



首都經濟貿易大學

CAPITAL UNIVERSITY OF ECONOMICS AND BUSINESS

## 本科生毕业论文（设计）

（二〇二四届）

### 基于融合注意力机制的 LSTM 神经网络 对地下水位预测研究

学生姓名	:	秦大程
学号	:	32020120132
学院	:	统计学院
专业（方向）	:	数据科学与大数据技术
指导教师	:	阮敬

## 摘要

全球有超过 30%的淡水来自地下水，地下水位的高低直接影响着这一宝贵资源的利用和可持续发展。地下水位预测成为应对水资源管理挑战的重要手段之一。对地下水位进行准确预测，可以更好地适应气候变化带来的挑战，保障人类生活和生态环境的可持续发展。

本文对一个流域内的地下水系统中的五个子系统分别进行地下水位的预测，其中这五个子系统分为承压地下水和无承压地下水两种类型。为实现对多个子系统的地下水位的精准预测，以及寻找对地下水位有影响的因素，本文首先对预处理后的数据分别进行 Grange 因果检验和斯皮尔曼相关性系数分析，结合两者的结果，初步对影响因素进行分析，之后根据结果建立一个融合注意力机制的多输入-多输出的 LSTM 神经网络，分别对南部的自流（或承压）地下水和北部的地下水位（或无承压）地下水进行预测，其中模型的超参数通过麻雀搜索算法在划定的范围内，通过比较 R 方确定最优的参数组合，经过模型的训练，通过数据可视化，MSE，MAE 和 R 方四种评价手段对结果进行评价，并且训练了传统的 LSTM 模型，XGBoost 极度梯度提升树，随机森林回归模型三个模型作为对照，用于评价融合注意力机制的 LSTM 神经网络的效果，其中改进后的 LSTM 模型，MSE 达到了 0.0001，MAE 达到了 0.016，R 方达到了 0.958，对同一水系中子系统之间的预测（南部承压地下水），模型的表现为 MSE 达到了 0.0025，MAE 达到了 0.0383，R 方达到了 0.872，相比直接非承压层的预测有一定的下降。相比于作为对照的三个模型无论是承压地下水或是非承压地下水，在模型的准确度和模型的拟合程度上都有较为显著的提升，其中在 MSE 和 MAE 最高减少了约 80%，R 方最高提升了约 35%。

而针对 Grange 因果检验和斯皮尔曼相关性系数分析，分别训练了全变量模型和逐个剔除低价值和无关变量后的变量的模型，将模型通过评价指标进行对比，经过剔除变量后的模型较全变量模型在 R 方提高了约 21%，在模型拟合程度上有较为明显的提高。

**关键词：**地下水位预测，LSTM，注意力机制，Grange 因果检验

---

## ABSTRACT

More than 30 per cent of the world's freshwater comes from groundwater, and the level of groundwater levels has a direct impact on the use and sustainability of this valuable resource. Groundwater level prediction has become one of the most important tools for addressing the challenges of water resources management. Accurate prediction of groundwater level can better adapt to the challenges brought by climate change and guarantee the sustainable development of human life and ecological environment.

In this paper, groundwater level predictions are made for each of the five subsystems of the groundwater system in a basin, which are divided into two types: pressurised and unconfined groundwater. In order to achieve accurate prediction of groundwater levels in multiple subsystems and to find the factors that have an impact on groundwater levels, this paper firstly carries out the Grange causality test and the Spearman correlation coefficient analysis of the preprocessed data, combines the results of the two, and preliminarily analyses the influencing factors, and then establishes a **multiple-input-multiple-output LSTM neural network** incorporating an attentional mechanism based on the results. The prediction of artesian (or pressurised) groundwater in the south and groundwater level (or unpressurised) groundwater in the north were carried out respectively, in which the hyper-parameters of the model were used to determine the optimal parameter combinations by comparing the R-square through the sparrow search algorithm in the delineated range, and after the training of the model, the results were evaluated through the four means of evaluation, namely, data visualisation, MSE, MAE and R-square, and the traditional LSTM was trained. model, XGBoost Extreme Gradient Boosting Tree, and Random Forest Regression Model were trained as controls to evaluate the effect of the LSTM neural network incorporating the attention mechanism, in which the MSE of the improved LSTM model reached **0.0001**, MAE of the improved LSTM model

---

reached **0.016** and the R-square reached **0.958**, especially for the prediction of the sub-systems in the same water system (the Southern Pressurised Groundwater), the performance of the model was greatly improved, in which the MSE reached **0.0025**, MAE reached **0.0383** and the R-square reached **0.872**. Compared with the three control models for both pressurised and non-pressurised groundwater, the model accuracy and model fit were significantly improved, with the MSE and MAE **reduced by up to 80%** and the R-squared improved by **up to 35%**.

For the Grange causality test and Spearman correlation coefficient analysis, the full-variable model and the model with the variables removed one by one after the removal of low-value and irrelevant variables were trained respectively, and the models were compared through the evaluation indexes, and the model after the removal of the variables was **improved by about 21%** in R-squared compared with the full-variable model, which was a more significant improvement in the degree of model fit.

**Keywords:** Groundwater Level Prediction, LSTM, Attention Mechanism, Granger Causality Test

## 目录

一 引言 .....	1
1.1 选题背景及意义 .....	1
1.2 国内外研究现状 .....	2
1.3 研究内容和研究思路 .....	4
1.4 本文主要工作内容 .....	6
1.5 本文主要创新点 .....	7
二 基础理论介绍 .....	8
2.1 Granger 因果检验 .....	8
2.2 斯皮尔曼相关性分析 .....	9
2.3 麻雀搜索算法 .....	10
2.3.1 麻雀初始化 .....	11
2.3.2 麻雀的位置更新机制 <sup>[10]</sup> .....	12
2.4 注意力机制 .....	13
2.5 LSTM 神经网络 .....	15
2.6 模型评价指标 .....	17
三 地下水位及其特征的数据预处理 .....	18
3.1 地下水位及特征流域的数据介绍 .....	18
3.2 数据预处理 .....	22
3.2.1 异常值处理与数据截取 .....	22
3.2.2 缺失值填补 .....	23
3.2.3 平滑异常值 .....	24
3.2.4 异常值检测 .....	25
3.2.5 划分训练集测试集 .....	25
3.2.6 数据标准化 .....	25
3.3 影响地下水位的特征变量选择 .....	26
3.3.1 Granger 因果检验 .....	26
3.3.2 斯皮尔曼相关性检验 .....	28
3.3.3 更正后的特征列 .....	29
四 基于 LSTM 模型预测的地下水位预测 .....	30
4.1 构建基础的 LSTM 模型 .....	30
4.1.1 基础 LSTM 的构建 .....	30
4.1.2 LSTM 神经元数量的选择 .....	31
4.1.3 L1, L2 正则化值的选择 .....	32
4.1.4 Dropout 率的选择 .....	32
4.2 基于融合注意力机制的 LSTM 神经网络预测地下水位 .....	33
五 改进后模型与其他模型的模型评价 .....	37
5.1 对比不同特征选择对改进型 LSTM 模型的影响 .....	37
5.2 改进型 LSTM 模型与其他模型对比 .....	38
5.2.1 改进型 LSTM 与传统 LSTM 模型对比 .....	38

---

5.2.2 改进型 LSTM 与 XGBoost 对比 .....	41
5.2.3 改进型 LSTM 与随机森林回归对比 .....	43
5.2.4 小结 .....	45
六 总结与展望 .....	47
6.1 研究总结 .....	47
6.2 展望 .....	48
参考文献 .....	49
致谢 .....	51

## 一 引言

### 1.1 选题背景及意义

随着全球气候变化的影响日益显现，地下水位预测成为应对水资源管理挑战的重要手段之一。气候变化可能导致降雨模式和蒸发蒸腾的变化，从而影响地下水的补给和消耗过程。因此，通过对地下水位进行准确预测，可以更好地适应气候变化带来的挑战，保障人类生活和生态环境的可持续发展。

据统计，全球有超过 30% 的淡水来自地下水，而地下水位的高低直接影响着这一宝贵资源的利用和可持续发展。

以我国为例，根据中国地下水监测数据显示，近年来全国地下水位呈现不同程度的下降趋势，其中包括北方地区的平原地带和南方的一些盆地地区。据《中国地下水水文监测年鉴》显示，部分地区地下水位下降速度已超过每年 0.5 米，严重威胁着当地的饮用水安全和农业灌溉需求。

此外，地下水位预测还对灌溉农业具有重要意义。农业是地下水的主要利用领域之一，而灌溉是农业生产中的重要环节。根据农业部的统计数据，中国农业用水占全国总用水量的 70% 以上，其中地下水占比较大。地下水位的下降使得部分地区的农业灌溉面临严峻挑战，据估算，中国每年因地下水位下降导致的农田面积减产达数百万亩，直接影响着粮食和农产品的生产供应。通过合理预测地下水位，可以更好地制定灌溉计划，优化农田水利设施的布局，提高水资源利用效率，从而保障粮食安全和农民收入。

另外，地下水位的变化也直接影响土壤的湿度和养分分布，进而影响作物的生长和产量。合理预测地下水位可以帮助农民调整种植结构，选择适应性更强的作物品种，降低农业生产的风险，提高农业生产的稳定性和可持续性。

在城市规划和土地利用方面，地下水位预测也起着重要作用。合理预测地下水位可以避免城市基础设施因地下水位变化而引发的问题，据中国城市地下水位监测数据，一些发达地区的城市地下水位也呈现着不同程度的下降趋势。这给城市基础设施的建设和运行带来了挑战，据有关专家预测，未来数十年内，城市地下水位的进一步下降可能会加剧地下管道、地铁和建筑物的沉降和损坏，

---

增加城市维护和修复的成本。

综上所述，加强地下水位的监测和预测显得尤为重要。通过建立科学的地下水位预测模型，并结合大数据和先进技术手段，可以更准确地预测地下水位的变化趋势，对水资源管理、灌溉规划和生态保护具有重要意义，还涉及到农业生产、城市规划等多个领域，对社会经济发展和人类生活具有深远影响。因此，开展地下水位预测的研究具有重要的理论和实践意义。

利用深度学习技术预测地下水水位，可以更好地理解和应对气候变化对地下水资源的影响，为相关应对措施的制定提供科学依据。

深度学习预测模型能够实现对地下水水位变化的精准预测，并能够及时地提供预测结果和分析报告。这些信息可以为政府部门、水利管理机构和决策者提供科学依据，支持其制定相关的水资源管理政策和应对措施。

## 1.2 国内外研究现状

地下水数据本质就是时间序列预测问题，地下水位受天气影响十分明显，特征数据均来自于长期检测记录的传感器数据，其中主要进行长期预测，主要用于对地下水和水资源的季节调配和预警。

国外目前采用各种数学建模和数据驱动的方法，如机器学习、人工神经网络<sup>[17]</sup>等，利用地下水位监测数据进行模型训练和预测。例如，一项由美国加州大学伯克利分校的研究<sup>[15]</sup>指出，在加州中部，研究人员利用机器学习算法分析了大量的气象和地下水位监测数据，成功预测了未来数月的地下水位变化情况，准确率超过 80%。学者 Emery A. Coppola Jr<sup>[16]</sup>利用人工神经网络（ANN）通过气候数据预测采样井的水位高度，获得了出色的预测能力。

在国内，对地下水位预测的研究现状显示出多样化和前沿性。国内研究者不仅致力于建立适用于中国地区的地下水数学模型和模拟技术，还开始尝试应用机器学习<sup>[18]</sup>、人工神经网络<sup>[6]</sup>等方法，对地下水位进行预测。部分学者采用水均衡法（Water Balance Method）<sup>[2]</sup>，逻辑回归等方法对地下水资源进行估算，取得较好的预测的效果，但主要考虑的为线性关系，而对其中的非线性关系有

---

所忽略。

另一方面，国内研究者也开始采用机器学习<sup>[3]</sup>和人工神经网络<sup>[9]</sup>等数据驱动的方法进行地下水位预测。这些方法利用大量的地下水位监测数据，通过训练模型来发现数据中的潜在规律，并建立高效的预测模型。目前，国内大多采用了注意力机制和 LSTM（长短期记忆网络）网络<sup>[12]</sup>的融合预测，如刘瑞学者<sup>[5]</sup>基于自注意力机制的 ConvLSTM 与 TCN 河流水位中长期预测模型<sup>[7]</sup>，以提高模型的预测能力和稳定性。还有王利军学者<sup>[8]</sup>基于 BP 神经网络模型，针对低渗透性含水层进行预测的方法，通过多个具有简单处理功能的神经元的复合作用，使网络具有非线性映射能力，并通过信息的正向传递与根据误差的逆向修改网络权重的过程来调整参数，来实现输入和输出之间的映射关系，具有广泛的实际应用性。同时也有使用极端梯度提升树（XGBoost）模型对地下水位进行预测，深入分析了冻融期潜水水位动态。

中国科学院水利部水文水资源研究所的研究团队通过建立基于地下水数学模型的预测系统，成功预测了黄土高原地区未来一年的地下水位变化趋势<sup>[11]</sup>，并提出了相应的水资源管理建议，为当地农业灌溉和生态环境保护提供了重要支持。

最后，将地下水位预测与决策支持系统相结合，为水资源管理者和决策者提供实时的、个性化的预测结果和建议，以制定科学合理的水资源管理策略。

### 1.3 研究内容和研究思路

本文基于从 kaggle 竞赛中下载的数据，记录了从 1998 年 3 月 5 日到 2020 年 6 月 30 日，每天的观测数据，共 8155 条数据，27 个特征列。具体内容如表 1.1 数据。

表 1.1 原数据内容

原列名称	列名称	数据含义及单位
Date	日期	该地下水流域内 10 各测量点每日各个降水测量数据
Rainfall*10	降雨*10	单位：毫米 mm
Depth_to_Groundwater_LT2	地下水深度（南部）	为本次预测的目标列，水体由两个子系统组成，分别称为北部和南部，前者在一定程度上影响后者的行为。
Depth_to_Groundwater_SAL		
Depth_to_Groundwater_PAG		
Depth_to_Groundwater_CoS		
Depth_to_Groundwater_DIEC	地下水深度（北部）	北部地区的水平由 SAL、PAG、CoS 和 DIEC 井的值表示，而南部地区的水平则由 LT2 井表示。
Temperature*4	温度*4	流域内气温监测数据点每日数据
		单位：摄氏度 °C
Volume*5	水体体积*5	由于流域内水体会互相相互补给，所以对测量其他水体体积十分重要
Hydrometry*2	水文监测*2	监测水体表面的蒸发和植被蒸腾的水量。数据通常通过气象站等设备来获取。测量水体的水位高度或水位变化情况，通过水位传感器等设备进行监测。

研究内容为地下水位预测和影响地下水位特征价值的筛选，经由图 1.1 所示的研究路线，在对数据进行经过异常值处理，剔除和填补缺失值，平滑异常值，数据标准化几步预处理之后，经过 Grange 因果检验，和斯皮尔曼相关性分析特征筛选之后，使用麻雀搜索算法对模型中的超参数进行筛选，最后使用传统 LSTM，改进型 LSTM，XGBoost，随机森林回归四个模型分别训练对南北两个子系统的地下水位预测模型，再通过测试集数据，验证模型在从未出现过的数据的表现情况，对地下水位进行预测，通过 R 方，MSE，MAE 三种评价指标对模型进行评价，筛选出准确度高的模型进行预测，以便相关部门对水资源的了解和预判更加充分，提供更多依据。

