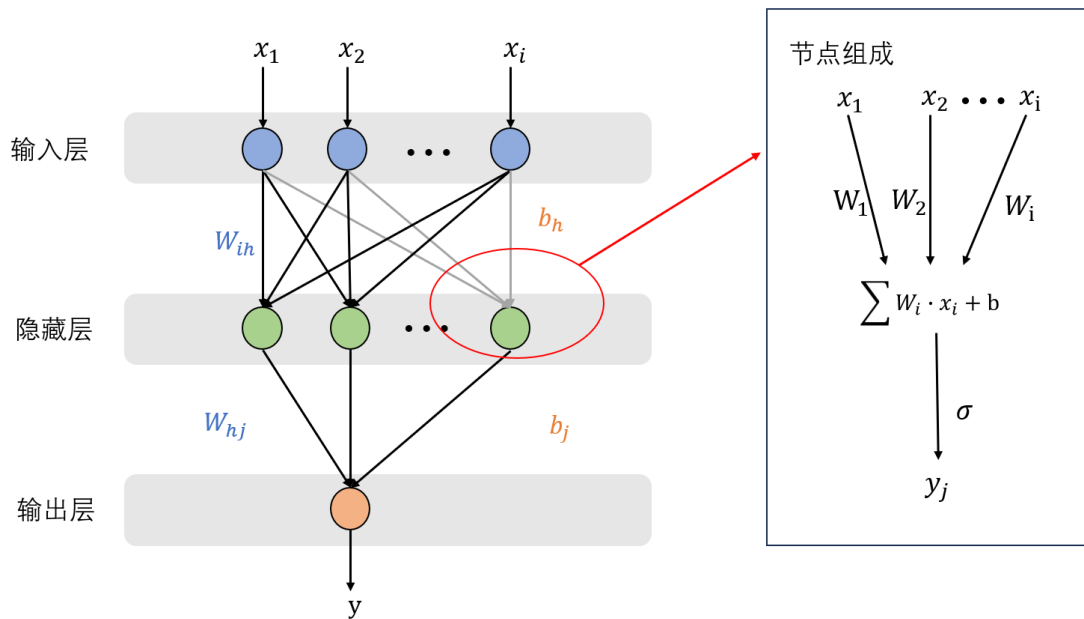


图 4 BP 神经网络结构图

我们将过去几年的一月份降水量  $x_i$  作为输入特征， $i$  表示第  $i$  年。每个  $x_i$  作为输入层节点。如有  $n$  个输入层节点，表示过去  $n$  年的 1 月份月降水量。根据 (1) 可以计算得出隐含层节点，若有  $j$  个隐含层节点，再由 (1) 计算得出的结果作为输出层输出节点，表示预测的降水量。最后使用反向传播算法来计算损失函数对于网络中所有权重和偏置的梯度，并利用梯度下降法来更新这些参数，使损失函数逐渐减小，直到达到停止条件为止。其余 11 个月份的降水预测相同。

具体的算法流程如下：

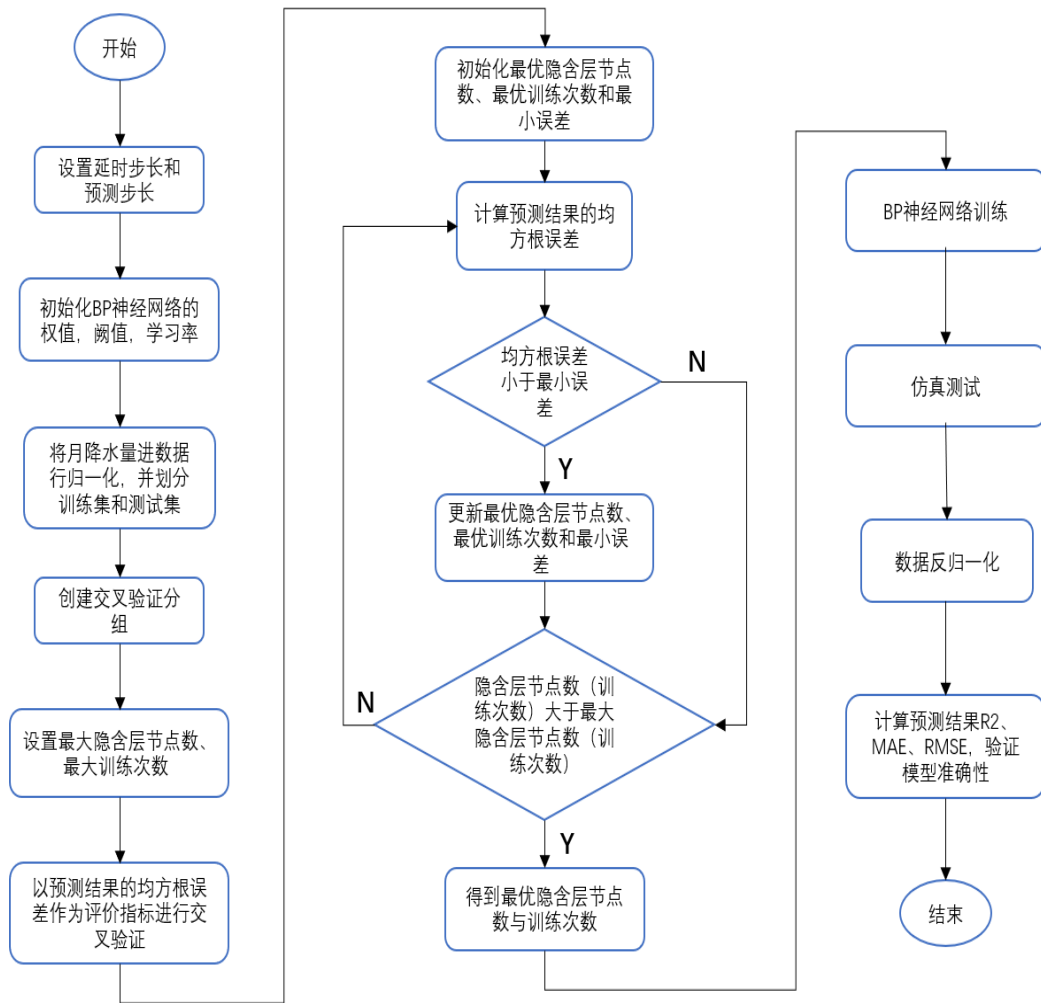


图 5 BP 神经网络算法流程图

## 2. 长短期记忆神经网络

BP 神经网络模型在处理长序列数据时，容易遇到梯度消失或数据爆炸等问题，为此，我们引入了长短期记忆神经网络（Long Short-Term Memory, LSTM），它是循环神经网络（Recurrent Neural Network, RNN）改进而得的神经网络，主要是为了解决此类问题以及难以捕捉长期依赖关系等难题而设计的。

LSTM 的核心思想是引入了一种称为“门”的机制，它是由细胞状态（Cell State）、遗忘门（Forget Gate）、输入门（Input Gate）和输出门（Output Gate）这四个重要结构组成。这些门对信息的流动进行控制，从而加强在处理长序列数据

时的表达能力和记忆能力，更加有效地捕捉长期依赖关系，实现对长期依赖关系的更好建模。

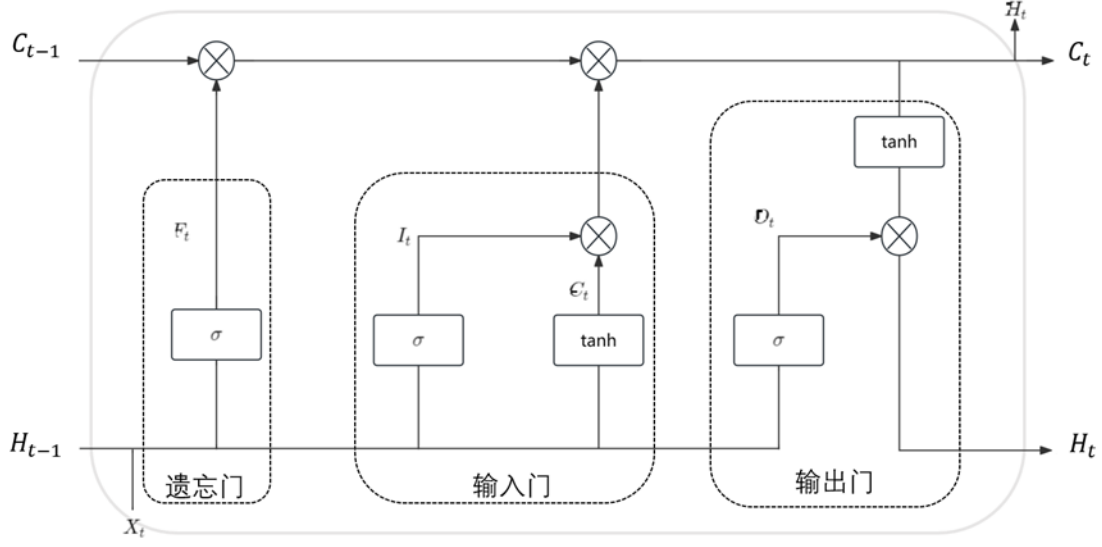


图 6 LSTM 结构图

$C_{t-1}$ 、 $C_t$  分别为存储单元的前一时刻和当前时刻的状态， $H_{t-1}$ 、 $H_t$  分别为前一单元和当前单元的输出，当前输入的数据为  $X_t$ 。

