Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 10, No. 2, 2024

Precipitation Prediction in Suqian City, Jiangsu Province of China Using LSTM Based on Grey Wolf Optimizer Algorithm

Yifan Zhang [1], Ruiwei Zhang [2], Shihan Zhang [1], Zhifu Jia [1, *]

[1] School of Mathematics and Sciences, Suqian University, Suqian, Jiangsu, China
[2] School of Information Engineering, Suqian University, Suqian, Jiangsu, China

*** Corresponding author:** Zhifu Jia (Email: jzflzbx@163.com)

Yifan Zhang and Ruiwei Zhang are co-first authors

Abstract: To enhance urban planning and disaster management in the context of global climate change, we focused on precipitation forecasting in Suqian City, Jiangsu Province, and established a Long Short-Term Memory (LSTM) model optimized by the Grey Wolf Optimizer (GWO) algorithm. Firstly, we collected five years of precipitation data to ensure a solid foundation for model training. Secondly, we integrated the GWO algorithm to optimize the LSTM model parameters, thereby improving prediction accuracy. Through data preprocessing, model establishment, and validation, we successfully constructed a precipitation forecasting model tailored for both rainy and non-rainy seasons in Suqian. Lastly, utilizing this model, we predicted the precipitation volume for the upcoming 12 months in Suqian and proposed targeted countermeasures and suggestions based on the forecast results to guide urban planning and disaster management efforts, ensuring that Suqian can effectively address the challenges posed by future precipitation changes. Additionally, by combining actual conditions with predictive data, we refined and optimized Suqian's precipitation management strategies.

Keywords: LSTM; GWO; ADF Test; ARMIA; Gray Prediction.

1. Introduction

As global climate change intensifies, precipitation forecasting becomes critical. Precipitation, as the core of the water cycle, is relevant to agriculture, water resources, urban planning and infrastructure. As a climate transition zone, Suqian City is frequently affected by floods and droughts, and high-precision precipitation modelling is crucial for disaster mitigation and cost reduction. Predicting precipitation fluctuations helps flood control, drought prevention and ecological maintenance, and also provides data support and decision-making basis for coping with extreme weather. The instability of precipitation not only affects all aspects of social production and life, but its quantity and period greatly influence the yield and quality of agriculture, people's daily travelling, and the frequency of natural disasters such as floods, landslides and mudslides.

Accurate precipitation prediction is critical to combating climate change, safeguarding people's livelihoods and promoting sustainable economic development. Research on precipitation prediction technology is essential to address the challenges of climate warming and to ensure food security, resource protection and economic stability. Efficient prediction models can warn of precipitation anomalies, optimise water resource management and mitigate the negative impacts of extreme weather. In Suqian City, Jiangsu Province, located in the transition from the subtropical monsoon zone to the temperate monsoon zone along the Huaihe River, it is strongly influenced by the monsoon. Because of its flat terrain, its water volume changes periodically, and the rise in rainfall greatly affects the growth and production of crops. The excessive rainfall in turn leads to urban flooding, causing unpredictable damage to local development and construction. Grasping the regular changes of precipitation in the region can strengthen the ability of the relevant departments to prevent floods and floods, so that people's production and life can be guaranteed.

In this paper, we use the grey wolf optimisation algorithm and the long and short-term memory neural network method for the changes of precipitation in Suqian, Jiangsu Province. We take the past its precipitation in the four years of 2019-2022 as a training set, and the precipitation in the year 2023 as a validation set. The number of network layers, the number of neurons per layer, and the learning rate are rationally selected in the genetic method, and the parameters are automatically selected to achieve a good balance of search efficiency and solution quality. The aim is to seek to get the optimal solution.

2. Status of Research

2.1. GWO Optimisation Model Study

Grey Wolf Optimizer (GWO), a swarm intelligence optimisation algorithm proposed in 2014 by Griffith University, Australia, academics Mirjalili et al.[1] . Inspired by grey wolf group predatory behaviour, the GWO algorithm simulates the leadership hierarchy and hunting mechanism of grey wolves in nature. The classification of grey wolves into four types was used to simulate hierarchical strata. In addition, the three main phases of finding prey, surrounding prey and attacking prey are also modelled.

2.2. LSTM Modelling Study

The Long Short Term Memory Neural Network, or LSTM, is a special type of recurrent neural network designed to solve the problem that the activation functions sigmoid and tanh functions in medium and long sequences training with large or small inputs will result in a convergence to 0 or infinite values, which can lead to gradient vanishing or gradient explosion. Unlike standard recurrent neural networks (RNN), LSTM introduces memory cells to control the flow of information in the network. Unlike other sequences, LSTM introduces control units of gates, i.e., input gates, forget gates, and output gates constitute[2] that are capable of retaining and transmitting information from longer temporal sequences in the past. With forgetting gates, the model can selectively ignore irrelevant information, making it more effective in dealing with long-term time series data.