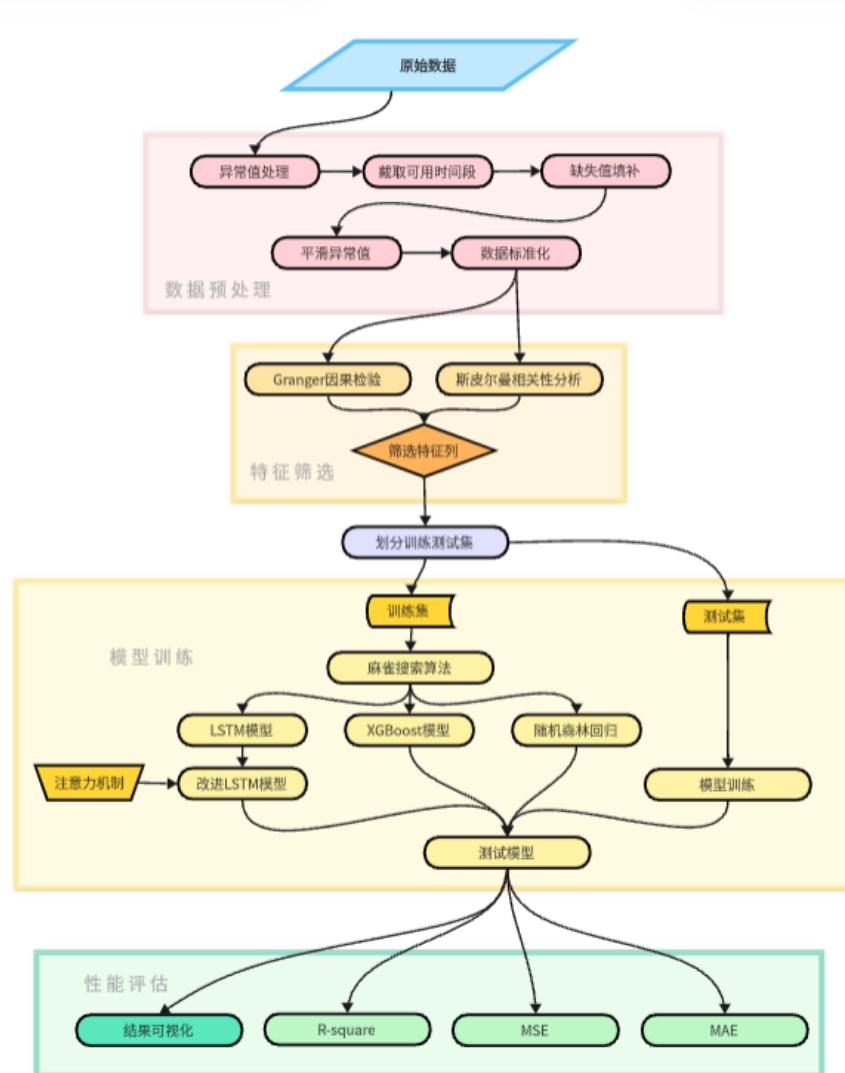


图 1.1 研究思路及内容



## 1.4 本文主要工作内容

本文第一章为绪论，阐述了本文选题背景及意义，以及国内外研究现状，提出研究的内容及主要思路，通过融合注意力机制的 LSTM 神经网络对地下水位进行预测。

第二章进行详细理论介绍，对本文中使用的 Grange 因果检验，斯皮尔曼相关性系数，LSTM 长短期记忆人工神经网络，注意力机制，这些方法中的数学原理以及数学公式进行原理阐述，为其后面在数据中的应用做好准备铺垫。

第三章为数据预处理，对本文使用的数据集进行处理，以满足模型输入数据的需求，并且保证模型最后预测结果的可用性和准确率，在这里详细阐述了处理的逻辑和原理，在处理数据时本文采用了插值填补，移动平均异常值，MinMax 标准化的方式对数据进行预处理。并且由于环境数据中含有大量的噪声和不相关的特征值，需要对其中的数据进行筛选，本文采用了 Grange 因果关系和斯皮尔曼相关性系数，来检验变量之间的非线性和线性关系，对数据集内的特征变量进行了筛选，达到了提高模型准确率的效果。

第四章为模型的构建，本文主要构建了一个融合了注意力机制的 LSTM 神经网络模型，通过流域内的气象数据和水文数据构建模型，完成了对一个水系的两个子系统的地下水位进行预测，从而可以研究地下水位和环境因素之间的关系，更好的了解和利用水资源。

第五章为模型的评价，通过数据可视化，R-Square，MSE，MAE 四种评价方式对所构建的神经网络，将改进后的 LSTM 模型与传统的 LSTM 模型、XGBoost、随机森林回归进行比较，验证了模型的可靠程度，完成了对南北两区一年地下水位的预测，通过可视化的方式，对结果进行了展示。

## 1.5 本文主要创新点

1. 本文通过融合注意力机制，提高了模型对于不同时间步的输入数据的关注程度和不同特征变量之间的影响的权重，同时也可以让模型在不同时间步之间和特征权重之间更加灵活地调整权重从而更好地捕捉地下水位变化的规律性和趋势，增强了预测的准确性和稳定性。
2. 本文采用 Grange 因果关系和斯皮尔曼相关性系数对数据集中的特征变量进行筛选，这两种方法的结合不仅考虑了特征变量之间的线性关系，还能够发现潜在的非线性关系，有效地提高了模型对于关键特征的识别能力，减少了噪声和干扰特征的影响，优化了运算的速度，进而提升了预测模型的准确性同时提高了运行的效率，减少了计算资源的消耗。
3. 在超参数的搜索上选择了麻雀搜索算法，相比于网格搜索增加了随机性，避免了陷入局部最优解的问题。

## 二 基础理论介绍

### 2.1 Granger 因果检验

Granger 因果检验是一种统计方法，用于检验两个时间序列之间是否存在因果关系。该检验由 Clive Granger 于 1969 年提出，被广泛用于经济学、金融学<sup>[14]</sup>和其他领域的时间序列分析中。

Granger 因果检验的基本思想是：如果一个时间序列能够帮助预测另一个时间序列的未来值，那么我们可以认为第一个时间序列对第二个时间序列有因果影响。

Granger 因果检验是用来确定时间序列数据中一个变量是否因果影响另一个变量<sup>[18]</sup>。它建立了一个模型将一个时间序列的过去值预测另一个时间序列的当前值。这可以通过向一个模型中添加或删除特征列来实现，然后利用 F 检验来评估这些列的整体显著性。

使用 F 检验确定添加或删除某个特征列是否会显著改变模型的性能。然后 F 检验来检验线性回归模型中的系数是否显著不为零。

当变量之间的关系具有一定的延迟效应时，Grange 可以检测到一些非线性关系。

Granger 因果检验因果关系检验的步骤：

**第一步：**将当前的  $y$  对所有的滞后项  $y$  以及别的变量（如果存在）做回归，即  $y$  对  $y$  的滞后项  $y_{t-1}, y_{t-2}, \dots, y_{t-q}$  及其他变量的回归，但在这一回归中没有把滞后项  $x$  包括进来，这是一个受约束的回归（1）。然后从此回归得到受约束的残差平方和（2）（Residual Sum of Squares, RSS）。

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_q y_{t-q} + \varepsilon_t \quad (1)$$

$$RSS_c = \sum_{t=q+1}^T \hat{\varepsilon}_t^2 \quad (2)$$

**第二步：**做一个含有滞后项  $x$  的回归，即在前面的回归式中加进滞后项  $x$ ，这是一个无约束的回归（3），由此回归得到无约束的残差平方和（4）。

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_q y_{t-q} + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \varepsilon_t \quad (3)$$

$$RSS_u = \sum_{t=q+1}^T \hat{\varepsilon}_t^2 \quad (4)$$

零假设是  $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ , 即滞后项  $x$  不属于此回归

**第三步：** 使用 F 检验 (5) 来检验 Granger 因果关系的假设

$$F = \frac{(RSS_c - RSS_u)/q}{RSS_u/(n-k)} \quad (5)$$

F 检验的统计量遵循自由度为  $q$  和  $(n - k)$  的 F 分布, 其中  $n$  是样本容量,  $q$  是  $y$  滞后项的个数 (即有约束回归方程中待估参数的个数),  $k$  是无约束回归中待估参数的个数。

**第四步：** 对于选定的显著性水平  $\alpha$  上计算的 F 值进行比较, 如果滞后项  $x$  的 F 值显著的大于临界值  $F_\alpha$ , 则可以拒绝原假设, 这样滞后项就属于此回归, 从而表明  $x$  是  $y$  的原因。在本文中显著性水平  $\alpha$  为 0.05。

## 2.2 斯皮尔曼相关性分析

斯皮尔曼相关性 (Spearman correlation) 是一种非参数统计方法, 用于衡量两个变量之间的单调关系, 特别是当变量的关系不是线性的时候, 两个变量之间的关系是按照一个方向变化的。具体来说就是如果一个变量的值随着另一个变量的增加而增加, 或者随着另一个变量的增加而减少。它基于变量的秩次而不是原始的数值。斯皮尔曼相关系数经常被称作“非参数”的。首先, 当 X 和 Y 的关系是由任意单调函数描述的, 则它们是完全皮尔逊相关的。与此相应的, 皮尔逊相关系数只能给出由线性方程描述的 X 和 Y 的相关性。其次, 斯皮尔曼不需要先验知识便可以准确获取  $x$  和  $y$  的采样概率分布。

斯皮尔曼相关系数被定义成等级变量之间的皮尔逊相关系数。对于样本容量为  $n$  的样本,  $n$  个原始数据被转换成等级数据, 相关系数  $\rho$  为 (6)

$$\rho = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot (y_i - \bar{y})^2}} \quad (6)$$

斯皮尔曼相关性系数计算过程中的一个重要步骤是数据的秩次转换。原始数据依据其在总体数据中平均的降序位置, 被分配了一个相应的等级, 如下表 2.1 所示内容, 变量  $x_i$  通过秩次转换, 最终变成等级  $x_i$ 。完成秩次转换后, 就可以计算这些秩次之间的皮尔逊相关系数, 从而得到斯皮尔曼相关性系数。

通过秩次转换，斯皮尔曼相关性系数能够更准确地衡量两个变量之间的单调关系，而不仅仅是线性关系，更加适用于时间序列数据。

表 2.1 斯皮尔曼相关性系数秩次转换

变量 $x_i$	降序位置	等级 $x_i$
0.8	5	5
1.2	4	3.5
1.2	3	3.5
2.3	2	2
18	1	1

实际应用中，变量间的连结是无关紧要的，于是可以通过简单的步骤计算  $\rho$  被观测的两个变量的等级的差值，则  $\rho$  为(7)

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (7)$$

### 2.3 麻雀搜索算法

麻雀搜索算法<sup>[19]</sup> (Sparrow Search Algorithm, SSA) 是基于模拟麻雀觅食行为的群体智能行为。通过模拟麻雀在寻找食物时的行为，来解决优化问题。麻雀搜索算法基于概率性和启发式搜索，在搜索空间中寻找全局最优解。该算法由 Ahmed M. Kamrul, Shovon, S. M. Abdullah, Tanzila Islam, 和 Kamrul Islam 教授于 2019 年提出。

相比于网格搜索，麻雀搜索算法具有一定的随机性<sup>[1]</sup>，能够根据搜索过程中的局部信息进行调整，避免陷入局部最优解或搜索空间的某些区域，并且会更加的高效。而网格搜索可能会受限于预先定义的参数网格，难以在搜索过程中动态调整搜索方向。

### 2.3.1 麻雀初始化

SSA 中麻雀种群分为觅食发现者，抢食物的加入者，两者角色可以互换。同时还选择了一定比例的麻雀作为侦察的警戒者，一有危险便飞向别处。麻雀种群在 d 维空间内 n 只麻雀的位置矩阵  $X$  及相应的适应度矩阵  $Fx^{[4]}$ 。

$X$  表示如下：

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix} \quad (8)$$

$Fx$  表示如下：

$$Fx = \begin{bmatrix} f_x([x_{11}, x_{12}, \dots, x_{1d}]) \\ f_x([x_{21}, x_{22}, \dots, x_{2d}]) \\ f_x([\dots, \dots, \dots, \dots]) \\ f_x([x_{n1}, x_{n2}, \dots, x_{nd}]) \end{bmatrix} \quad (9)$$

假设需要搜索的变量个数为 10，则  $d = 10$ ， $n$  则为麻雀的数量，麻雀的数量为搜索空间中可能存在的解的数量，麻雀的数量并不一定，但与搜索的迭代次数和搜索空间有关，当迭代次数增加时，搜索空间中可能存在的解的数量可能也会增加，但是迭代次数的增加同时会伴随着程序的复杂程度的提高，同时也受限于计算资源，所以应当进行平衡，选择一个合适的迭代次数。

在公式 (8-9) 中， $x_{ij}$  表示第  $j$  维空间内第  $i$  只麻雀的位置， $x_{ij}$  为一个变量的参数值，而在公式 (9) 中， $f_x([x_{11}, x_{12}, \dots, x_{1d}])$  表示第  $i$  只麻雀的适应度值。这里本文采用 R 方 (R-Square) 作为评价指标，所以这里的适应度即为 R 方的值。

### 2.3.2 麻雀的位置更新机制<sup>[10]</sup>

整个更新机制分为三部分，为发现者觅食过程，消化食物过程，迁徙和更新最优解过程。同时设定最大迭代次数，用于搜索最优解。流程如下图 2.1 所示。

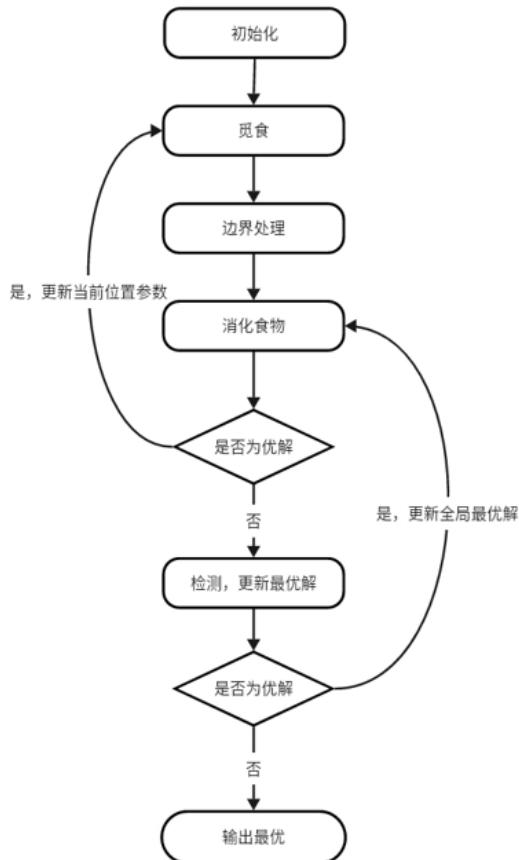


图 2.1 麻雀搜索算法位置更新机制

#### 1. 觅食过程

首先，由发现者负责觅食和指导整个种群移动，其中对当前位置的每个参数执行一些更新操作。这个过程被称为“觅食过程”。通过设定两个随机数  $\alpha$  和  $Q$  来更新当前位置的参数值，并且设置  $ST$  为警戒阈值，一般取值为 0.6<sup>[10]</sup>，取随机数，当随机数小于 0.6 时，使用指数衰减函数更新速度和位置（10-11）；否则，使用正态分布的随机数来更新位置（12）。

设  $itermax$  为最大迭代次数， $v$  为速度， $Q$  为正态分布的随机值， $position$  为当前位置。

当  $ST < 0.6$  时，

$$v = v \times \exp(-\alpha \times itermax) \quad (10)$$

$$position = position + v \quad (11)$$

否则，

$$position = position + Q \quad (12)$$

## 2. 边界处理

确保更新后的参数的值在参数空间内。

## 3. 消化食物过程

遍历当前位置的每个参数。如果当前参数值在一段时间内比最优位置对应的参数值更好，就尝试通过内部参数调整改善当前位置的效果，朝着更好的方向改善，否则保留当前解。这样可以及时更新全局最优解，在搜索过程中更好地指导搜索方向，以尽可能找到更优的解决方案。