这些门机制使 LSTM 能够有效地捕捉长期依赖关系，并在处理长序列数据时具有更强的表达和记忆能力。所以当前语音识别、自然语言处理和图像描述生成等领域广泛应用 LSTM 进行预测处理数据。

LSTM 的核心算法步骤有：

① 遗忘门计算：决定在当前单元应该丢弃多少以前的细胞状态信息。它通过一个  $\sigma$ （sigmoid 激活函数）将输入和前一个单元的隐藏状态结合起来，得到公式（2），其中  $W_f$  和  $b_f$  分别为遗忘门的权值向量和损坏变量，最后计算得到一个在 0 到 1 之间的值，表示每个细胞状态的哪些部分应该被遗忘。

$$F_t = \sigma(W_f[H_{t-1}, x_t] + b_f) \quad (3)$$

② 输入门计算：决定当前单元应该更新多少新信息到细胞状态中。它通过一个  $\sigma$  和一个  $\tanh$  激活函数来确定两个部分：一个确定哪些值将被更新，另一个确定更新值的大小。

$$I_t = \sigma(W_i[H_{t-1}, x_t] + b_i) \quad (4)$$

$$C_t = \tanh(W_c[H_{t-1}, x_t] + b_c) \quad (5)$$



③ 细胞状态更新：在每个单元都会被更新，通过将上一个单元的细胞状态乘以遗忘门的输出以丢弃一些信息，并加上输入门的输出以更新新信息。得到公式（5），其中\*为矩阵的 Hadamard 积。

$$C_t = F_t * C_{t-1} + I_t * \bar{C}_t \quad (6)$$

④ 输出门计算：决定当前单元应该输出多少信息给下一个单元。它同样时通过一个  $\sigma$  和  $\tanh$  激活函数来确定输出的部分。

$$I_t = \sigma (W_o[H_{i-1}, x_t] + b_o) \quad (7)$$

$$H_t = O_t * \tanh(C_t) \quad (8)$$

### 3. 遗传算法

遗传算法（Genetic Algorithm, GA）是优化领域中一种基于计算机科学的启发式算法。其主要原理是借助对生物进化过程的模拟来寻找最优解。在该算法中，解决方案一般会被编码为染色体或基因型，同时利用适应度函数来衡量每个解决方案的好坏程度。通过实施选择、交叉以及变异等操作，使遗传算法能够持续迭代，产生新的解，并逐渐提升解的质量，直至达到终止条件。具体算法流程见图 7。

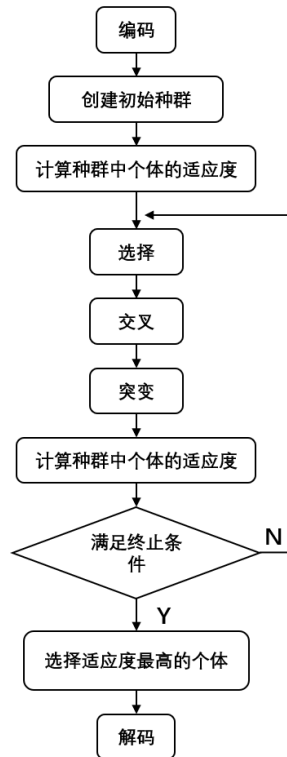


图 7 遗传算法流程图

### (三) 模型构建

#### 1. BP 神经网络预测模型

(1) 构建数据集：所用数据为典型的时间序列数据，于是使用滑动窗口进行划分，将延时步长设为 12 和时间跨度设为 1，每个窗口包含连续的一段时间数据，例如过去几个月的降水量数据，对于每个滑动窗口，将窗口内的降水量数据作为输入特征，下一个时间点（即下一个月）的降水量数据作为目标值。

(2) 创建 BP 神经网络模型：隐含层的传递函数采用了双曲正切函数（tansig），输出层的传递函数使用了线性函数（purelin），训练方法则是 Levenberg-Marquardt 反向传播算法（trainlm）。

(3) 交叉验证确定最佳参数：使用交叉验证确定最佳的模型参数，如隐含层节点数和训练次数。交叉验证可以减少因数据分割不合理而引入的偏差。在每一轮交叉验证中，使用训练集数据对神经网络模型进行训练，同时在测试集上进行性能评估。针对不同的隐含层节点数和训练次数组合，重复进行 K 次交叉验证，计算每次交叉验证的均方根误差。选择均方根误差最小的隐含层节点数和训练次数作为最佳参数。最后得到最佳隐含层节点数为 5，最佳训练次数为 3500。

(4) 训练 BP 神经网络模型：使用训练集对 BP 神经网络模型进行训练，以  $R^2$  作为评价指标辅助模型的权重和偏置的调整，以最小化损失函数。然后通过反向传播算法更新模型参数，不断优化模型的拟合能力。

(5) 模型评估：使用测试集评估训练好的模型在未知数据上的性能。

综合评价指标显示，训练数据的预测表现明显优于测试数据。具体来说，训练数据的  $R^2$  较高，但测试数据的  $R^2$  为负值，MAE 和 RMSE 在测试数据上表现相对较好。这种情况可能是由过拟合造成的，即模型在训练过程中过度适应了训练数据的特点，导致在测试数据上的泛化能力下降。总体而言，模型的预测效果一般，如果只需要了解降水量的大概趋势，则可以使用该模型进行预测，但结果可能不是非常准确。

#### 2. LSTM 预测模型

(1) 构建数据集：与 BP 神经网络模型相同，仍使用滑动窗口。

(2) 创建 LSTM 网络模型：利用 MATLAB 中的深度学习工具箱创建 LSTM 层、

全连接层和回归层 LSTM 层用于处理时间序列数据，并学习其内部模式和依赖关系。全连接层用于将 LSTM 层的输出映射到预测结果的空间。回归层用于计算损失函数，并优化模型参数以最小化损失。使用均方误差（MSE）作为损失函数。

（3）确定最佳参数：仍采用交叉验证方法确定最佳隐藏单元数和最大迭代次数。得到最佳隐含层节点数为 6，最佳训练次数为 5000。

（4）指定训练选项：采用 Adam 优化器以加快收敛速度，而学习率初始设置为 0.01。

（5）训练 LSTM 网络模型

（6）模型评估：仍采用 R2、MAE、RMSE 来评估模型的预测精度和拟合程度。

### 3. GA-LSTM 预测模型

将遗传算法（GA）和长短期记忆神经网络（LSTM）相结合，建立基于 GA-LSTM 模型的武汉月降水量分析预测模型，将 LSTM 模型的超参数（即学习率、隐藏单元数、最大迭代次数等）表示为染色体，并使用 GA 的进化过程来搜索最优的超参数组合。适应度函数可以根据模型在验证集上的性能来评估每个个体的适应度。通过不断迭代优化，可以找到使 LSTM 模型性能最佳的超参数组合，进一步提高模型解决非线性问题的能力。将遗传算法优化过的参数直接运用到 LSTM 模型中，进一步对武汉的降水量进行拟合预测。