In the recent development of deep learning, LSTM, as part of the machine learning method, simulates the structure and function of the neural network of the human brain, with powerful computational and data analysis capabilities[3] .The LSTM model solves the problem of gradient explosion through a gate structure that controls the degree of information retention, thus enabling the processing of complex nonlinear time-series data. Such networks are able to efficiently process long sequential data and learn to remember long-term dependencies, which excels in tasks such as natural language processing and time series prediction[3] .

The emergence of LSTM has greatly boosted the application of deep learning in the field of sequence modelling and prediction. Compared with the traditional RNN, LSTM is able to better capture and utilise long-term dependencies in time series data, enabling the model to achieve remarkable results in dealing with language, audio, video and other facing natural languages. At the same time, it has shown good results in predicting weather and climate change, in order to meet the challenges posed by climate change.

2.3. Precipitation Forecasting Studies

Climate change in China is receiving increasing attention from experts and scholars in the context of global warming. China's climate system has continued to warm, and it has become a sensitive area and an area of significant impact of global climate change. Extreme heavy precipitation events have increased, extreme low-temperature events have decreased significantly, and extreme high-temperature events are frequent. In this context, precipitation prediction has become particularly important.

And statistical theory has developed rapidly since the 20th century, combined with advances in computer technology, to provide strong support and guarantee for precipitation prediction. However, it is often difficult to obtain high-precision prediction results only by relying on statistical methods, such as solving low-dimensional and linear problems only with basic theories. Therefore, many scholars have proposed various new prediction methods, such as fuzzy theory, grey prediction system, chaotic prediction system, machine learning and deep learning. Based on these models and theoretical methods, new paths are provided for predicting precipitation in the future.

The prediction of precipitation in weather meteorology involves knowledge from several subject areas and requires a combination of factors. Such as geographic location, precipitation cyclicity, local hydrological conditions and sudden weather changes. For this reason, more meteorologists, by combining the evolving computer technology with other disciplines, we have proposed many new methods based on computer technology and statistics. Since the late 1960s, scholars based in this field have started to apply many methods from operations research, applied mathematics and statistics to precipitation in various places, listed as ARAIM model, SVM model and LSTM model.

The prediction of precipitation is often heavily constrained by factors from many sources. For example, changes in atmospheric circulation, differences in topography and geomorphology, differences in the location of land and sea, and the impact of human activities and so on. The changes of these factors usually have obvious instability, and the rule of change is difficult to accurately grasp. This instability is likely to lead to significant deviations in the predictions made by the relevant models, thus affecting the judgement of the precipitation situation and the formulation of response strategies.

In order to minimise the serious damage caused by meteorological disasters, we must make effective use of the precious water resources brought about by precipitation. For example, through the construction of reasonable water conservancy projects and the deployment of water resources, we can meet the demand for water for agricultural irrigation, industrial production and residential use.

Therefore, this paper will improve the accuracy of precipitation prediction based on machine learning as well as statistical and forecasting tools of big data for the precipitation condition in Suqian City, Jiangsu Province.

2.4. Content of this Paper

The aim of this paper is to study the monthly precipitation forecasts for Suqian City, Jiangsu Province, eastern China. This region is geographically unique and is subject to significant cyclical differences in the evolutionary characteristics of the rain belt and the northward push process of the climate-averaged monsoon in eastern China. Specifically, the centre of precipitation anomalies is located in the Jianghuai-North China region in July, while it moves southward to Central China in August. This phenomenon of anomalously high precipitation in the eastern region with inter-monthly differences is mainly affected by the north-east (south-west) deviation of the subtropical high pressure in the northwestern Pacific Ocean (Western Pacific Sub-High), the north (south) deviation of the subtropical westerly rapids in East Asia, and the persistent eastward extension of the high pressure in South Asia in July (August), etc.[4] .

As can be seen from Fig. 1, summer July precipitation in eastern China is mainly concentrated in the Jianghuai basin. From Fig. 2, it can be seen that the precipitation changes in Suqian City also show this trend. Precipitation in Suqian City is concentrated in June-August, and excessive rainfall and concentration in a short period of time are likely to cause problems such as flooding and urban waterlogging. Therefore, timely prediction of precipitation is crucial for the relevant departments to provide precipitation information to help agriculture, water conservancy and other departments to take timely measures to prevent floods and droughts, and to minimise the disaster losses caused by climate change[5] The following is a summary of the results of this study.