## 4. 检测过程

设置概率为 50%，对当前位置有概率的进行评估，并根据评价函数计算其适应度，这样可以增加搜索的多样性，避免陷入局部最优解。引入随机性，算法有更高的可能性跳出局部最优解，并继续搜索全局最优解，有助于提高算法的全局搜索能力。

## 5. 更新过程

如果当前参数的适应度优于全局最优适应度，则更新全局最优位置和适应度。

## 6. 经过多次迭代，最终输出最优参数。

## 2.4 注意力机制

注意力机制是一种在机器学习和人工智能领域中常用的技术，它模拟了人类在处理信息时的注意力过程。该机制使得模型能够在处理数据时有选择性地关注其中的某些部分，而忽略其他部分，从而提高模型的性能和效率。注意力机制通常被用于引进处理序列数据。它可以帮助模型在处理长序列时更好地捕捉到序列中的重要信息，而不是简单地对整个序列进行处理。

注意力机制的核心思想是根据输入的上下文动态地分配不同部分的权重，在处理信息时更加集中地关注与当前任务相关的部分。在本文中，这种机制可以帮助模型在预测时更加关注需要预测的相同时间段的往年数据或临近数据，增加模型预测的精确度，提高模型的性能和泛化能力。

在本文中主要引用了两个方面的注意力机制：

一方面，注意力机制会更加关注于当前预测任务中对目标列影响力较高的特征列，增加该列在预测结果之中的权重，并且在训练的过程中不断地更新权重矩阵的参数，另一方面，注意力机制同样也会更加专注于当前任务上时间序列的关系，对于一些对于当前预测任务中的高价值数据，例如，历史临近数据的注意力和前几年同样时间段的数据等，注意力机制会不断的更新不同时间步的权重矩阵，使得高价值数据会比其他时间段的数据更加的专注。

第一步，注意力权重计算：首先，通过将输入数据经过权重矩阵和偏置项的线性变换，得到特征维度上的注意力得分 ( $et_{feature}$ ) (13)，以及时间维度上的注意力得分 ( $et_{time}$ ) (14)。然后，将两者相加，并添加偏置项，提高模型的灵活性，得到最终的注意力得分 ( $et$ ) (15)。

第二步，注意力权重归一化：利用  $softmax$  函数对注意力得分进行归一化 (16)，以确保注意力权重的和为 1，并且保留了各个时刻的相对重要性。

第三步，加权求和：将归一化的注意力权重与输入数据相乘，得到加权后的输出 (17)。这样，输出中的每个时间步都受到了对应时刻的注意力权重的影响，突出了重要时刻的信息。

设输入数据为  $x$ ，特征维度的权重矩阵为  $W_{feature}$ ，时间维度的权重矩阵为  $W_{time}$ ，偏置项为  $b$ ，数学公式如下：

$$et_{feature} = \text{ReLU}(x \cdot W_{feature}) \quad (13)$$

$$et_{time} = \text{ReLU}(t \cdot W_{time}) \quad (14)$$

$$et = et_{feature} + et_{time} + b \quad (15)$$

$$at = softmax(et) \quad (16)$$

$$output = x \odot at \quad (17)$$

## 2.5 LSTM 神经网络

长短期记忆网络（Long Short-Term Memory, LSTM）是一种循环神经网络（Recurrent Neural Network, RNN）的变体，专门设计用于处理和预测时间序列数据，尤其是具有长期依赖关系的序列数据。

LSTM 解决了 RNN 中存在的存在梯度爆炸和消失的问题，以及对于长距离的学习效果不好的问题。

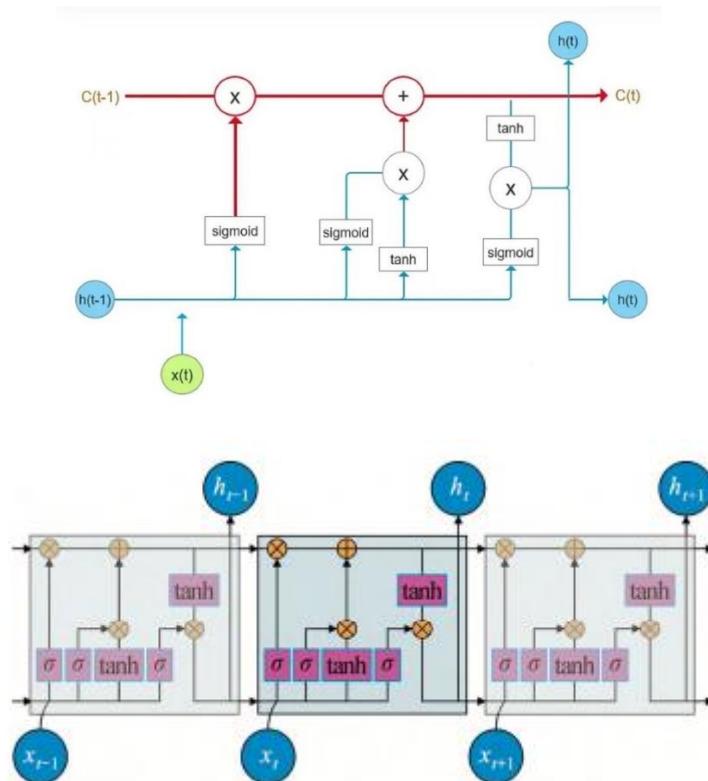


图 2.2 LSTM 示意图

LSTM 核心是细胞状态，穿过图 2.2 的顶部的长横线，长直线称之为细胞状态（Cell State），决定什么样的信息会被保留，什么样的信息会被遗忘，记为  $C_t$ 。

LSTM 网络能通过一种被称为门的结构如图 2.3，对细胞状态进行删除或者添加信息。门能够有选择性的决定让哪些信息通过。门的结构为一个 sigmoid 层和一个点乘操作的组合，sigmoid 层输出 0 到 1 之间的数，描述每个部分有多少量可以通过，0 代表不允许任何量通过，1 表示允许任何量通过。

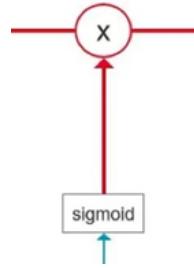


图 2.3 LSTM 门单元示意图

LSTM 实现了三个门计算，即遗忘门、输入门和输出门，用来保护和控制细胞状态。

遗忘门（Forget Gate）：负责决定保留多少上一时刻的单元状态到当前时刻的单元状态，计算公式如 (18)， $h_{t-1}$  为前一时间步的隐藏状态， $x_t$  为当前时间步的输入。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (18)$$

输入门（Input Gate）：负责决定保留多少当前时刻的输入到当前时刻的单元状态，包含两个部分，第一部分为 sigmoid 层，如公式 (19)，第二部分为 tanh 层，如公式 (20)，该层把需要更新的信息更新到细胞状态里。并创建一个新的细胞状态值向量  $\tilde{C}_t$ 。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (19)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (20)$$

并将遗忘门  $f_t$  与旧细胞状态  $C_{t-1}$  相乘加新细胞状态向量  $\tilde{C}_t$  与  $i_t$  相乘，更新旧细胞状态，得到  $C_t$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (21)$$

输出门（Output Gate）：负责决定当前时刻的单元状态有多少输出，通过一个 sigmoid 层来确定细胞状态的哪个部分将输出出去。把细胞状态通过 tanh 进行处理，得到一个 -1 到 1 之间的值，并将它和 sigmoid 门的输出相乘，最终输出确定输出的部分。

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (22)$$

$$h_t = o_t \times \tanh(C_t) \quad (23)$$

LSTM 的参数训练算法与 RNN 相同，依然是反向传播算法。

第一步：前向计算每个神经元的输出值。对于 LSTM 而言，依据前面介绍的算法，分别进行计算。

第二步：确定优化目标函数。在训练前期，输出值和预期值会不一致，于是计算每个神经元的误差项值，构造出损失函数。

第三步：根据损失函数的梯度指引，更新网络权值参数。与传统 RNN 类似，LSTM 误差项的反向传播包括两个层面：一个是空间上层面的，将误差项向网络的上一层传播。另一个是时间层面上的，沿时间反向传播，即从当前 t 时刻开始，计算每个时刻的误差。

跳转第一步，重复做第一、二和三步，以此类推。

## 2.6 模型评价指标

对于回归问题中，通常采用 MSE 和 MAE 作为误差评价指标，采用 R-Square 作为模型精确度评价指标。用 MSE 和 MAE 衡量模型的预测误差大小，值越接近于 0 则表示模型预测越准确。R 方是一种衡量统计模型拟合程度的常用指标之一，R 方的取值范围在 0 到 1 之间，越接近 1 表示模型拟合得越好，0 表示模型不具备任何预测能力。选择 MSE、MAE 和 R 方可以从模型精确度和模型拟合程度两个角度，更加客观准确的来评价模型的拟合程度。

公式如下：

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{true,i} - y_{pred,i})^2 \quad (24)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |(y_{true,i} - y_{pred,i})| \quad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{true,i} - y_{pred,i})^2}{\sum_{i=1}^n (y_{true,i} - \bar{y})^2} \quad (26)$$

其中  $y_{true,i}$  代表第 i 个预测结果的真实值， $y_{pred,i}$  代表第 i 个预测结果的预测值。

### 三 地下水位及其特征的数据预处理

#### 3.1 地下水位及特征流域的数据介绍

本文数据均来自于 ACEA 集团，ACEA 集团是意大利一家综合性的能源、环保和基础设施集团。数据颗粒为一天，其中记录了 Auser 流域地区从 1998 年 3 月 5 日到 2020 年 6 月 30 日每天的观测数据，包括降雨，地下水深度，温度，水体体积，水文监测 5 类数据的详细信息，共 8155 条数据，27 个特征列。其中地下水深度数据来自于 5 个不同的地下水位探测点，作为本文的目标列。

Auser 是意大利托斯卡纳大区的一个地区<sup>[17]</sup>，如图 3.1。Auser 河流经该地区，它的源头位于阿尔卑斯山脉，流经卢卡 (Lucca) 和皮斯托亚 (Pistoia) 等城市，最终注入地中海。

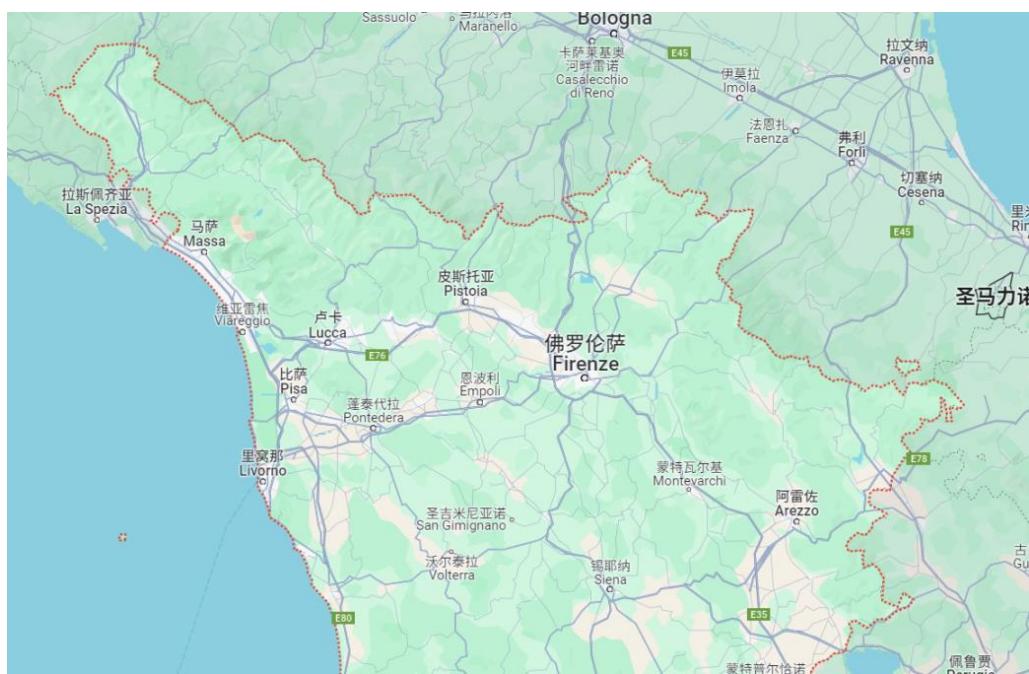


图 3.1 流域地图

本文将 5 列地下水探测数据作为目标数据，该流域由两个子系统组成，称为“北”和“南”，北部子系统是地下水位（或无承压）含水层，而南部子系统是自流（或承压）地下水。前者部分影响后者的行为。