#### （四）模型性能评价指标

为了更加简单明了的观察模型对武汉月降水量数据的拟合效果，本文建立了 BP 神经网络模型、长短期记忆神经网络（LSTM）模型和基于遗传算法的长短期记忆神经网络（GA-LSTM）模型这三种模型。用他们分别对训练数据和测试数据进行拟合，然后用预测值与真实值的对比的结果来检验模型预测效果。本文主要利用以下几种方法来检验模型拟合程度，衡量标准主要包括了三个指标：R2、MAE、RMSE

1. R2（决定系数）：

$$R2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (\hat{y}_i - \bar{y})^2} \quad (8)$$

R2 的值在 [0, 1] 区间内，越接近 1 说明拟合效果越好，反之越差。

2. MAE(平均绝对误差)：

$$MAE = \frac{1}{m} \sum_{i=1}^n |(y_i - \hat{y}_i)| \quad (9)$$

3. RMSE（均方根误差）：

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (10)$$

其中 $n$ 表示样本个数， $\hat{y}_i (i = 0, 1, \dots, n)$ 表示预测值， $y_i (i = 0, 1, \dots, n)$ 表示真实值，范围 $[0, +\infty)$ 。MAE 与 RMSE 的数值越大，则说明模型拟合效果越差，此外 RMSE 相对与 MAE 对异常数据值将更敏感。

## 四、结果分析

### （一）统计描述

#### 1. 中国降水量的空间分布

如图 8 中国地面气候示意图(1981-2010)<sup>[14]</sup>所示，中国的降水量在空间分布上具有显著的地域特征。东部沿海区域，涵盖华东、华南以及东北地区，因受到季风及海洋的影响，降水量较为丰富。像广东、福建、浙江等省份，年平均降水量通常都在 1000 毫米以上，部分地区甚至超过 2000 毫米。西南地区，包括云南、贵州、广西和西藏等地方，由于靠近喜马拉雅山脉，对来自东部海洋的湿润水汽起到明显的阻挡作用，从而导致降水量较大。华北平原地区，主要包括北京、天津、河北等地，降水量相对较少，年降水量一般在 400 毫米以下，是我国较为干燥的区域之一。西北地区，如新疆、甘肃、青海等地，降水量则是最少的。在新疆维吾尔自治区的西部，年降水量甚至可以低至 50 毫米以下。东北地区，包括黑龙江、吉林、辽宁等省份，降水量相对适中，不过在冬季会有较多的降雪。其年降水量一般在 500 毫米至 800 毫米之间。总体而言，中国降水量的空间分布呈现出明显的东部多西部少、南部多北部少的特点，这主要是受到地形、地理位置及气候等因素的综合影响。

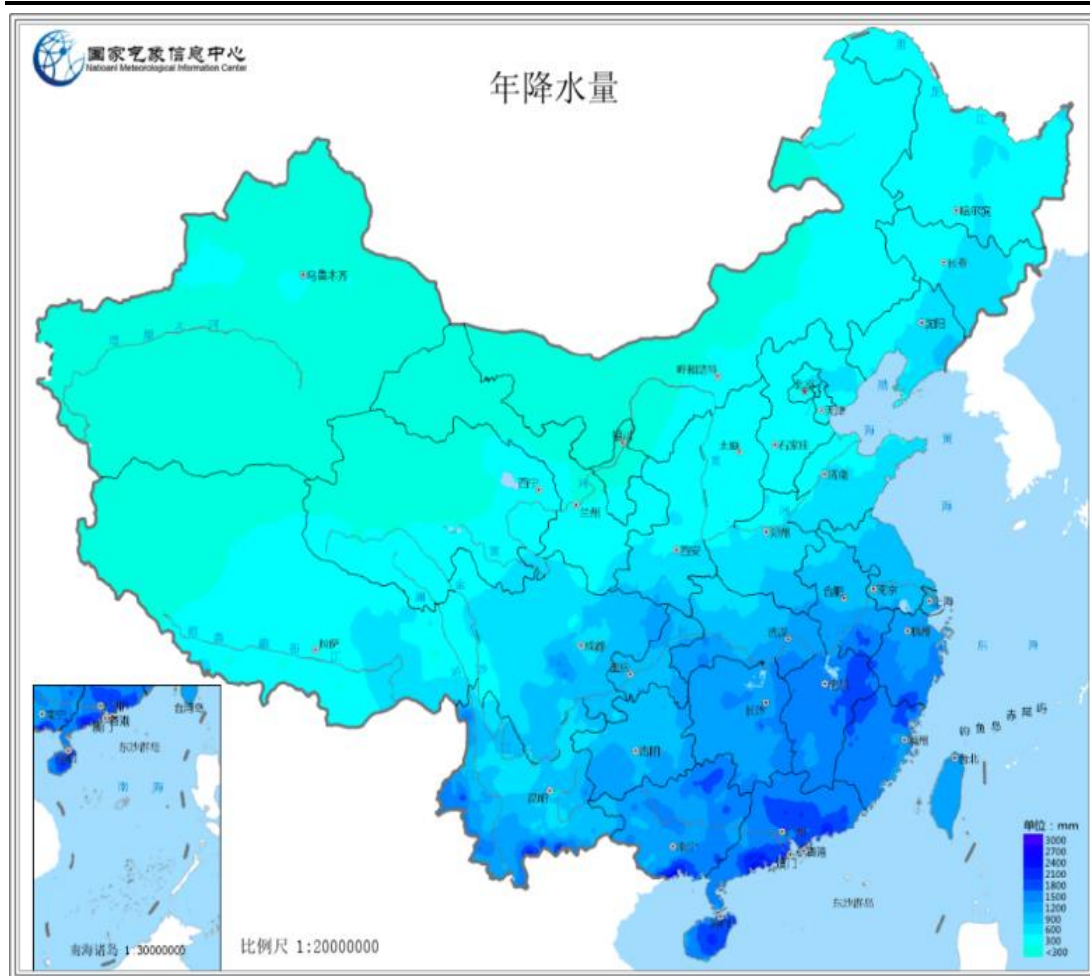


图 8 中国地面气候示意图(1981-2010)

## 2. 武汉历史降水量的描述与分析

武汉市 1901 年-2022 年全年的月降水量如图 4、图 5 所示。从图中可以看出，武汉市历史降水量自二十世纪初期至二十世纪中叶呈现总体上升趋势，二十世纪中叶至二十一世纪初期呈现总体下降趋势，通过查询武汉市历史上因异常降水量而引发自然灾害的相关资料可知，武汉在近百年的历史上经历过三次超大洪水的侵袭，分别是 1931 年、1954 年和 1998 年<sup>[15]</sup>。由图 5 可以看出，在 1931 年、1954 年和 1998 年，均发生过连续数月持续性大量降水，其中长江流域受灾尤为严重，导致武汉遭受了大规模的洪灾，致使许多人员伤亡和财产损失。同时，从图中可以明显看出，自二十世纪末期降水量开始减少，在 1999 年、2010 年、2022 年等发生过严重干旱，影响了农田灌溉和居民生活用水，导致农作物大量减产，河流干涸。