“北”扇区的液位由 SAL、PAG、CoS 和 DIEC 井的值表示，而“南”扇区的液位由 LT2 井的值表示。

## 1、地下水位数据

下图 3.2 中展示了该流域 2006 年-2020 年地下水位的数据，可以明显地看出数据中存在明显的季节性，周期性，周期为年。并且目标列中存在缺失值。

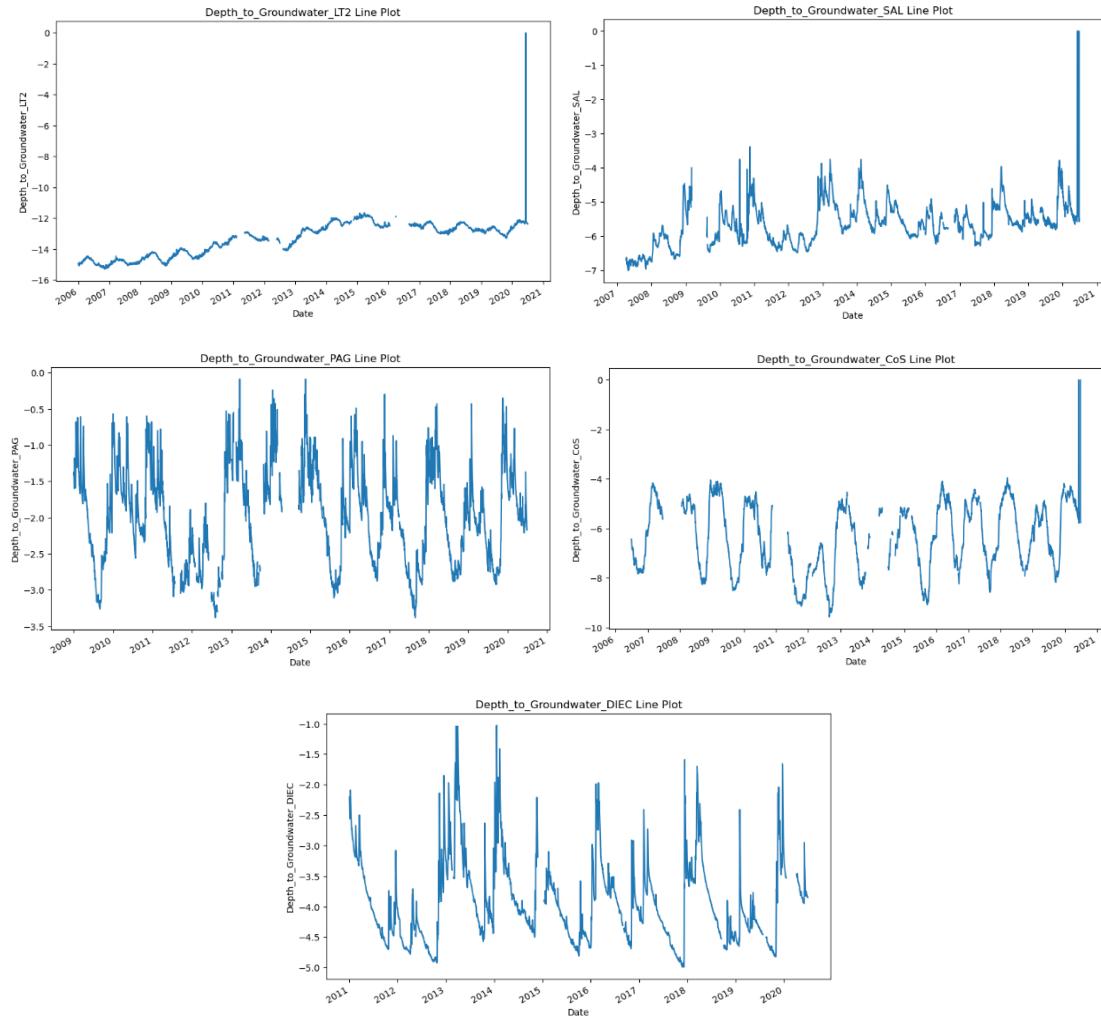


图 3.2 地下水位数据

## 2、气象数据

下图 3.3 为北部 4 个气温观测点的气象数据，由图轻松可知，有明显的季节性和周期性，

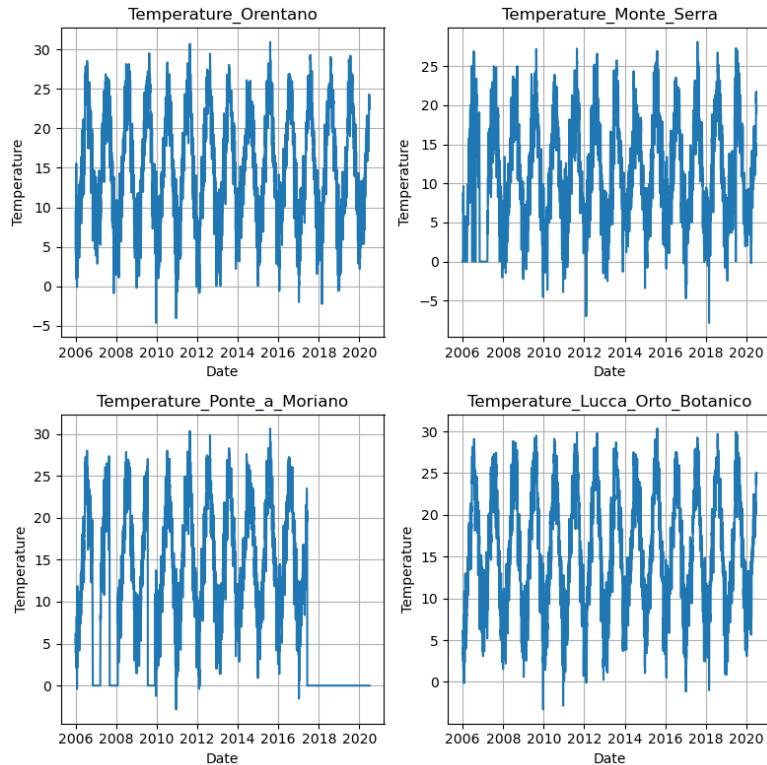


图 3.3 气温数据

由气温图像可知，该流域气温基本均高于 0 摄氏度，积雪量少，不存在春季冰雪融水对地下水位的补给，导致降雨量对地下水的补给没有滞后性。

### 3、水文观测数据

该流域共有 10 个降水量监测点，具体数值由下图 3.4 所示，

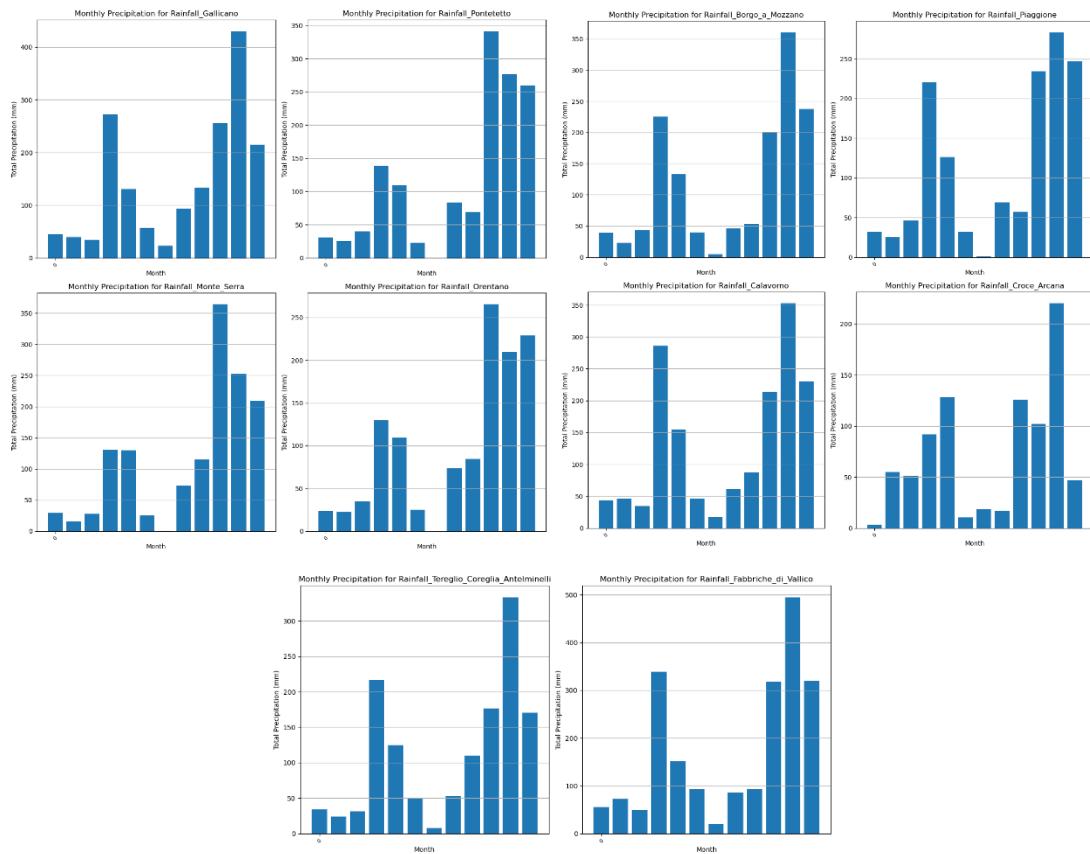


图 3.4 水文数据

由图 3.4 可知，该地区降水季节性较强，年降水量大，每年降水主要集中在 11 月-次年 3 月，为典型的地中海气候，夏季高温少雨，冬季寒冷多雨。

由于当地河流的主要补给方式为季节性降水补给，而地下水的主要补给方式为河流水等地表水的下渗，所以降水量的影响对于地下水位的影响会尤为的突出，同时在后续的 Grange 因果检验和斯皮尔曼相关性检验中可以进行验证，所以在后续数据预处理的时候，应当保留更加完善和更多的细节信息，可以使模型在预测的能力上更加广泛，学习更多的细节信息，提高模型的预测能力。

### 3.2 数据预处理

#### 3.2.1 异常值处理与数据截取

由于在 2011 年以前，部分特征列数据存在大量的缺失和明显数值为 0 的异常值，如图 3.5。

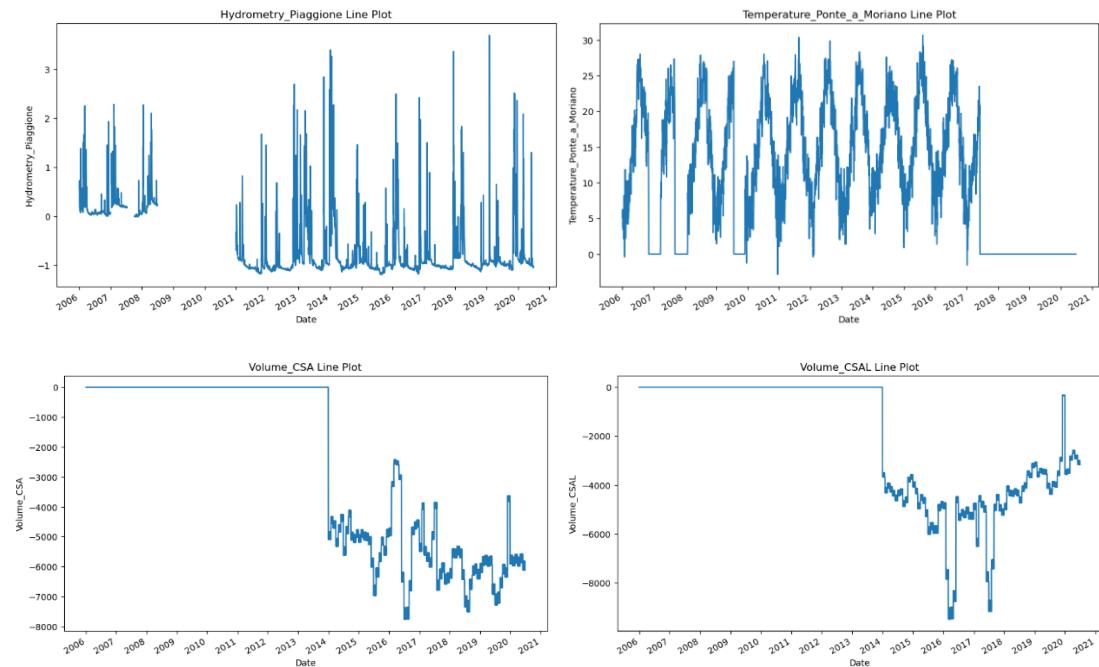


图 3.5 缺失数据

所以在平衡数据的可用程度和缺失程度来说，对缺失值较多的两列（Volume\_CSA Line Plot, Volume\_CSAL Line Plot）数据进行删除，对多列数据中，同一时间段上缺失值较为集中的数据，对数据集时间段进行截取，截取了 2011-08-31 到 2019-12-31 之间的数据，以保证数据集的可用性和特征数量，提高模型的精准度。

同时对数据中一些反常规的数据的数值进行判定，如：降水量为负值，天气气温长时间为同一温度等，在处理缺失值时将数据的值置为空，等待后续在缺失值填补时进行填补，提高数据的准确性。

### 3.2.2 缺失值填补

在截取后数据中依然含有缺失值，同时也有异常值识别之后产生的缺失值数据，需要对数据继续进行缺失值填补。

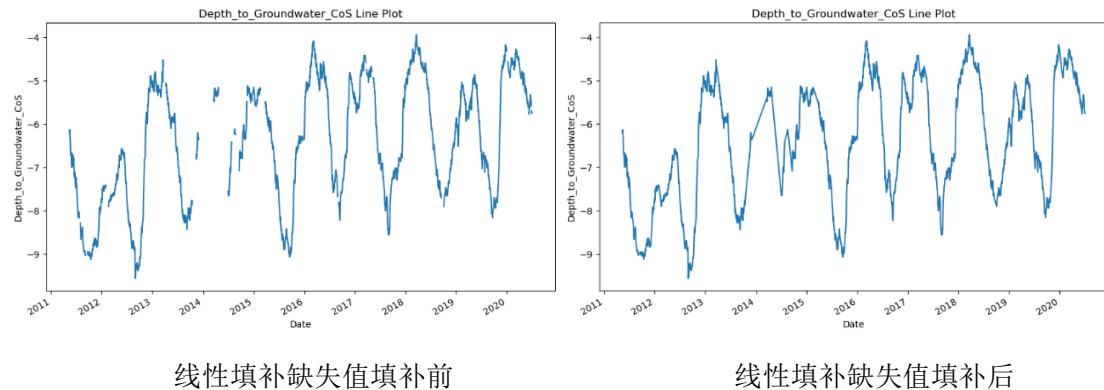


图 3.6 线性填补缺失值对比

本文在缺失值填补上采用了样条插值和线性填补多种填补的方法。对所有的降水数据和部分地下水数据，如 Depth\_to\_Groundwater\_LT2, CoS, SAL, DIEC 进行线性填补缺失值，如上图 3.6。

线性填补使用线性回归的方式进行填补，用已知的数据拟合线性回归模型，并使用该模型来预测缺失值，数学公式为(27)。

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (27)$$

其中， $\beta_0$ 和 $\beta_1$ 是模型的截距和斜率， $\varepsilon$ 为误差项。之后通过已知变量对缺失值进行拟合。

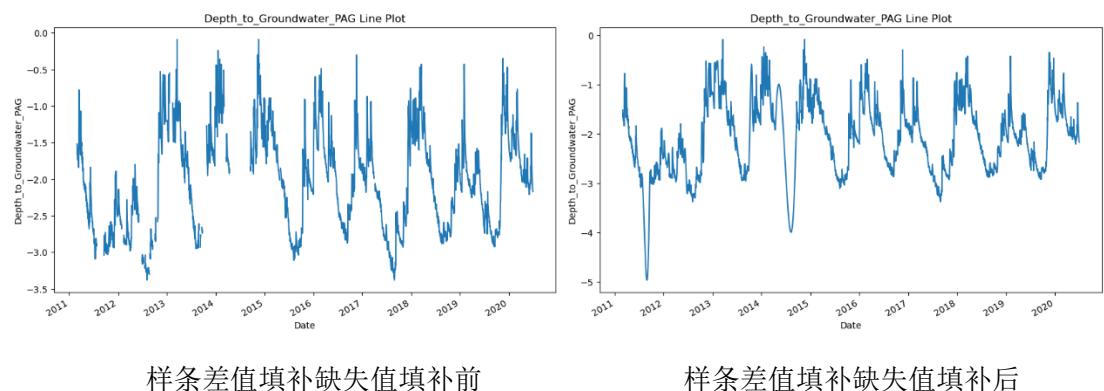


图 3.7 样条插值填补缺失值对比

其中对 PAG 列进行了样条插值的填补，如图 3.7，样条插值是一种常用的数据插值方法，它通过在数据点之间拟合出多项式片段（通常是低阶多项式），

---

从而在整个数据范围内进行平滑的插值。这些多项式片段称为样条函数，它们在每个数据点处具有连续性和光滑性。样条插值通过这些片段，尽可能准确地逼近原始数据，同时保持光滑性和连续性。

样条插值的常见形式是三次样条插值，其中每个片段都是三次多项式。

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (28)$$

其中  $a_i, b_i, c_i, d_i$  为待定系数，需要通过插值条件求解。

为保证样条函数的光滑性和连续性需要满足：在每个小段，样条函数必须连续，样条函数的一阶导数必须连续，样条函数的二阶导数必须连续，如 (29–32)，以及边界条件 (33)，自然边界条件是为了确保样条插值函数在端点处是平滑的，避免出现异常的振荡行为。

$$s_i'(x_{i+1}) = y_i \quad (29)$$

$$s_i'(x_{i+1}) = y_{i+1} \quad (30)$$

$$s_i''(x_{i+1}) = s_{i+1}'(x_{i+1}) \quad (31)$$

$$s_i''(x_{i+1}) = s_{i+1}''(x_{i+1}) \quad (32)$$

$$s''(x_0) = s''(x_n) = 0 \quad (33)$$

### 3.2.3 平滑异常值

由于观测数据中可能存在噪声和观测异常值，影响模型精度，所以需要对数据进行异常值检测并进行平滑处理。由于本文数据的特征列较多，所以使用移动平均的方法可以保持较高的计算效率，同时可以灵活的设置其窗口，适应不同的特征列。

本文中使用移动平均的方法对一系列连续数据点平均值进行计算，用于平滑数据并去除噪声。其中  $m$  为移动平均窗口的大小，由于数据的单位是每天一条，所以本文中设置窗口大小为 7。

$$MA_i = \frac{1}{m} \sum_{j=i-m+1}^i x_j \quad (34)$$

计算残差以及标准差

残差表示原始数据与移动平均之间的差异

$$residual_i = x_i - MA_i \quad (35)$$

标准差是衡量数据的离散程度的指标，表示数据点与其均值之间的平均偏差

$$\sigma_i = \frac{1}{m} \sum_{j=i-m+1}^i \sqrt{(x_j - MA_i)^2} \quad (36)$$

### 3.2.4 异常值检测

使用一个阈值判断残差是否大于标准差的倍数，若超过阈值则标记为异常值，并用平均值进行替换。在本文中使用了三倍标准差作为阈值。

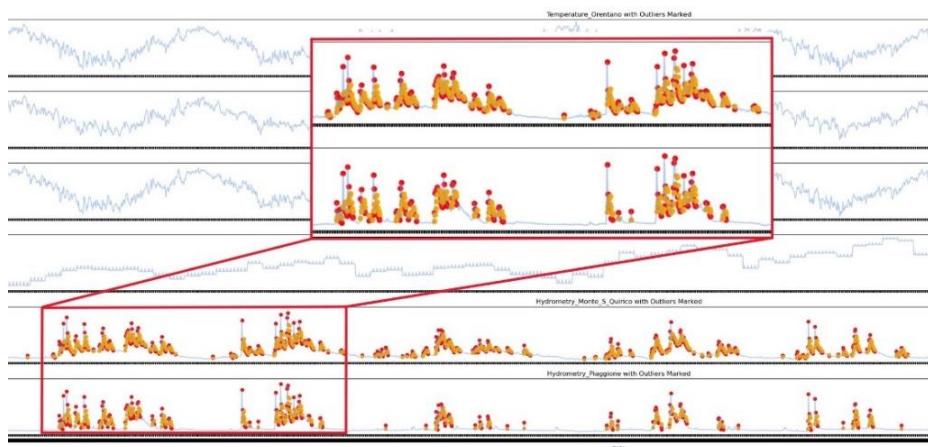


图 3.8 异常数据

部分列中存在较多异常值，图 3.8 中红色标记为异常值，黄色标记为修正后值，均全部进行替换。

### 3.2.5 划分训练集测试集

经过上述全部数据处理步骤后，数据时间范围为 2011-03-01 至 2020-06-30，为了保证完整的检验模型的性能将 2019-06-30 至 2020-06-30 整年的数据作为测试集，其余的作为训练集进行训练模型。

### 3.2.6 数据标准化

分别对训练集和测试集两个数据集，每个特征变量独立进行缩放，统一不同的量纲和单位，使得每个特征都落在指定的范围内，从而消除不同特征之间的单位差异，使得特征权重更加均衡，使得数据更加统一和可比较。

本文使用 Min-Max 标准化 (37)，对每列特征列进行处理。

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (37)$$

### 3.3 影响地下水位的特征变量选择

#### 3.3.1 Granger 因果检验

对每个因变量逐个进行 Granger 因果检验，以确定它们对各个特征变量对其的影响。对每个特征变量进行 F 检验，原假设  $H_0$  为  $x$  不是  $y$  的原因，当 p 值小于 0.05 时，拒绝原假设，则  $x$  是  $y$  的原因。

表 3.1 对于北部地下水位的格兰杰因果检验 P 值

列名	SAL P 值	PAG P 值	CoS P 值	DIEC P 值
<b>Rainfall_Gallicano</b>	0.00000	0.00000	0.00000	0.00000
Pontetutto	0.00000	0.00000	0.00000	0.00000
Monte_Serra	0.00000	0.00000	0.00000	0.00000
Orentano	0.00000	0.00000	0.00000	0.00000
Borgo_a_Mozzano	0.00000	0.00000	0.00000	0.00000
Piaggione	0.00000	0.00000	0.00000	0.00000
Calavorno	0.00000	0.00000	0.00000	0.00000
Croce_Arcana	0.00000	0.017	0.001	0.00000
Tereglio_Coreglia	0.00000	0.00000	0.00000	0.00000
Fabbriche_di_Vallico	0.00000	0.00000	0.00000	0.00000
<b>Temperature_Orentano</b>	0.041	0.00000	0.00000	0.00000
Monte_Serra	0.00000	0.00000	0.00000	0.00000
Lucca_Orto_Botanico	0.012	0.00000	0.00000	0.00000
<b>Volume_POL</b>	0.024	0.00000	0.00000	0.016
CC1	<b>0.763</b>	0.006	0.00000	<b>0.486</b>
CC2	<b>0.997</b>	0.005	0.00000	<b>0.917</b>
<b>Hydrometry_Monte_S_Quirico</b>	0.00000	0.012	0.00000	0.00000
Piaggione	0.00000	0.001	0.00000	0.00000

如表 3.1 可知，对于北部的 DIEC 和 SAL 各目标列，其中特征值

Volume\_CC1 和 Volume\_CC2 的 p 值均大于 0.05，并且超出 0.05 过多，显然应当接受原假设，可得 Volume\_CC1 和 Volume\_CC2 对目标列不存在因果关系。

由于目标列 LT2 为南部子系统的承压地下水，所以对于 LT2 应当将其余的四个北部子系统的地下水作为 LT2 的特征列划为考虑范围之内。

表 3.2 对于南部地下水位的格兰杰因果检验 P 值

列名	LT2 P 值
<b>Rainfall_Gallicano</b>	0.00000
Pontetetto	0.00000
Monte_Serra	0.00000
Orentano	0.00000
Borgo_a_Mozzano	0.00000
Piaggione	0.00000
Calavorno	0.00000
Croce_Arcana	0.00000
Tereglio_Coreglia	0.00000
Fabbriche_di_Vallico	0.00000
<b>Temperature_Orentano</b>	0.00000
Monte_Serra	0.00000
Lucca_Orto_Botanico	0.00000
<b>Volume_POL</b>	0.001
CC1	<b>0.436</b>
CC2	<b>0.149</b>
<b>Hydrometry_Monte_S_Quirico</b>	0.00000
Piaggione	0.00000
<b>Depth_to_Groundwater_PAG</b>	0.00000
CoS	0.00000
DIEC	0.00000
SAL	0.00000

如表 3.2 可知，Volume\_CC1 和 Volume\_CC2 的 p 值明显大于 0.05 且明显

大于其他特征变量，接受原假设，可得 Volume\_CC1 和 Volume\_CC2 对南部的 LT2 不存在因果关系。

### 3.3.2 斯皮尔曼相关性检验

由于数据中存在非线性关系，Granger 因果检验和斯皮尔曼相关性检验同时可以检验线性和非线性关系，将 Grange 因果关系检验和斯皮尔曼相关性检验的结果进行汇总分析，将分析出的无用的为噪声的特征列进行剔除，减少其影响，增加模型准确度和模型预测效率和性能。

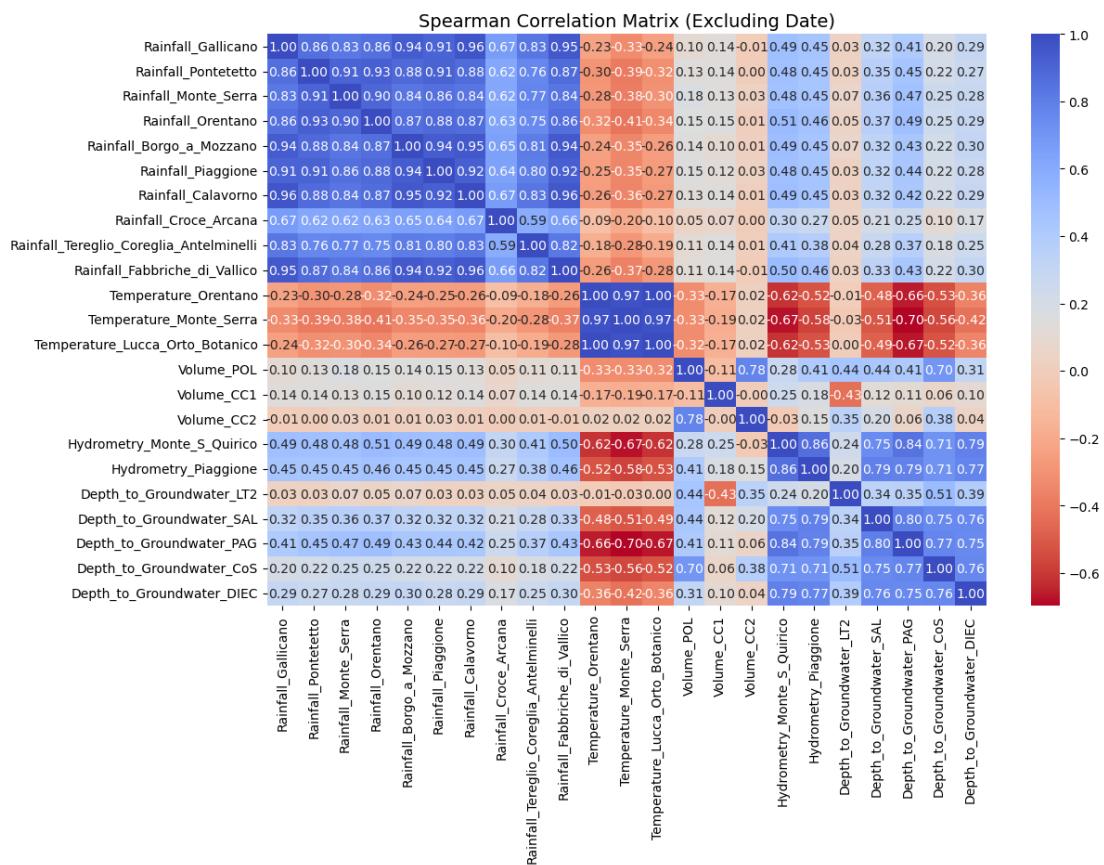


图 3.9 斯皮尔曼相关系数热力图

表 3.3 对于 Volume\_CC1 和 CC2 的相关性系数

目标列	Volume_CC1	Volume_CC2
Depth_to_Groundwater_SAL	0.12	0.20
Depth_to_Groundwater_PAG	—	—
Depth_to_Groundwater_CoS	—	—
Depth_to_Groundwater_DIEC	0.10	0.04
Depth_to_Groundwater_LT2	-0.43	0.35

根据上表 3.3 可知，PAG 和 CoS 两个目标列不存在不相关的特征列，而 SAL 和 DIEC 两列均存在 Volume\_CC1 和 Volume\_CC2 两个不相关的列，所以接下来需要对移除特征列的模型准确度加以判断，权衡模型复杂度和预测性能之间的关系，选择最优的特征列选择方案。

综上所述，在构建两个模型时，需要将 Volume\_CC1 和 Volume\_CC2 两列无关变量进行排除在外，从而减小其对预测结果和模型拟合的影响。

### 3.3.3 更正后的特征列

经过 Grange 因果检验和斯皮尔曼相关性检验，最终移除了 Volume\_CC1，Volume\_CC2 两列。

表 3.4 更新后的数据内容

原列名称	列名称	数据含义及单位
Date	日期	
Rainfall*10	降雨*10	该地下水流域内 10 各测量点每日各个降水测量数据 单位：毫米 mm
Temperature*4	温度*4	流域内气温监测数据点每日数据 单位：摄氏度 °C
Volume*3	水体体积*3	流域内水体会互相相互补给，所以对测量其他水体体积十分重要 (移除 Volume_CC1, Volume_CC2) 单位：升 L
Hydrometry*2	水文监测*2	监测水体表面的蒸发和植被蒸腾的水量。数据通常通过蒸发皿、蒸发计或其他设备来获取，通过水位传感器等设备进行监测。

## 四 基于 LSTM 模型预测的地下水位预测

### 4.1 构建基础的 LSTM 模型

#### 4.1.1 基础 LSTM 的构建

由于气候和地下水位数据为时间序列数据，LSTM 模型适用于时序数据建模和预测，如自然语言处理、时间序列预测等任务。LSTM 的门控机制，它能够更好地捕捉序列中的长期依赖关系，从而在处理长序列数据时表现更好。而地下水预测则是需要通过过去数据对未来数据进行预测，所以本文采用 LSTM 模型进行预测。

通过遗忘门来决定保留多少过去的信息，即 $f_t$ ，通过输出门来控制当前信息进入网络的数量，即 $C_t$ 。最后，使用输出门输出的信息用于预测，并与全连接层连接以生成最终的输出 $h_t$ 。在这个过程中， $x_{t-1}, x_{t-2}$ 分别表示 t 天前一天和前两天的数据。全连接层有四个节点，每个神经元对应一个目标列，全连接层接收前面几层的数据并将其转换为与目标相匹配的形式，在训练集中模型动态调整输出层的权重，以最小化预测值与真实目标之间的损失，从而更准确地学习目标之间的关系。

$$f_t = \sigma(W_f \cdot [h_{t-1}, (x_{t-1}, x_{t-2})] + b_f) \quad (38)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, (x_{t-1}, x_{t-2})] + b_i) \quad (39)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, (x_{t-1}, x_{t-2})] + b_c) \quad (40)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (41)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, (x_{t-1}, x_{t-2})] + b_o) \quad (42)$$

$$h_t = o_t \times \tanh(C_t) \quad (43)$$

在构建 LSTM 模型后，通过麻雀搜索算法（SSA）对 LSTM 的部分超参数进行搜索，其中包括 LSTM 层的神经元个数，L1 正则化，L2 正则化，Dropout 层丢弃的概率，存在四个搜索对象，所以搜索空间的维度  $d = 4$ ， $n$  为解空间内可能存在解的数量。

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} \quad (44)$$

通过多次迭代，麻雀搜索算法的筛选，根据评价指标 R 方选择其中的最大值，寻找到最优的参数值组合。

#### 4.1.2 LSTM 神经元数量的选择

受设备的限制，较大的迭代次数会造成运行速度过慢，同时运存过高，所以需要对模型训练的迭代次数进行缩小，以进行麻雀搜索算法，减少程序的运行时间，

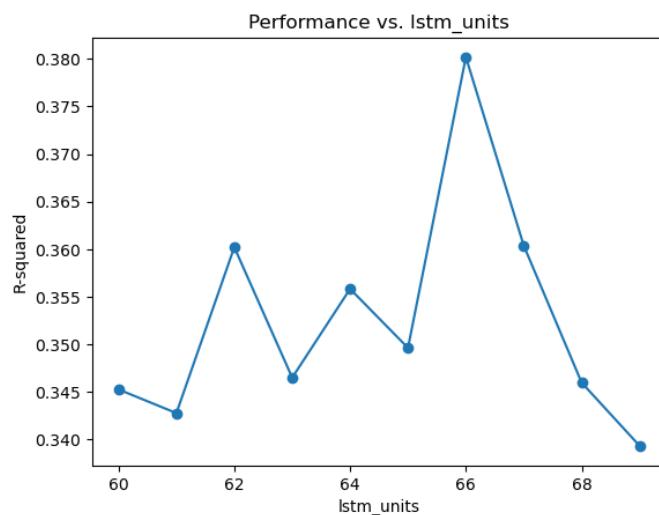


图 4.1 不同 LSTM 神经元数量对应 R 方值

经过几次初步的测试，决定将 LSTM 神经元的搜索节点个数范围设置为 60–70，经麻雀搜索算法进行搜索发现，如图 4.1，在增加到 66 个节点之后准确率下降较快，模型可能存在过拟合，并且相比 62 和 64 模型的 R 方值有明显的提高，但是模型复杂度并没有明显的上涨，所以综合来看，模型的节点数应当选择 66 更为合适。

#### 4.1.3 L1, L2 正则化值的选择

深度学习模型很容易过拟合 L1, L2 可以在模型训练过程中对模型参数进行惩罚，使得模型倾向于学习简单的模式，从而减少模型的复杂度，以防止过拟合，提高模型的泛化能力。在本文的搜索算法之中，将 L1 和 L2 的搜索范围设置为 0–0.03 之间，在不同情况下模型的拟合结果如下图 4.2 所示。