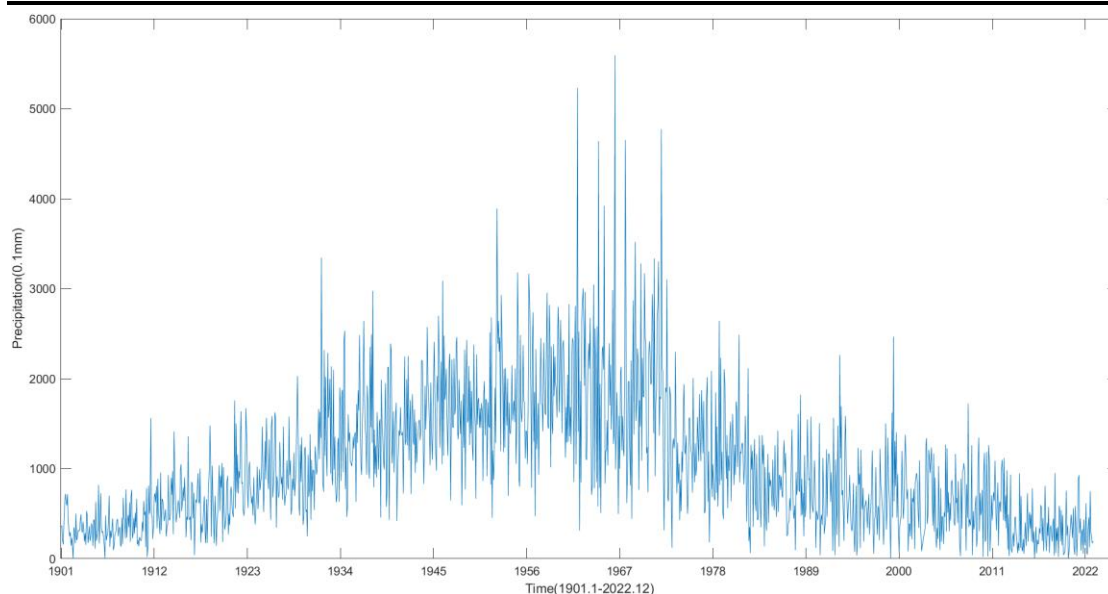


图 9 武汉市 1901 年-2022 年的月降水量趋势图

## (二) 预测效果

### 1. BP 神经网络模型预测结果

BP 神经网络以其强大的非线性处理能力和自学习能力在众多领域得到应用。它能够自动从输入数据中提取特征，适应性强，泛化性能好。BP 网络结构灵活，可通过调整神经元数量和层级来适应复杂问题，且具有较好的容错性和鲁棒性，即使在数据存在噪声或不完整的情况下也能保持预测性能。

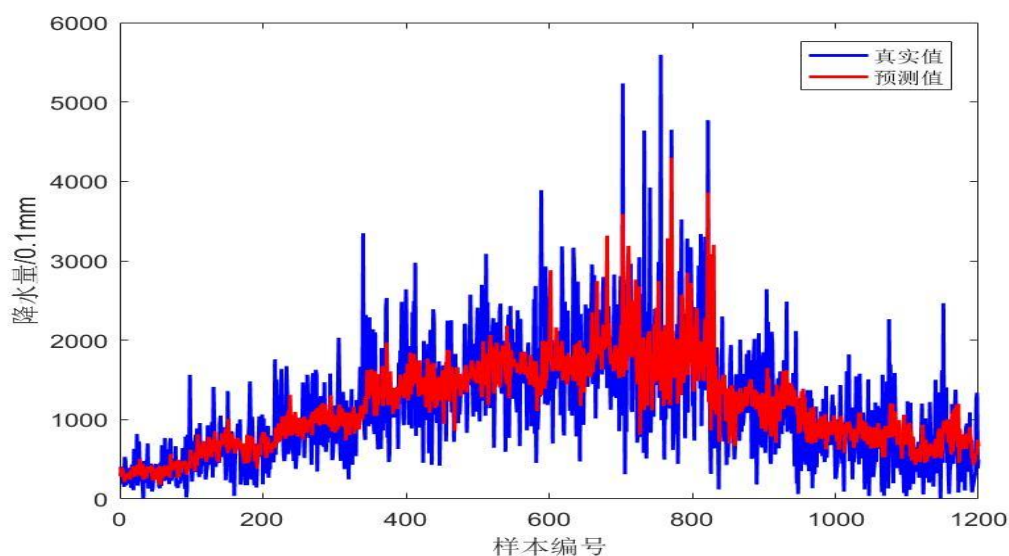


图 10 BP 神经网络模型下训练集真实值与预测值拟合折线图

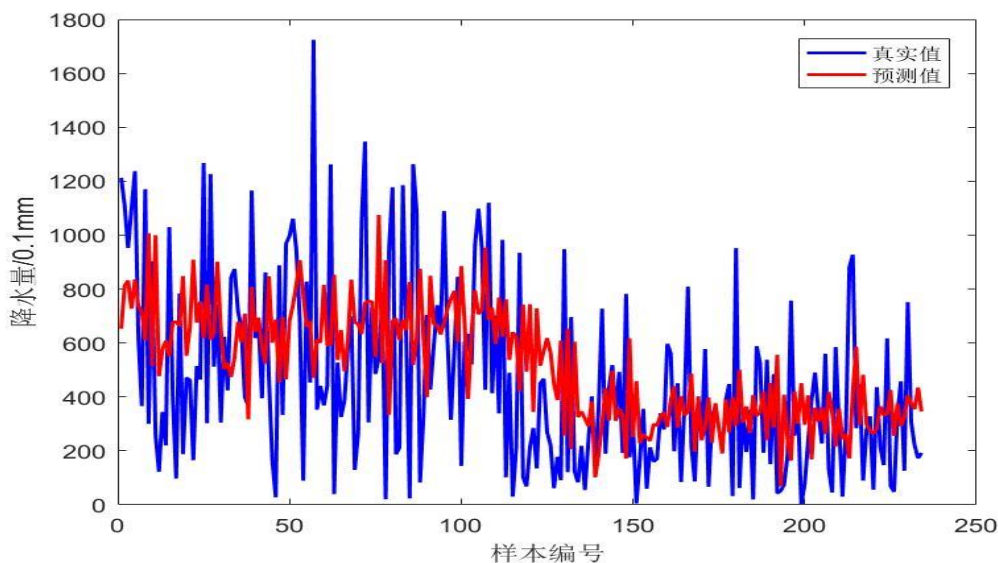


图 11 BP 神经网络模型下测试集真实值与预测值拟合折线图

由图 10 和图 11 可见，BP 神经网络模型对武汉市年降水量的预测效果并不好，训练集和测试集的真实值与预测值偏差很大。

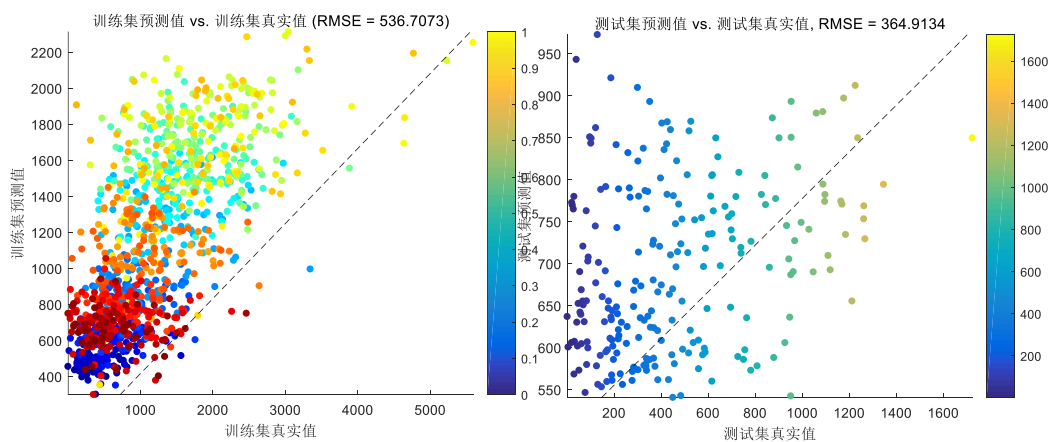


图 12 BP 下训练集预测值和真实值（左）、测试集预测值和真实值（右）

由图 12 可以看出，BP 神经网络模型下训练和测试的武汉市年降水的均方根误差分别为 536.7073、364.9134，从数值结果来看拟合值和真实值之间的偏差还是比较大的。