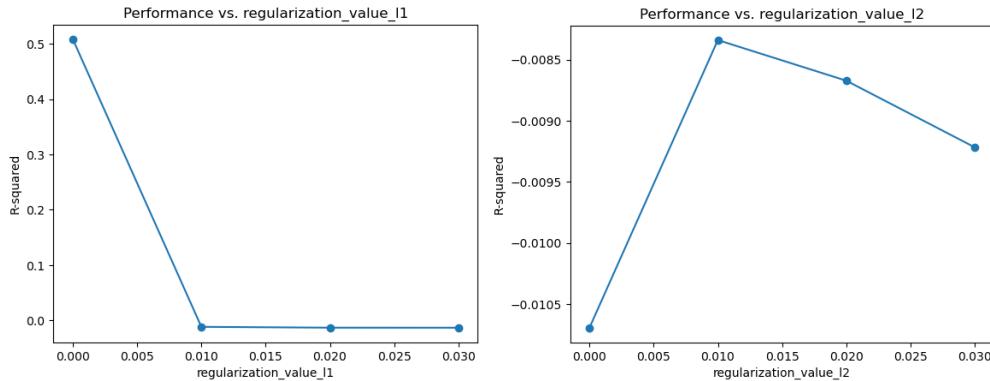


图 4.2 不同 L1, L2 正则化值对应 R 方值

L1 正则化：

$$L'_1 = L_{data} + \lambda_1 \sum_{i=1}^n |\omega_i| \quad (45)$$

L2 正则化：

$$L'_2 = L_{data} + \lambda_2 \sum_{i=1}^n \omega_i^2 \quad (46)$$

其中  $\lambda_1$  和  $\lambda_2$  分别是 L1 和 L2 正则化的参数， $L_{data}$  为数据损失， $\omega_i$  为模型的权重，n 为模型的参数数量。

针对不同的 L1 和 L2 正则化值，如图 4.2，当  $L1 = 0$  时， $L2 = 0.01$  时模型效果最好，L2 越大之后模型拟合越差，可能是过大的 L2 值导致模型复杂度降低，无法充分拟合训练数据中的复杂模式导致效果变差。

#### 4.1.4 Dropout 率的选择

Dropout 层同样是一种防止过拟合的方法，当 Dropout 设置过大时会导致网络中大量的神经元被随机丢弃，使得网络学习过于不稳定，影响模型的收敛速度，同时也可能会导致模型欠拟合，无法很好地捕捉训练数据中的特征和模式，影响模型的性能。过小可能会导致模型过度依赖某些神经元，忽略了其他神经

元的作用，导致模型过拟合，泛化能力较差。所以也需要通过麻雀搜索算法对寻到找其合适的参数值。

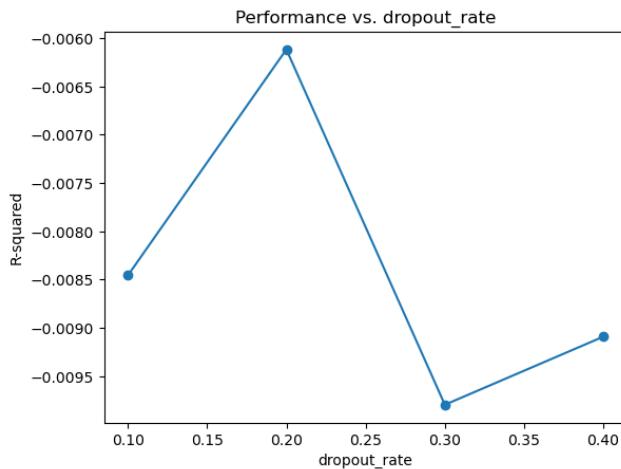


图 4.3 不同 Dropout 值对应 R 方值

Dropout 率越大，随机丢弃的神经元数量就越多，如图 4.3，当概率为 0.2 时，模型效果最好。

综上，在对几个超参数进行搜索后，超参数的组合选择为

表 4.1 LSTM 超参数选择

参数名称	参数值
LSTM 数量	66
L1 正则化	0.00
L2 正则化	0.02
Dropout 率	0.2
学习率	0.0001
Epoch	200
Batch size	2

## 4.2 基于融合注意力机制的 LSTM 神经网络预测地下水位

引入注意力机制可以显著提高模型的性能，特别是在处理长序列数据时。例如，在预测地下水位 T+1 天的数据时，注意力机制可以使模型更加专注于 T、T-1、T-2、T-365 等时间点的数据。通过动态地调整不同时间步的权重，模型能够更有效地捕捉与当前任务相关的信息，从而提高预测精度。注意力机制不仅

降低了模型的复杂度，还减少了被其他相关性较低数据干扰的可能性。

本文首先将原始数据输入 LSTM 层，再将输出的数据通过注意力机制进行加权计算，通过对 TimeAttentionLayer 类的调用和改写，首先重写了父类 tf.keras.layers.Layer 中的 build 方法，用于构建层。在这里，定义了层的权重，包括  $W_{feature}$ ,  $W_{time}$  两个权重矩阵，分别用于特征注意力权重和时间注意力权重。其次，重写了父类 tf.keras.layers.Layer 中的 call 方法，用于执行前向传播操作。在这个方法中，计算了特征注意力分数和时间注意力分数，并经过激活函数进行非线性变换。随后，通过 softmax 函数将注意力分数转换为注意力权重，并将其应用于输入以获得加权后的输出。计算过程如下：

$$et_{feature} = \text{ReLU}(x \cdot W_{feature}) \quad (47)$$

$$et_{time} = \text{ReLU}(t \cdot W_{time}) \quad (48)$$

$$et = et_{feature} + et_{time} + b \quad (49)$$

$$at = \text{softmax}(et) \quad (50)$$

$$output = x \odot at \quad (51)$$

改进后的 LSTM 模型结构示意图如下图 4.4 所示

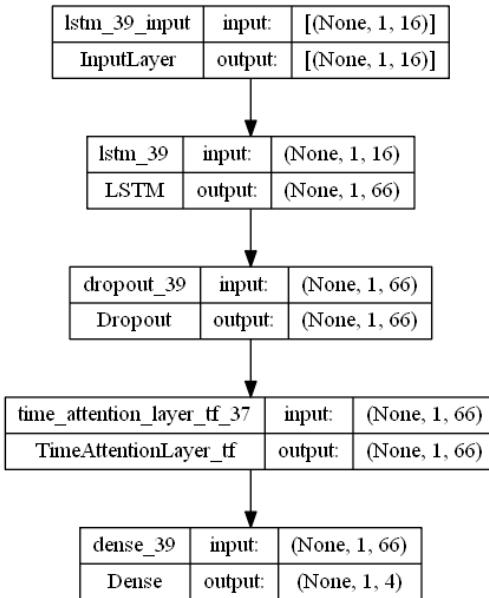


图 4.4 改进后的 LSTM 模型结构示意图

对模型经过 200 次迭代后，损失函数（loss）和验证损失（val\_loss）均逐渐趋近于 0 和 0.05，并且验证损失（val\_loss）在损失函数（loss）之上，上下

波动，说明验证集的精确度略有低于训练集，证明模型的训练效果良好，如图 4.5。

模型在预测北部 4 列（SAL, PAG, CoS, DIEC）结果的测试集上

MSE 值为 0.0001，MAE 值为 0.016，R 方达到 0.958。

预测南部 1 列（LT2）结果的测试集上

MSE 值为 0.0025，MAE 值为 0.0383，R 方达到 0.872。

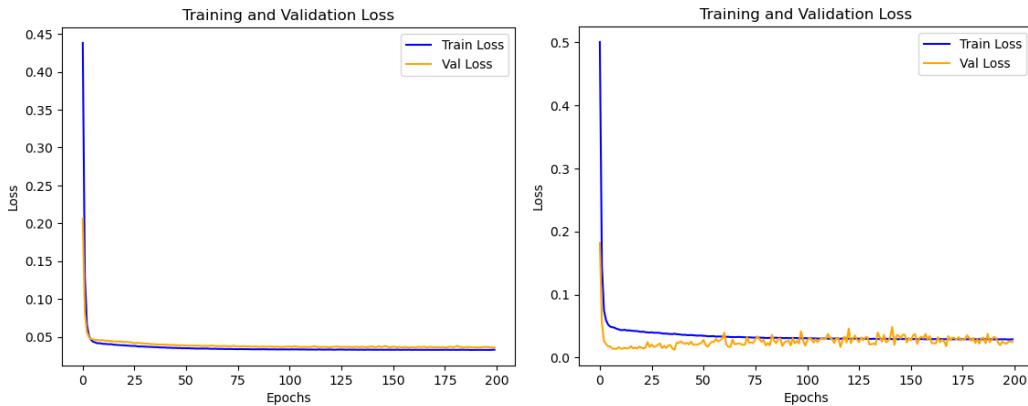


图 4.5 改进后的 LSTM 训练曲线，左侧为预测 4 列，右侧为 1 列

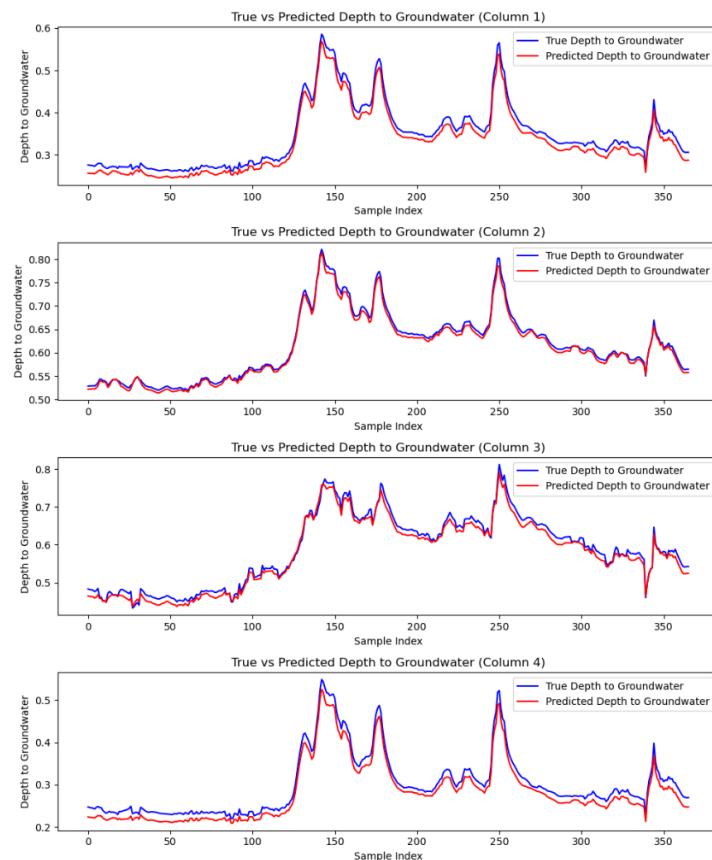


图 4.6 改进后的 LSTM 预测值与真实值对比（北部）

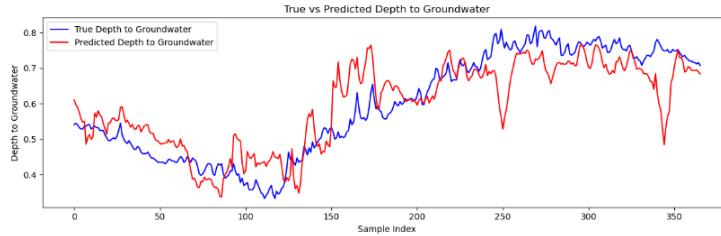


图 4.7 改进后的 LSTM 预测值与真实值对比（南部）

对于预测结果的可视化图 4.6, 4.7 中，可以明显地发现，在北部非承压层的四列预测中，改进后的 LSTM 预测值与真实值对比拟合的十分出色，无论是极值或是其他值预测结果都比较接近真实值，而对于南部承压层的预测基本拟合，在评价指标方面有所下降，在部分值的预测中有较大的波动，尤其是在极值方面，但总体的趋势大致相同，与非承压层相比表现略有不足。

造成这个的原因是因为，由于南部是位于承压层，特征列相比北部的特征列要更多，并且所受的影响因素相比北部更多更复杂，但是数据集只提供了基础的气象和水文特征，可能存在其他的影响因素无法在模型中进行拟合，导致预测结果相较于北部 4 列不佳，后续可以通过增加特征列进行改善。

## 五 改进后模型与其他模型的模型评价

### 5.1 对比不同特征选择对改进型 LSTM 模型的影响

R 方是一种衡量统计模型拟合程度的常用指标之一。它表示因变量方差中被自变量解释的比例。相比 MSE 和 MAE 来说，R 方可以更好地表示在不同情况下改进型 LSTM 模型对数据的拟合情况。

所以本文通过比较 R 方来对比在不同特征选择下，改进型 LSTM 模型在预测集上的表现情况，结果如下图 5.1 所示。

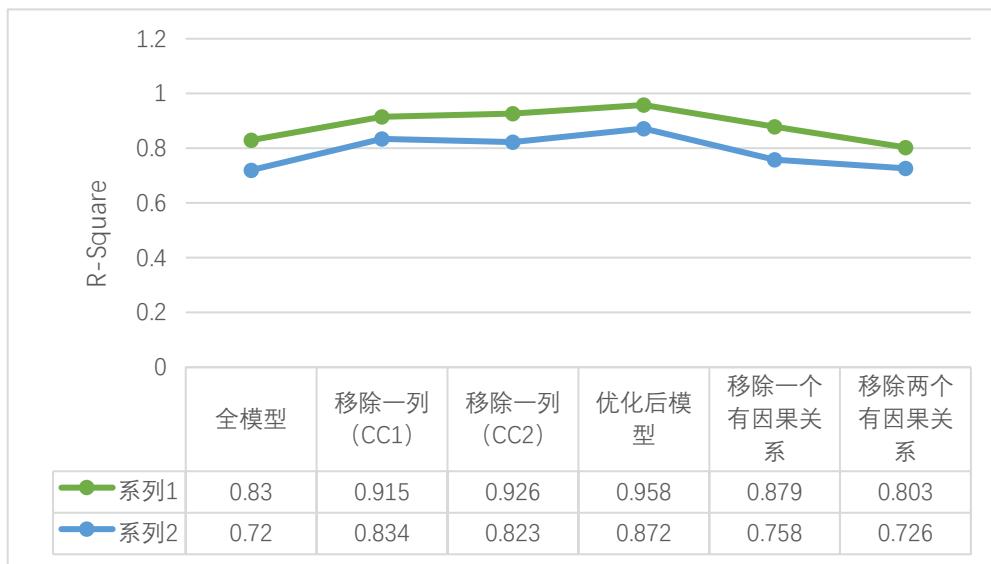


图 5.1 不同特征选择对应不同 R 方值

显然，在预测两个子系统地下水位时，可以观察到，每移除一个经过 Grange 因果关系检验中无因果关系的特征列后，模型结果的 R 方值都会逐渐增加，从 0.78 左右上升至 0.90 左右，上升了约 15%。进一步移除一个无关特征列后，模型的拟合程度继续增加，上升了较全模型约 21%。这一现象充分说明了无关特征列对模型性能的显著影响。这些无关列可能包含与南部地下水位无关的信息和噪声，严重干扰了模型的学习过程。此外，它们可能与其他关键特征列存在共线性问题，导致模型参数估计的不稳定性，进而影响模型的泛化能力，最终使模型性能大幅下降。

为了排除是由于减少因变量，使得数据复杂度降低所导致的拟合程度增加，所以随机删除了两列在检验中具有因果关系的两列，发现预测的结果无论是从

精确度或是拟合程度上都有大幅的下降，

因此，在构建和优化模型时，应谨慎选择特征列，在前期使用 Grange 因果检验是十分有效的，可以避免引入无关和共线的特征，以确保模型的有效性和准确性。

可见 Grange 因果关系检验和斯皮尔曼相关性系数对地下水位目标列的特征列的筛选是准确的，在删除不相关列后，无论是在北部四列和南部一列的预测模型中效果又有所提高。

## 5.2 改进型 LSTM 模型与其他模型对比

在国内外研究现状中，有其他不少文章采用 XGBoost 梯度提升树或是随机森林回归等方式对地下水位等数据进行预测，为了全面评估融合注意力机制的 LSTM 模型的性能，本文采用 MSE, MAE, R 方作为评价标准，旨在从模型拟合程度和预测误差大小两个维度进行综合考量。同时为了展示模型的优越性，我们同时将注意力机制 LSTM 模型分别与传统 LSTM 模型、XGBoost 和随机森林回归进行比较，评价模型的预测能力。

### 5.2.1 改进型 LSTM 与传统 LSTM 模型对比

传统模型仅包含传统的 LSTM 层，dropout 层，全连接层，与改进型相比，如图 5.2，剔除了对因变量因果关系较小的两列 (Volume\_CC1, Volume\_CC2)，缺少了对时间序列和特征列的注意力机制，对部分重要的时间序列数据和特征列不敏感，并且容易受不相关因素的影响，影响模型运行速度和模型的预测准确率。

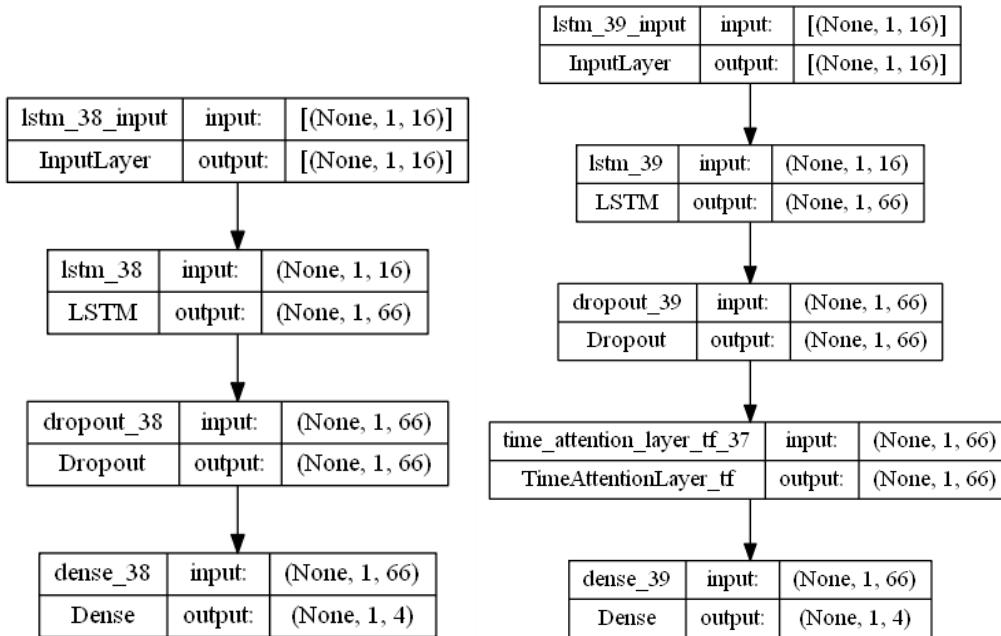


图 5.2 传统 LSTM 和改进型 LSTM 模型结构对比

经过麻雀搜索算法对最佳超参数进行搜索，进行 5 次迭代，其中最佳超参数设置如表 5.1 所示

表 5.1 传统 LSTM 模型参数设置

参数名称	参数值
LSTM 数量	66
L2 正则化	0.02
Dropout 率	0.2
学习率	0.0001
Epoch	200
Batch size	2

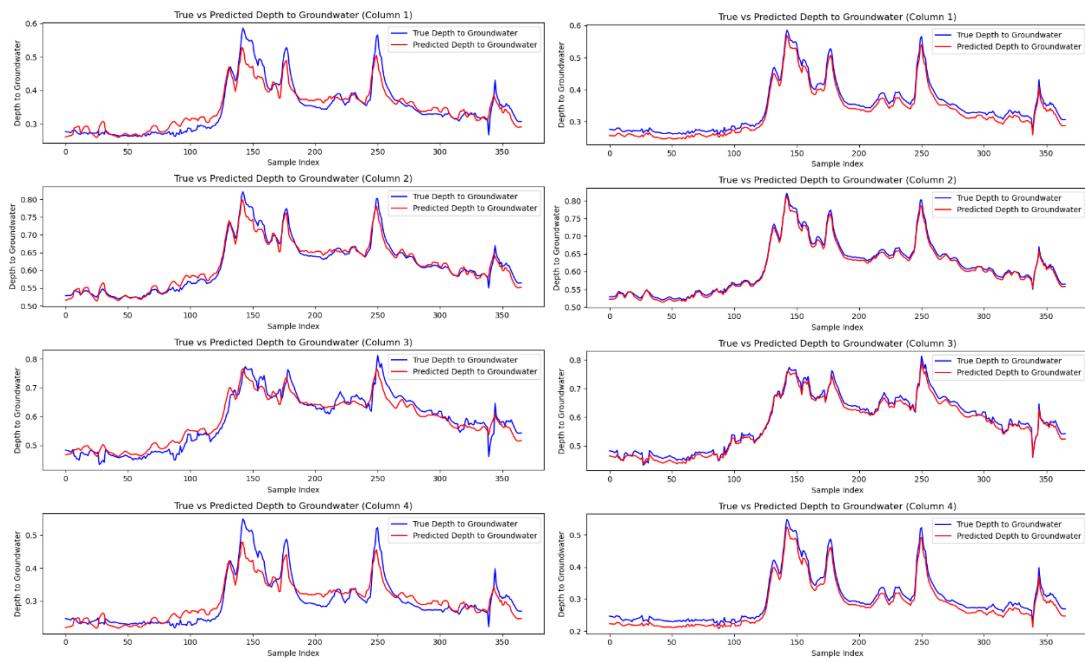


图 5.3 传统 LSTM 模型（左侧）预测值和真实值与优化后 LSTM（右侧）对比（北部四列）

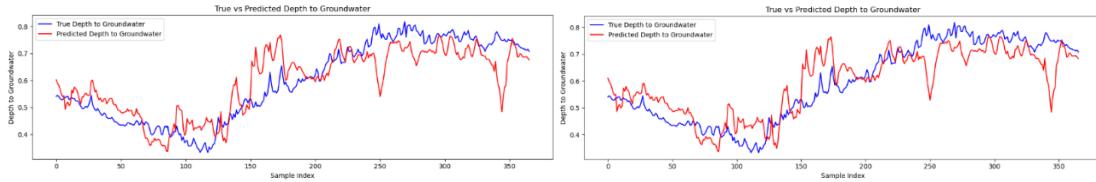


图 5.4 传统 LSTM 模型（左侧）预测值和真实值与优化后 LSTM（右侧）对比（南部一列）

在预测北部 4 列 (SAL, PAG, CoS, DIEC) 模型上，结果如图 5.3 所示，数据方面如图 5.5 所示，

MSE 0.0005 与改进 LSTM 相比增大了 0.0004，

MAE 0.018 与改进 LSTM 相比增大了 0.002，

R 方 0.907 与改进 LSTM 相比降低了 0.051。

在南部 1 列 (LT2) 模型上，结果如图 5.4 所示，数据方面如图 5.5 所示，

MSE 0.0028 与改进 LSTM 相比增大了 0.0003，

MAE 0.0400 与改进 LSTM 相比增大了 0.0017，

R 方 0.870 与改进 LSTM 相比降低了 0.002。

从图中可以明显观察到，在部分时间段内，改进型的 LSTM 模型相较于传统 LSTM 模型展现出了更为贴切的预测性能，尤其是在一些极值的预测结果上。

这主要来自于注意力机制在模型中的有效运用，使得模型能够更好地聚焦于关键信息，从而提高预测的准确性和泛化能力。

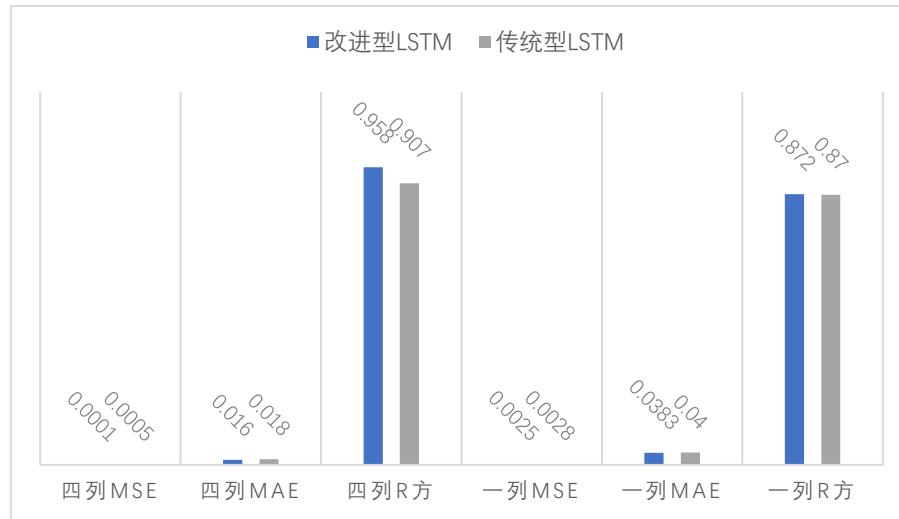


图 5.5 改进型 LSTM 和传统型 LSTM 评价指标对比

### 5.2.2 改进型 LSTM 与 XGBoost 对比

XGBoost 是一种基于梯度提升树的机器学习算法，它在大规模、稀疏数据集上的性能表现优秀。XGBoost 是一种集成学习算法，通过集成多个弱分类器，从而构建一个强分类器。它广泛使用在解决结构化数据问题方面，也被广泛用于数据挖掘的使用中。

经过麻雀搜索算法对最佳超参数进行搜索，进行 5 次迭代，其中最佳超参数设置如表 5.2 所示

表 5.2 XGBoost 参数

参数名称	参数值
基础学习器	dart
L1 正则化	0.5
L2 正则化	0.5
Num estimators	400

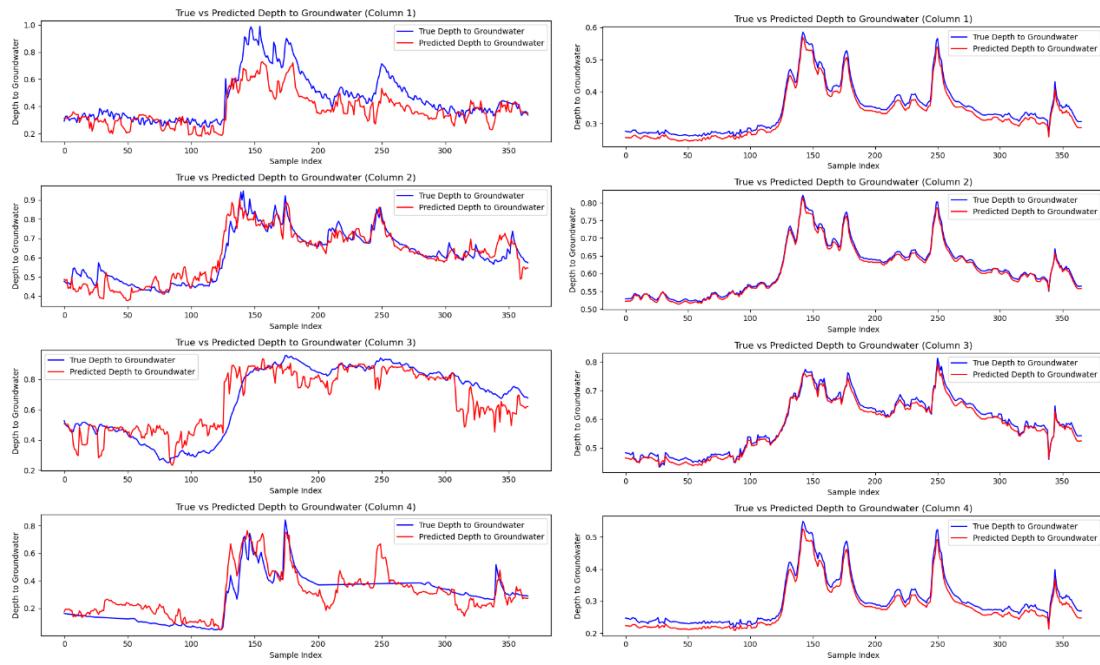


图 5.6 XBGooest 模型（左侧）预测值和真实值与优化后 LSTM（右侧）对比（北部四列）

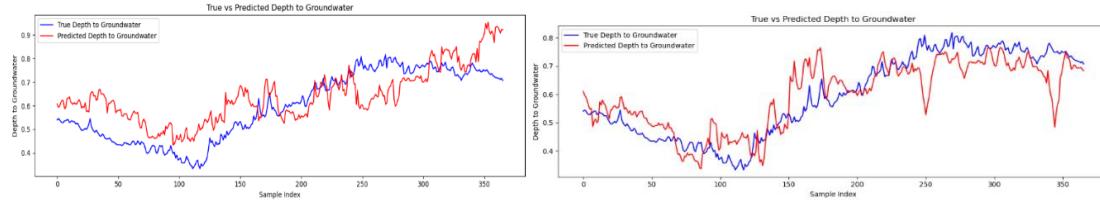


图 5.7 XBGooest 模型（左侧）预测值和真实值与优化后 LSTM（右侧）对比（南部一列）

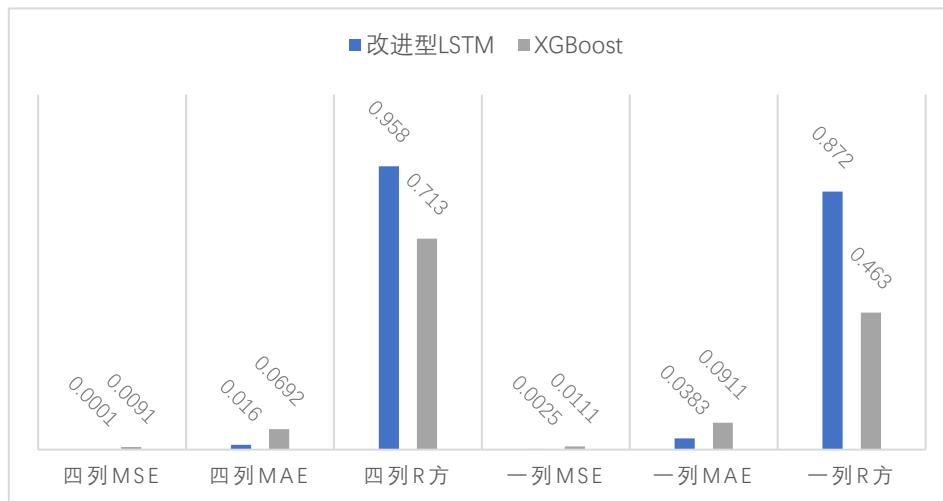


图 5.8 改进型 LSTM 和 XGBoost 各个评价指标对比

在预测北部 4 列 (SAL, PAG, CoS, DIEC) 模型上, 结果如图 5.6 所示, 数据方面如图 5.8 所示,

MSE 0.0091 相比改进型 LSTM 增大了 0.0090,

MAE 0.0692 相比改进型 LSTM 增大了 0.0532,

R 方 0.713 相比改进型 LSTM 减少了 0.245。

在预测南部 1 列 (LT2) 模型上, 结果如图 5.7 所示, 数据方面如图 5.8 所示,

MSE 0.0111 相比改进型 LSTM 增大了 0.0086,

MAE 0.0911 相比改进型 LSTM 增大了 0.0528,

R 方 0.463 相比改进型 LSTM 减少了 0.409。

在结果图像上可以很明显的看到, XGBoost 与 LSTM 模型相比, 缺少了在细节方面的精确度, 相比 LSTM 来说, XGBoost 主要侧重于预测整体趋势, 而在捕捉具体、高精度数据方面则显得力不从心。