## 2. 长短期记忆神经网络模型预测结果

长短期记忆神经网络 (LSTM) 模型攻克了信息在网络单元中以链式形式流动时所产生的梯度消失及爆炸这一难题，极大地提升了模型学习的效率，在处理时

间序列预测及非线性映射等相关问题时，因其具备强大的记忆能力而展现出了良好的优势，所以被广泛应用于各个领域，而且经实践验证，在模型训练中取得了比较理想的效果。

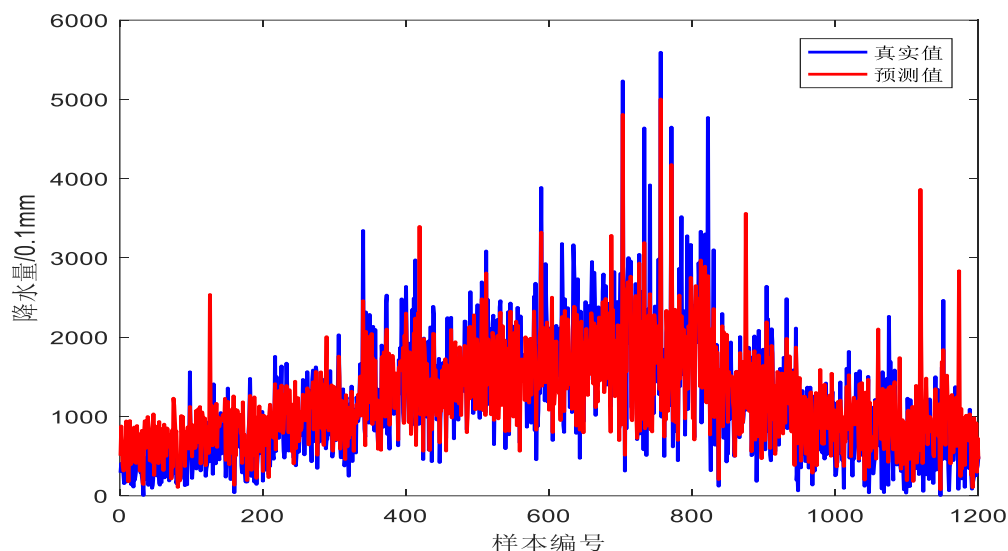


图 13 LSTM 模型下训练集真实值与预测值拟合折线图

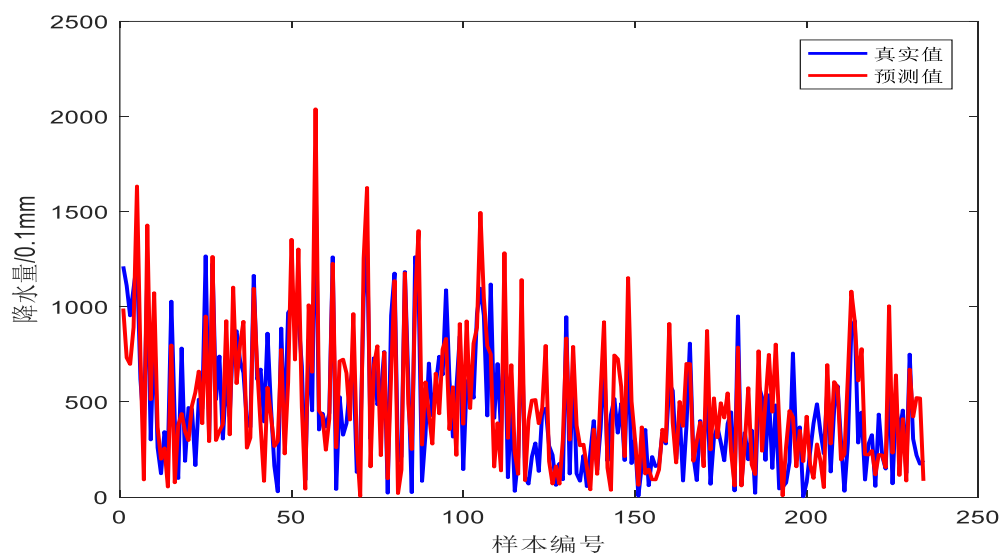


图 14 LSTM 模型下测试集真实值与测试值拟合折线图

图 13 和图 14 展示了武汉市降水量训练样本和测试样本的拟合结果，能够看出，武汉市月降水量的拟合值与真实值的趋势大致雷同，总体拟合效果不错，同时也有拟合误差相对较大的时段，该拟合效果与一般实验中数据的拟合情形相



符。

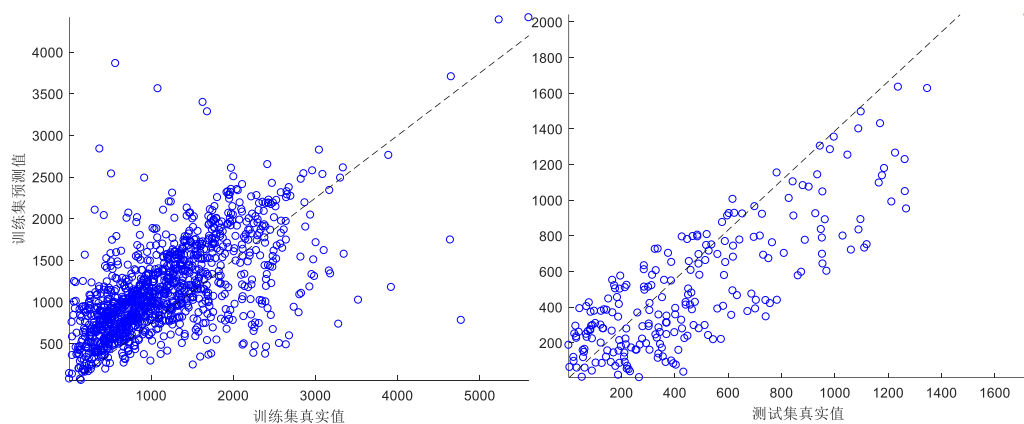


图 15 LSTM 模型下训练集预测值和真实值（左）、测试集预测值和真实值（右）

从图 15 可以看出，LSTM 模型下测试的武汉市近几十年降水量预测值和真实值拟合效果 BP 神经网络模型的好，整体线性关系强度较大。

### 3. GA-LSTM 模型预测结果

本文把遗传算法（GA）与长短期记忆神经网络（LSTM）进行了结合，构建起基于 GA-LSTM 模型的武汉市月降水量分析预测模型。通过遗传算法来优化 LSTM 的参数，获得学习率、隐层神经元个数以及训练次数的最优组合，从而进一步增强了模型解决非线性问题的能力。随后，把遗传算法优化好的参数直接应用到 LSTM 模型中，实现对武汉市降水量的拟合预测。

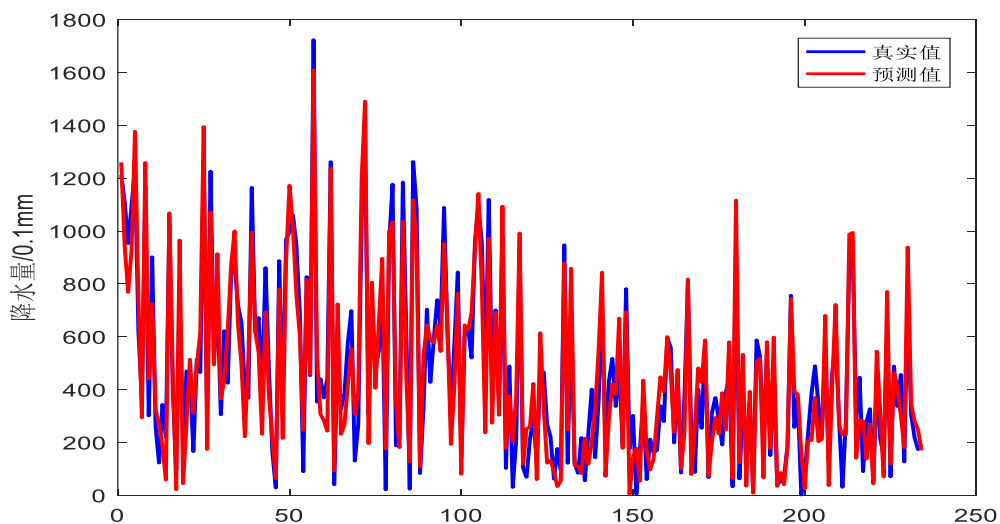


图 16 GA-LSTM 模型下训练集真实值与预测值拟合折线图

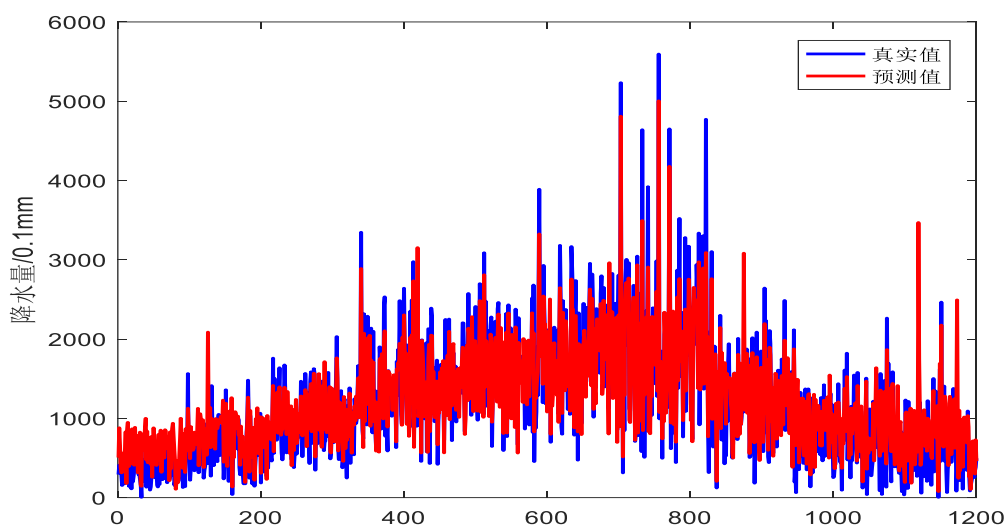


图 17 GA-LSTM 模型下测试集真实值与预测值拟合折线图

根据图 16 和图 17 对比观察 GA-LSTM 模型在训练集和测试集的拟合值和真实值之间的趋势以及拟合程度的差异，能够明显看出，在两个数据集上，GA-LSTM 模型的拟合值与真实值的走势大致相同。通过对比图 14 和图 15，能够发现两者拟合的程度相比无优化参数的 LSTM 模型的拟合程度，在肉眼观察下更为接近。