在预测南部一列的情况时, 前半部分预测较为精准, 而后半部分趋势上则呈现了相反的的趋势, 可能是由于 XGBoost 模型可能难以适应这种变化, 从而导致预测偏差。所以相比 XGBoost, LSTM 更具有优势。

### 5.2.3 改进型 LSTM 与随机森林回归对比

随机森林回归是一种集成学习算法, 它基于决策树构建多个决策树模型, 并通过取平均值来提高预测性能。与传统的决策树算法相比, 随机森林具有更好的泛化能力和抗过拟合能力。

经过麻雀搜索算法对最佳超参数进行搜索, 进行 5 次迭代, 其中最佳超参数设置如表 5.3 所示

表 5.3 随机森林回归参数

参数名称	参数值
Max_depth	15
Num estimators	500

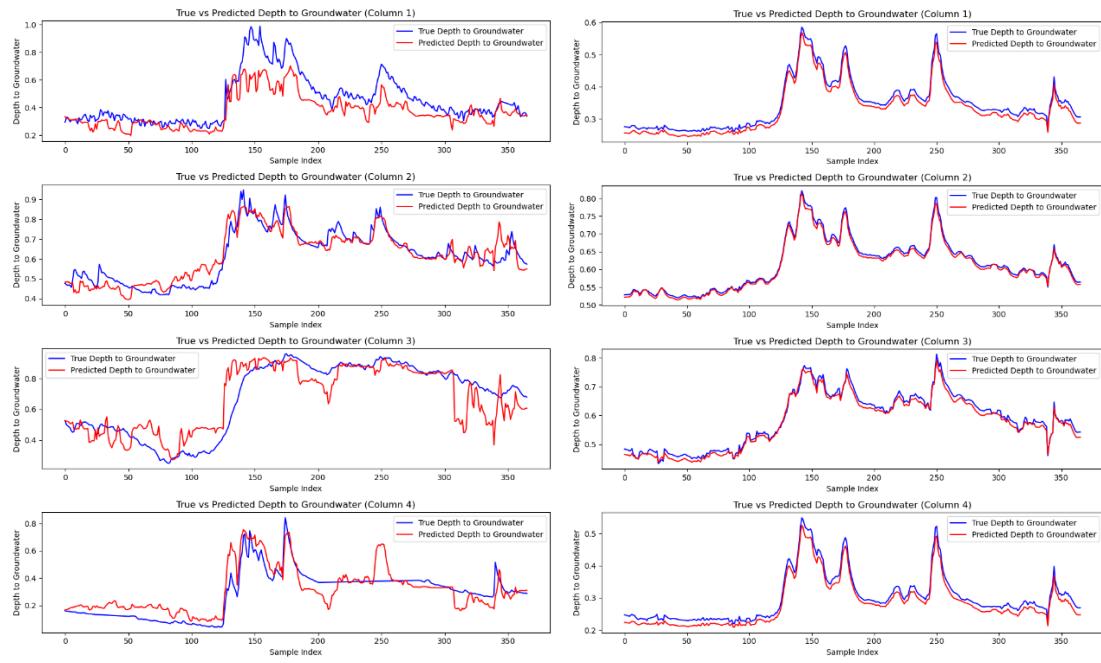


图 5.9 随机森林模型（左侧）预测值和真实值与优化后 LSTM（右侧）对比（北部四列）

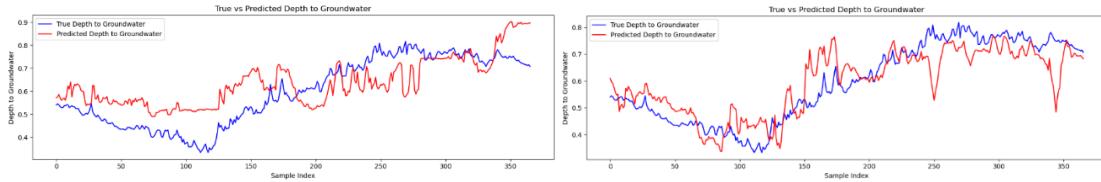


图 5.10 随机森林模型（左侧）预测值和真实值与优化后 LSTM（右侧）对比（南部一列）



图 5.11 改进型 LSTM 和随机森林各个评价指标对比

在预测北部 4 列 (SAL, PAG, CoS, DIEC) 模型中, 结果如图 5.9 所示, 数据方面如图 5.11 所示,

MSE 0.0090 相比改进型 LSTM 增大了 0.0089,

MAE 0.0690 相比改进型 LSTM 增大了 0.076,

R 方 0.717 相比改进型 LSTM 减小了 0.241。

在南部 1 列 (LT2) 模型中, 结果如图 5.10 所示, 数据方面如图 5.11 所示,

MSE 0.010 相比改进型 LSTM 增大了 0.0099,

MAE 0.0900 相比改进型 LSTM 增大了 0.293,

R 方 0.475 相比改进型 LSTM 降低了 0.397。

由图可知, 随机森林和 XGBoost 模型一样, 与改进型 LSTM 模型相比, 同样缺少了在细节方面的精确度, 主要侧重于预测整体趋势, 在预测极值方面也有不足, 在捕捉具体、高精度数据方面则显得力不从心。

#### 5.2.4 小结

表 5.4 北部南部不同模型各个评价指标汇总

		MSE	MAE	R 方
	<b>改进型 LSTM</b>	<b>0.0001</b>	<b>0.016</b>	<b>0.958</b>
预测北部 四列	传统 LSTM	0.0005	0.018	0.907
	XGBoost	0.0091	0.0692	0.713
	随机森林	0.0090	0.0690	0.717
	<b>改进型 LSTM</b>	<b>0.0025</b>	<b>0.0383</b>	<b>0.872</b>
预测南部 一列	传统 LSTM	0.0028	0.0400	0.870
	XGBoost	0.0111	0.0911	0.463
	随机森林	0.0108	0.0902	0.475

总体来说, 如表 5.4 所示, 在预测水体子系统中的承压地下水, 即南部 (LT2) 地下水水位的预测方面, LSTM 无论是改进型还是传统型, 相比 XGBoost 和随机森林回归都更加准确, 这表明 XGBoost 和随机森林回归模型在

---

学习地下水系统内部关系和处理复杂具体的精确的数据方面可能存在一定不足，导致其拟合程度相比 LSTM 都有显著下降。此外，在预测误差方面，这些模型也显示出了一定程度的增加，所以对于一个水系中多个子系统的地下水的水位的复杂问题中，传统机器学习模型显得较为简单，LSTM 神经网络相比传统机器学习的方法可以更好的学习并拟合模型，但是相比预测北部（SAL, CoS, DIEC, PAG）四列的结果来看，南部的改进后的模型均不如北部预测的效果，可能是由于影响因素的增加，但数据量的大小并未改变，数据量较少所导致模型的拟合不够充分。

在预测北部（SAL, CoS, DIEC, PAG）四列地下水水位的预测中，LSTM 网络与 XGBoost 和随机森林回归的差距相比预测南部一列上有所缩小，同时 R 方也有一定的提升，但是融合注意力机制的 LSTM 神经网络在这方面的预测更加出色。

这也印证了在一个水系中不需要考虑其他子系统为较为简单的任务时，传统机器学习的性能有所改善，与 LSTM 神经网络的性能差距有所缩小，但融合了注意力机制的 LSTM 依旧出色，相比其余三个模型在三个评价指标上均有较大提升，并且无论是在可视化图像上或是数据上均有所改善。

## 六 总结与展望

### 6.1 研究总结

水资源是人类生活和生态环境最重要的一部分，而地下水是水资源不可缺少的一部分，对于地下水位的预测是水资源调配和生态可持续不可缺少的一部分。

本文首先通过对地下水区域所在流域内的气象数据和水文数据进行处理，将处理过的时间序列数据，使用 Grange 因果检验和斯皮尔曼相关性系数对特征变量进行筛选，通过 F 检验，保留当前系统中对预测任务价值较大的列，对无关变量进行剔除，减少噪声和无关变量对任务的干扰，提高结果预测的准确程度，维护数据集的可用性。

之后通过麻雀搜索算法对模型进行优化，寻找属于模型的最佳超参数，算法模拟了麻雀的捕食行为，将搜索空间划分为多个区域，并通过随机选择和交换来探索不同区域，更简单的达到全局搜索的目的。之后引入了一定的随机性，避免搜索过程中，陷入局部最优解，并且当陷入局部最优解时会主动跳出。随机性的加入，使得算法有相对较短的时间内找到较优解，从而有较快的收敛速度，达到优化模型的效果。

构建模型，对南北两个子系统的地下水位进行拟合。首先将数据输入 LSTM 层，将 LSTM 层输出的数据经过注意力机制的筛选，使模型更加专注于与当前预测任务相关的时间序列数据，并且增加对当前预测任务影响较大的特征值的权重，提高模型的预测能力，最后将数据输入全连接层，输出成所需要的形状。

相比于传统的 LSTM 模型、XGBoost 和随机森林回归模型，无论是解决同一流域内地下水子系统之间的关系，或是气候和水文特征与地下水位之间的关系，融合了 LSTM 的改进型模型都有着显著的提升，尤其是针对流域内地下水子系统之间问题的解决是传统机器学习方法所不具备的。

---

## 6.2 展望

本次的研究还有不足之处，在未来的研究还可以从以下方面改进：

1. 在麻雀搜索算法中，受限制于性能的局限，迭代次数有所欠缺，所以所得到的可能并不是最具代表型的最优参数。
2. 由于部分列缺失数据过多，已经无法进行缺失值填补，所截取的可用数据数据量较少，可能导致信息损失，存在模型拟合不够充分的问题。
3. 以上的预测是基于已知的因变量来进行预测，未来可以考虑使用生成对抗网络，将历史数据输入训练模型，将文中的特征变量作为协变量，预测未来的地下水位，提供更大的作用。
4. 在对南部承压层预测的结果相比较差，可能是由于南部承压层相比北部所受的影响因素过多，但是数据集中给出特征列并不能充分的使模型进行拟合，未来可以增加特征列，以更好地预测承压层水位。

## 参考文献

- [1] 曹还君, 李长云. 基于SSA-LSTM模型的空气质量预测研究[J]. 现代信息科技, 2024, 8(04): 142-146+152. DOI:10.19850/j.cnki.2096-4706.2024.04.030.
- [2] 崔静芝. 水量均衡法在地下水水资源评估中的应用分析[J]. 地下水, 2022, 44(03): 51-52+91. DOI:10.19807/j.cnki.DXS.2022-03-015.
- [3] 蔡逸. 基于机器学习的季节性冻融过程中潜水水位动态模拟及预测研究[D]. 吉林大学, 2023. DOI:10.27162/d.cnki.gjlin.2023.001595.
- [4] 黄鹤, 李文龙, 杨澜等. 跳跃跟踪SSA交叉迭代AP聚类算法[J/OL]. 电子学报, 1-14[2024-03-26]. <http://kns.cnki.net/kcms/detail/11.2087.TN.20240312.1044.106.html>.
- [5] 刘瑞. 基于深度学习的河流水位预测研究[D]. 青海师范大学, 2023. DOI:10.27778/d.cnki.gqhzy.2023.000459.
- [6] 廖绍欢, 赵乃千, 詹旭. 基于BP神经网络的地下水位预测系统设计[J]. 地震科学进展, 2023, 53(04): 165-170. DOI:10.19987/j.dzksjz.2022-118.
- [7] 田园, 孙梦觉, 周植高等. 一种基于信息熵的LSTM时间序列数据预测模型[J]. 科技创新与应用, 2024, 14(07): 28-34. DOI:10.19981/j.CN23-1581/G3.2024.07.008.
- [8] 王利军. 基于神经网络算法分析和预测低渗透性含水层矿坑涌水对地下水位的影响[J]. 治淮, 2024, (01): 24-26.
- [9] 王永强, 张森, 谢帅等. 基于深度学习的三峡未来坝前最大最小水位预测研究[J/OL]. 长江科学院院报, 1-6[2024-03-26]. <http://kns.cnki.net/kcms/detail/42.1171.TV.20231218.1021.008.html>.
- [10] 魏乐, 李承霖, 房方等. 小样本下基于改进麻雀算法优化卷积神经网络的飞轮储能系统损耗[J/OL]. 电网技术, 1-9[2024-03-26]. <https://doi.org/10.13335/j.1000-3673.pst.2023.2001>.
- [11] 汪璐璐, 谢小伟, 幸茂仁, 等. 基于GRACE数据评估黄土高原煤炭开采对地下水储量的影响[J]. 江西科学, 2024, 42(02): 264-270. DOI:10.13990/j.issn1001-3679.2024.02.008.
- [12] 许博文, 毕敬, 苑海涛等. 基于季节性分解与长短期记忆网络的水质动态预警[J]. 智能科学与技术学报, 2021, 3(04): 456-465.
- [13] 薛子钊, 许琴, 马理辉等. 榆阳区风沙草滩区地下水动态特征与影响因素分析[J]. 地下水, 2024, 46(01): 71-73+96. DOI:10.19807/j.cnki.DXS.2024-01-022.
- [14] 张玉鹏, 娄云深. 基于时变传染网络和分位数Granger因果检验的系统性风险传染研究[J/OL]. 金融经济学研究:1-19[2024-04-29]. <http://kns.cnki.net/kcms/detail/44.1696.F.20240326.1724.016.html>.
- [15] Agarwal V, Akyilmaz O, Shum C K, et al. Machine learning based downscaling of GRACE-estimated groundwater in Central Valley, California[J]. Science of the Total Environment, 2023, 865: 161138.

- 
- [16] Coppola Jr E A, Rana A J, Poulton M M, et al. A neural network model for predicting aquifer water level elevations[J]. *Groundwater*, 2005, 43(2): 231–241.
  - [17] Franceschini F. The groundwater monitoring system in Tuscany according to the EU Directive 2000/60 and the groundwater As contamination in the Versilia coastal plain[J]. *Acque Sotterranee–Italian Journal of Groundwater*, 2020, 9(3).
  - [18] Hiemstra C, Jones J D. Testing for linear and nonlinear Granger causality in the stock price - volume relation[J]. *The Journal of Finance*, 1994, 49(5): 1639–1664.
  - [19] Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. *Systems science & control engineering*, 2020, 8(1): 22–34.

## 致谢

至此论文已经全部结束了，大学的4年时间也要结束了，在这四年中我学到了许多的专业知识，结识了很多朋友，找到了自己的兴趣所在，挑战了不少困难，提高了自我学习的能力，真真正正改变了我很多。

首先感谢我的指导老师，阮敬老师在我的初期选题方面，为我指明了方向，在论文写作方面对我起到了督促和指导的作用，同时在论文的内容上提供了宝贵的意见和建议，在此表达我深深的感谢。

同时也要感谢刘心洁同学，在论文写作期间，和大学4年期间对我的帮助，她的帮助带给了我灵感，同时也带给我了力量。

也要感谢我的家人和朋友们，在任何时候，无论是在学习上还是生活上都对我起到了很大的鼓励和支持。

最后向审阅我的论文和参加答辩的老师表示衷心的感谢，感谢您们抽出宝贵的时间审阅我的论文，感谢您们对论文的不足之处提出的宝贵意见和建议。