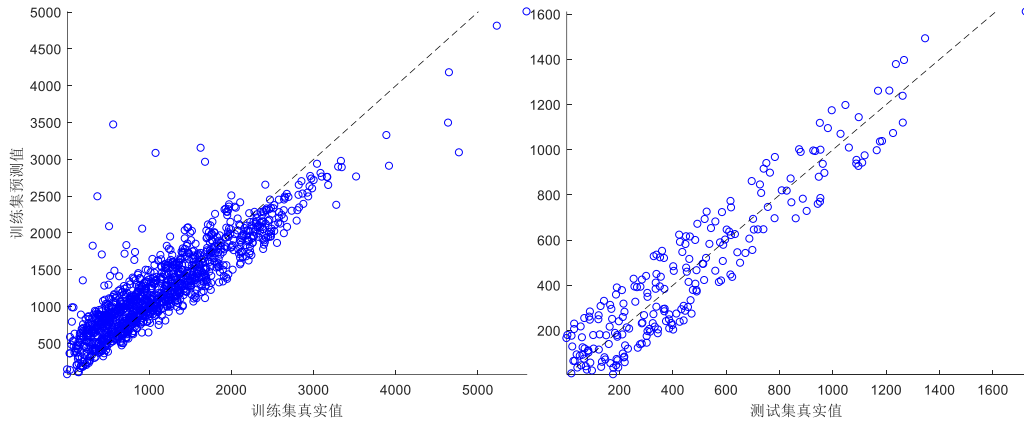


图 18 GA-LSTM 训练集预测值和真实值（左）、测试集预测值和真实值（右）

由图 18 明显可以看出，GA-LSTM 模型下武汉市近几十年降水量预测值和真实值线性关系强度更为明显。

数据结果见下表 5、表 6，对比 BP 神经网络，LSTM 和优化后 GA-LSTM 模型的训练集与测试集拟合结果性能指标，可以看出 GA-LSTM 模型优于 LSTM 模型和 BP 神经网络模型，LSTM 模型又优于 BP 神经网络模型。

根据表 5，可知在训练集上三种模型都显示出了一定程度的拟合能力，但其性能有所不同。BP 模型的  $R^2$  为 0.51284，MAE 为 379.3371，RMSE 为 519.543。LSTM 模型稍微提升了性能，其  $R^2$  为 0.67526，MAE 为 294.9144，RMSE 为 423.9305。而 GA-LSTM 模型在训练集上表现最佳，其  $R^2$  为 0.79131，MAE 为 250.2452，RMSE 为 339.8425。这表明 GA-LSTM 模型能够更好地拟合训练数据，并且在训练集上达到了更高的准确性。

在测试集上，BP 模型的性能明显下降，其  $R^2$  为负值，为 -0.131715，MAE 为 293.1939，RMSE 为 364.7533。这表明 BP 模型在测试集上的泛化能力较差，可能存在过拟合的情况。相比之下，LSTM 模型在测试集上表现略有提升，其  $R^2$  为 0.59821，MAE 为 183.6398，RMSE 为 215.4342。然而，GA-LSTM 模型在测试集上表现最佳，其  $R^2$  达到了 0.87707，MAE 为 103.7974，RMSE 为 119.1623。这表明 GA-LSTM 模型不仅在训练集上表现良好，而且具有较强的泛化能力，在测试集上也能够取得出色的预测结果。

表 5 模型训练集拟合结果性能指标统计表

模型	R2	MAE	RMSE
BP	0.51284	379.3371	519.543
LSTM	0.67526	294.9144	423.9305
GA-LSTM	0.79131	250.2452	339.8425

表 6 模型测试集拟合结果性能指标统计表

模型	R2	MAE	RMSE
BP	-0.131715	293.1939	364.7533
LSTM	0.59821	183.6398	215.4342
GA-LSTM	0.87707	103.7974	119.1623

## 五、结论与分析

由预测结果分析可见,用遗传算法优化的长短期记忆神经网络模型预测值更接近真实值,因此,本文选用遗传算法优化的长短期记忆神经网络模型对武汉市未来 8 年的月降水量进行预测,预测结果如下表所示。

表 7 武汉市 1-3 月的降水量预测值

年份	一月	二月	三月
2023	1609.585449	528.9425049	786.1481934
2024	1000.371155	715.7216187	558.1605225
2025	589.621521	451.0218506	541.2828979
2026	519.3468628	1185.75647	484.0160522
2027	685.0041504	46.45234299	1106.079224
2028	754.0256348	490.7265625	208.1041412
2029	373.6797485	313.5640869	355.380127
2030	941.6375122	121.1936798	1213.016235

表 8 武汉市 4-6 月的降水量预测值

年份	四月	五月	六月
2023	463.8602905	559.4152222	452.2602539
2024	1400.763062	960.0050659	557.9750366
2025	1410.494385	80.40632629	396.3469849
2026	1050.763306	1195.948364	1129.187134
2027	665.274231	267.4071045	1466.953125
2028	1280.467896	1590.877075	1017.676819
2029	838.0737915	338.177124	76.90693665
2030	639.979248	547.3666382	786.991272

表 9 武汉市 7-9 月降水量预测值

年份	七月	八月	九月
2023	1230.710327	514.6680908	632.6878662

2024	387.7002563	676.4685059	1016.800049
2025	434.6422119	390.7424927	1119.034302
2026	936.6804199	651.90625	973.7875366
2027	751.817627	136.4983673	2264.416992
2028	700.7636108	394.1859741	552.8895264
2029	323.0232544	809.1022339	37.33820724
2030	191.1218414	581.2926025	596.6486206

表 10 武汉市 10-12 降水量预测值

年份	十月	十一月	十二月
2023	866.3701782	1382.031616	917.1861572
2024	870.4503784	197.7863922	562.1123047
2025	683.1790161	297.1955566	419.6344604
2026	586.9915161	127.5888214	1597.214233
2027	1111.486694	1693.324585	712.6224976
2028	689.6236572	932.0308838	532.7225342
2029	175.440567	1226.760498	829.2374268
2030	650.6960449	532.2283936	340.7070923

由上表可以看出，武汉市在 2023 年-2030 年降雨主要集中在 6、7、8 月份，且基本稳定在 1000-3000mm。由可以看出，降雨量同样集中在 6、7、8 月份，且降雨量超过 3000mm 的天数仅 19 天，在整个历史数据中约占 1.6%，所占比极低，故未来八年降雨量基本稳定在 1000-3000mm 的预测结果可信度高，可以作为参考。

由此可见，未来八年内武汉市的降水量几乎不会发生洪灾，但是有关部门要做好防洪的准备，提前做好防洪工作。

## 六、总结

本研究基于国家青藏高原科学数据中心收录的全国 1901-2022 年 500 亿条降水量数据，通过提取指定经纬度范围区域内共计 4 千万条月降水量数据，对武汉市月降水量进行分析和建模。

首先对降水量的空间和时间分布进行了可视化展示，分析了我国降水量的空间分布情况以及武汉市降水量的历史变化趋势，直观的表现了降水量状况。然后，在武汉市范围内，结合经过预处理后的降水量数据，通过对历史数据的划分学习，建立 BP 神经网络预测模型和 LSTM 神经网络预测模型，根据测试集数据验证模型，得出 LSTM 神经网络更适合用于预测长期时序模型的结论。接着使用训练完成的 LSTM 神经网络预测模型预测了武汉市未来 8 年的月降水量。最后，为进

一步提高预测精度，得到一个更加完善的预测模型，我们采用遗传算法，对 LSTM 神经网络预测模型的参数进行了寻优操作，然后通过优化后的 GA-LSTM 模型进行模型评估，验证了经优化后的 GA-LSTM 模型具有更高准确度预测能力的论断。

根据 MAE, RMSE 和 MAPE 等指标发现无论是在长时间的数据预测还是大规模的时间序列数据的处理上，GA-LSTM 模型所得到的结果都要更优一些，对未来降水量的预测效果也更好，为未来降水量的预测研究提供了参考依据。

遗憾的是，由于降水量受到多种因素的影响，仅仅通过降水量的历史规律，很难做到对未来降水量的精确预测，为了更加准确的预测降水量，通常还需要结合气象观测、地理特征等多种方法，本研究仅就武汉市降水量的时序规律进行了相对粗糙的预测，并且所选用的地理区域为武汉市最大边界区域，与实际的武汉市降水量存在些许偏差。

## 参考文献

- [1]柯杰. 基于卷积神经网络的图像描述算法研究[D]. 广西师范大学, 2023. DOI:10.27036/d.cnki.ggxsu.2023.002048.
- [2]欧红师. 基于多级 LSTM 和信息融合的视频行为识别技术研究[D]. 华南理工大学, 2021. DOI:10.27151/d.cnki.ghnlu.2021.005003.
- [3]卢官明, 袁亮, 杨文娟, 等. 基于长短期记忆和卷积神经网络的语音情感识别[J]. 南京邮电大学学报(自然科学版), 2018, 38(05):63-69. DOI:10.14132/j.cnki.1673-5439.2018.05.009.
- [4]唐颖, 曹春萍. 基于 LSTM 的生物医学核心实体提取模型研究[J]. 软件导刊, 2018, 17(05):132-137.
- [5]陈斌, 周勇, 刘兵. 基于卷积双向长短期记忆网络的事件触发词抽取[J]. 计算机工程, 2019, 45(01):153-158. DOI:10.19678/j.issn.1000-3428.0049801.
- [6]秦伟刚. 基于人工神经网络的智能交通系统检测与控制[D]. 山东大学, 2005.
- [7]袁美英, 周秀杰, 陈威. 人工神经网络在温度和降水预报中的应用[J]. 黑龙江气象, 2000, (03):17-19. DOI:10.14021/j.cnki.hljqx.2000.03.005.
- [8]吴新余, 孙力娟. 改进交叉方式的遗传算法在求解通信网优化问题中的应用[J]. 通信学报, 1997, (10):16-22.
- [9]王宏刚, 曾建潮. 遗传算法的改进及应用[J]. 太原重型机械学院学报, 1996, (04):9-14.
- [10]程昊天, 韩曦, 王运智, 等. 人工神经网络的现状与发展——以雾霾预测研究为例[J]. 现代信息科技, 2020, 4(01):20-22.
- [11]曹昕, 马拓, 杨俊坡, 杨耿, 郑兴. 神经网络模型在混凝效果预测中的应用现状及展望[J]. 环境科学学报, 2023, 43(12):186-193.
- [12]陈明金, 欧阳祖熙. 遗传算法的 BP 神经网络对滑坡孔隙水压力时间序列恢复及预测研究[J]. 地壳构造与地壳应力文集, 2006, (00):22-29.
- [13]彭守璋. (2020). 中国 1km 分辨率逐月降水量数据集 (1901-2022). 国家青藏高原数据中心. <https://doi.org/10.5281/zenodo.3185722>.
- [14]中国气象数据网 - WeatherBk Data (cma.cn) <https://data.cma.cn/data/weatherBk.html>

- [15]武汉：百年的洪水记忆-湖北省水利厅 - Hubei [http://slt.hubei.gov.cn/slyw/slwh/slls/201912/t20191213\\_1761914.shtml](http://slt.hubei.gov.cn/slyw/slwh/slls/201912/t20191213_1761914.shtml)
- [16]湖北干旱成因与对策-湖北省水利厅 - Hubei [http://slt.hubei.gov.cn/slyw/dcyj/201211/t20121112\\_967568.shtml](http://slt.hubei.gov.cn/slyw/dcyj/201211/t20121112_967568.shtml)
- [17]Peng, S.Z., Ding, Y.X., Wen, Z.M., Chen, Y.M., Cao, Y., & Ren, J.Y. (2017). Spatiotemporal change and trend analysis of potential evapotranspiration over the Loess Plateau of China during 2011-2100. *Agricultural and Forest Meteorology*, 233, 183-194.
- [18]Ding, Y.X., & Peng, S.Z. (2020). Spatiotemporal trends and attribution of drought across China from 1901-2100. *Sustainability*, 12(2), 477.
- [19]Peng, S.Z., Ding, Y.X., Liu, W.Z., & Li, Z. (2019). 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth System Science Data*, 11, 1931 - 1946.
- [20]Peng, S., Gang, C., Cao, Y., & Chen, Y. (2017). Assessment of climate change trends over the loess plateau in china from 1901 to 2100. *International Journal of Climatology*.



## 附录

表 11 变量名称及解释 (完整)

变量	解释
$x_t$	t 时刻长短期记忆神经网络的输入参数
$n_t$	t 时刻长短期记忆神经网络的输出参数
$W$	输入层和记忆细胞之间的连接权值
$b$	记忆细胞和输出层之间的神经元偏置值
$i_t$	t 时刻的输入门
$f_t$	t 时刻的遗忘门
$o_t$	t 时刻的输出门
$c_t$	t 时刻的记忆细胞所包含的一个输出

表 12 武汉市 1901-2022 年历史月降水量数据集

Time	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1901	357	20	462	1107	1267	2681	5233	844	353	1103	84	102
1902	367	139	838	1634	1313	455	1410	316	63	662	148	490
1903	195	811	1759	1342	1997	1206	2525	1580	1267	263	197	30
1904	158	498	557	3347	1258	874	312	1901	988	1480	246	210
1905	421	556	1504	1342	2252	1080	1975	1849	1199	758	298	935
1906	618	1564	735	845	1092	1900	846	3103	352	132	339	103
1907	722	490	1161	744	1485	1287	2645	680	995	2264	1253	68
1908	625	458	724	2319	1764	2598	2856	638	1328	1117	1339	213
1909	593	220	974	1145	719	3890	3005	1020	658	1694	384	285
1910	718	379	897	2019	1513	2390	2584	1911	527	705	496	135
1911	397	687	1314	1209	1828	2642	1920	1795	1163	761	1213	450
1912	314	730	1638	917	1314	2298	2963	781	433	485	1111	467
1913	250	626	834	2288	1267	2460	1122	122	434	199	953	269
1914	295	809	850	1565	1659	1197	1097	1244	710	1272	1097	222
1915	146	339	565	1735	954	2929	1709	1338	1371	1587	1237	62
1916	230	890	476	1998	1519	2401	2183	1225	687	1014	619	177
1917	135	506	570	946	1067	2080	2394	1333	350	697	366	91
1918	2	390	1113	2138	1266	1183	2111	2299	772	389	1170	948
1919	348	276	1673	898	1412	2109	2678	1157	1371	549	301	121
1920	307	957	1353	817	1527	1235	1700	726	1050	685	903	697
1921	170	314	787	2098	1543	2117	709	1290	1267	930	260	123
1922	502	665	565	990	1316	1699	816	835	136	528	122	84

基于中国百年月降水量大数据的统计建模与预测模型研究

1923	206	651	966	1354	1432	1479	1909	1057	1018	372	344	219
1924	235	552	1077	832	2205	2003	3045	427	486	307	220	55
1925	317	417	819	625	2203	698	781	879	311	355	1030	232
1926	309	476	706	701	1533	1735	2559	527	671	840	404	402
1927	312	549	609	1268	833	1423	2110	1192	1165	335	97	142
1928	404	246	695	1341	1226	1386	853	961	399	75	783	360
1929	493	346	478	631	1921	1564	2582	1697	960	321	188	727
1930	327	393	905	1902	1436	2150	581	1940	805	811	471	190
1931	292	930	600	1705	1754	1535	4640	1210	875	35	463	433
1932	410	433	379	864	2574	1697	785	1162	763	175	166	518
1933	308	428	528	1844	1942	1758	1943	1371	577	1231	515	336
1934	192	516	730	1878	1842	1517	509	857	1101	825	465	493
1935	284	900	1023	961	1504	2116	1510	498	652	945	1268	192
1936	153	709	549	2425	1038	1055	2311	689	540	116	303	782
1937	531	973	679	2532	1956	1477	2359	1546	1262	1213	1227	176
1938	480	271	994	774	1492	3182	2100	1571	730	639	512	303
1939	173	1413	763	1236	1100	2498	3921	1309	464	547	741	5
1940	256	992	876	466	1361	894	835	1198	726	788	305	166
1941	363	650	889	589	2047	800	1196	1155	1423	189	623	356
1942	196	403	1468	1602	2411	2485	1200	732	519	584	424	60
1943	263	492	1026	1300	1689	1653	1541	2008	929	594	843	213
1944	393	541	517	1404	1211	1518	1036	1214	926	650	874	160
1945	352	648	821	1078	977	1702	1944	847	764	534	719	170
1946	140	589	1308	1063	2030	2373	2394	1284	916	342	653	339
1947	436	453	880	1027	1743	1748	1702	922	705	538	399	279
1948	419	981	1566	1109	2698	1741	2155	1026	680	720	368	597
1949	111	1050	1216	1354	2219	1112	1436	1576	1148	268	1165	560
1950	632	528	915	1225	1235	1422	1779	1333	1315	482	619	198
1951	199	661	771	1374	1766	1412	2986	1081	955	566	673	451
1952	264	794	1326	1407	2190	1050	1275	1827	1175	831	395	85
1953	347	639	842	631	1049	1638	1391	1381	702	1202	862	375
1954	820	905	430	1718	3088	3166	5595	1084	253	402	476	809
1955	169	652	1465	1285	1144	2879	993	1875	267	57	164	217
1956	388	236	1586	1871	2479	2525	1409	1541	394	427	28	87
1957	727	467	682	1647	1767	1747	1849	1170	424	466	888	402
1958	354	429	802	2484	1844	788	1315	1750	673	1015	333	254
1959	290	1360	977	1606	2110	2393	497	507	533	516	969	578
1960	305	453	1626	1043	1611	2738	1589	522	927	265	997	67
1961	231	432	1563	930	1145	1300	1002	1143	1436	586	1061	313
1962	163	469	346	1523	1545	1854	2058	1803	941	553	944	370
1963	4	205	916	1787	2164	473	2134	2016	539	206	688	287
1964	481	853	824	2641	2278	2330	1204	631	444	1222	89	191
1965	203	828	679	1874	646	1218	1585	1193	593	818	828	386
1966	251	447	715	1439	1340	1412	1230	182	94	746	453	448

基于中国百年月降水量大数据的统计建模与预测模型研究

1967	283	732	1297	1360	2213	1730	1140	740	964	598	1724	33
1968	386	39	986	934	1460	934	2316	789	709	465	353	952
1969	701	470	826	1926	1457	1680	4651	2087	805	156	441	62
1970	132	659	912	1713	2226	1929	2220	685	1609	483	369	339
1971	313	620	657	1015	1600	2449	617	1053	760	859	441	196
1972	233	770	1470	889	1666	1718	1003	812	743	1502	1263	352
1973	336	950	778	2352	2381	1708	1961	669	1821	324	40	20
1974	447	646	592	1460	2461	1412	1974	649	475	541	528	588
1975	96	614	704	2493	1404	2401	1242	1849	891	1344	325	525
1976	135	1002	712	1288	1326	1963	812	729	489	899	387	193
1977	264	181	1091	2976	1992	1617	2203	1469	730	690	582	538
1978	282	383	825	796	1682	1599	443	579	249	502	699	151
1979	327	289	768	1436	1385	2428	1846	897	1152	4	130	485
1980	424	369	1579	941	1756	2955	2869	2643	464	746	266	43
1981	441	577	927	1258	594	1467	1379	818	594	1623	1048	52
1982	251	694	1094	901	1227	1449	2420	2231	862	457	1347	74
1983	337	277	489	1629	1830	2820	3521	871	1546	2466	306	184
1984	353	174	568	1263	1234	2482	1530	1190	1108	637	732	757
1985	160	661	889	990	2323	1015	1740	515	875	1307	485	257
1986	135	174	763	1504	770	2359	2336	442	891	805	544	369
1987	458	557	1177	1547	1937	1515	2178	2107	437	1407	765	2
1988	165	809	695	457	2431	1386	765	1944	1579	457	21	78
1989	679	1004	852	1990	1513	2384	1762	1281	959	942	954	237
1990	448	1481	1008	1469	1545	1805	1469	888	586	346	1177	387
1991	444	953	1513	1360	2142	2250	3280	1375	511	65	188	490
1992	226	245	2031	1075	1262	1953	1088	560	721	363	211	345
1993	769	1035	1395	993	1975	1850	2233	1145	1090	441	1185	228
1994	237	728	625	1652	1391	1568	1797	975	832	476	617	560
1995	398	384	480	1952	1863	2581	1797	1336	121	1197	24	132
1996	485	173	1268	577	1243	2798	3171	1522	531	764	1263	46
1997	351	697	854	730	1090	1626	2473	638	486	444	1089	585
1998	626	467	1349	2130	2377	1967	2314	1291	429	594	83	293
1999	188	144	843	2126	1698	2652	1174	1608	564	1144	329	31
2000	690	394	377	431	1487	2229	1240	1188	1507	1377	705	284
2001	766	535	479	1316	667	1397	736	818	39	964	427	880
2002	228	754	1218	2390	2269	2352	2225	1266	432	663	590	927
2003	276	1040	1242	2295	1488	2428	2417	928	489	769	740	284
2004	455	531	520	1060	1754	2185	2285	1747	373	82	643	447
2005	278	924	544	1239	1786	1556	1934	1665	1119	373	1089	90
2006	512	789	248	1639	1457	1124	1968	1018	696	392	612	278
2007	343	1065	1157	957	1672	1683	2939	1365	275	247	315	328
2008	665	185	687	1361	1453	1290	2617	2487	372	782	573	56
2009	158	1017	779	1561	1730	2043	1393	846	460	214	845	436
2010	208	589	1384	1734	1632	1299	3337	1186	1216	674	144	216

基于中国百年月降水量大数据的统计建模与预测模型研究

2011	137	322	432	420	866	2828	918	1171	466	629	603	147
2012	238	386	1098	958	1827	1359	1702	1263	1067	777	634	617
2013	216	506	751	1162	1971	1685	1865	1289	1168	107	520	70
2014	198	854	988	1422	1369	1271	2655	1105	1149	842	963	48
2015	216	922	975	1367	1775	2262	2824	825	975	923	1098	337
2016	411	269	540	1683	1764	2452	3303	1068	629	951	954	458
2017	303	499	1247	1385	916	2400	1368	1587	960	829	427	126
2018	642	372	1103	1372	1739	853	1798	1050	618	534	1120	751
2019	485	733	936	1410	1482	2273	1773	411	86	342	414	306
2020	523	800	997	726	1455	2807	4772	958	1565	1092	702	219
2021	128	622	1290	1382	2517	1050	2115	2116	661	482	339	173
2022	786	512	1665	2247	823	2185	2076	196	41	271	982	192

表 13 武汉市 2023-2030 年预测月降水量数据集

Time	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2023	1610	529	786	464	559	452	1231	515	633	866	1382	917
2024	1000	716	558	1401	960	558	388	676	1017	870	198	562
2025	590	451	541	1410	80	396	435	391	1119	683	297	420
2026	519	1186	484	1051	1196	1129	937	652	974	587	128	1597
2027	685	46	1106	665	267	1467	752	136	2264	1111	1693	713
2028	754	491	208	1280	1591	1018	701	394	553	690	932	533
2029	374	314	355	838	338	77	323	809	37	175	1227	829
2030	942	121	1213	640	547	787	191	581	597	651	532	341