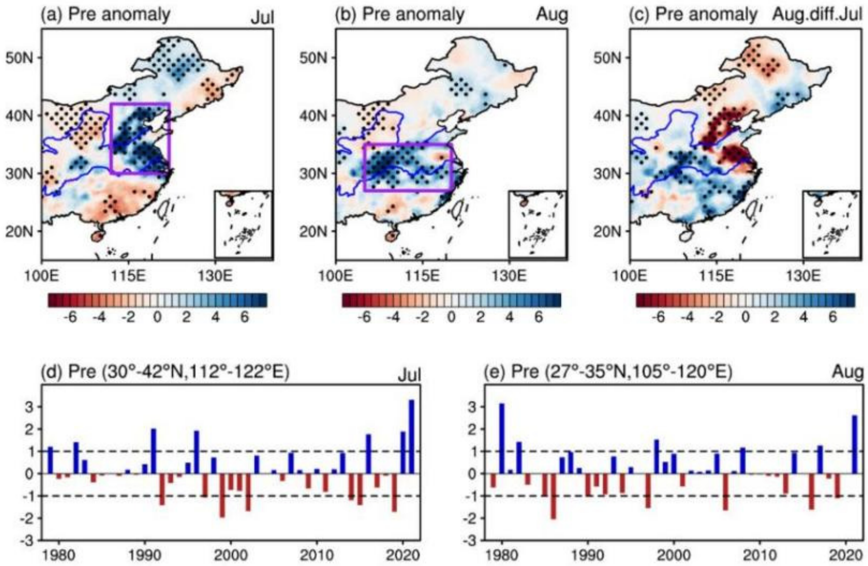

Fig 1. Eastern 2021 July-August Precipitation Schematic[4]

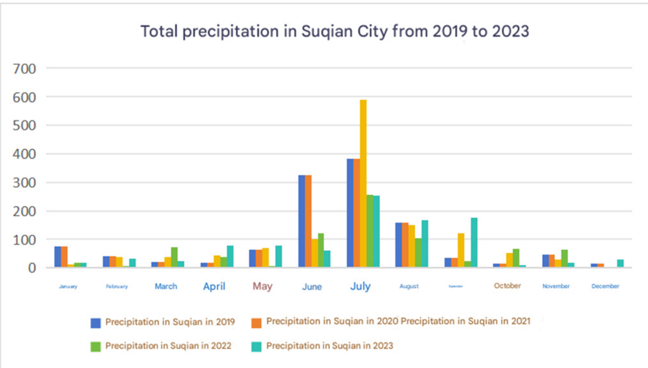

Fig 2. Schematic diagram of precipitation totals for Suqian City, 2019-2023

In this paper, the long and short-term memory neural network model based on the grey wolf optimisation algorithm is used to train the model on the monthly precipitation collected during the period of 2019-2022 and to predict the precipitation in Suqian in the coming year, so as to provide the corresponding numerical basis for the relevant departments to make strong protection measures against the occurrence of different natural disasters, to reduce the people's economic losses, and to promote the solid development of the economy.[5] The model is trained to predict the precipitation of Suqian in the coming year.

3. Modelling Preparation

3.1. Description of Symbols

3.2. Data Sources

The current data was obtained from the Xihe Energy Meteorological Big Data Platform, and the hourly precipitation (mm/h) for 2019-2023 was obtained from the local detection station in Suqian City (33.9645°N, 118.2697°E). This precipitation is the depth to which liquid or solid (after melting) precipitation falling from the sky to the ground accumulates on the horizontal surface without evaporation, infiltration, or loss[6] .

The data were collected from 0:00 on 1 January 2019 to 23:00 on 31 December 2023, with a total of 60 sets of data samples. In this paper, we will take the data from 0:00 on 1 January 2019 to 23:00 on 31 December 2022, with 744 hourly precipitation samples in each group, with a total of 48 data samples as the training set, and 12 sets of sample data for the whole year of 2023 as the test dataset, and the model will be trained and fitted to the test dataset. We will train the model and then give the predicted values and fit them to the test data set.

Table 1. Description of symbols

notation	Description of symbols
mm/h	Hourly precipitation
I_t	input gate
F_t	Oblivion Gate
O_t	output gate
C_t	Memory cells (candidate memory cells)
$\odot$	elemental multiplication
W_{xi},W_{xf} ,, ,$W_{xo}W_{hi}$ W_{hf} .W_{ho}	weighting parameter
b_i,b_f .b_o	Deviation parameters
H	Number of hidden units
X_t	time step
a	The leader of the pack, the first in the social hierarchy.
β	βWolves, in the second tier of the social hierarchy
δ	δWolves are in the third tier of the social hierarchy, consisting of pups, sentinel wolves comprise
ω	ωWolves, in the fourth tier of the social hierarchy

After that, we will give precipitation projections for the next 12 periods, i.e., rainy and non-rainy season precipitation data for the year 2024.

Table 2. Data features of the LSTM model dataset

data set	Starting and ending time of data	sample size
Training data	January 2019-December 2022	Group 48
Test data	January 2023-December 2023	12 groups

4. Data Pre-processing

4.1. Data Cleaning Accumulation

We took the acquired data, first sieved and washed it, and accumulated the sieved data into monthly precipitation totals to get the following table:

and draw images for analysis:

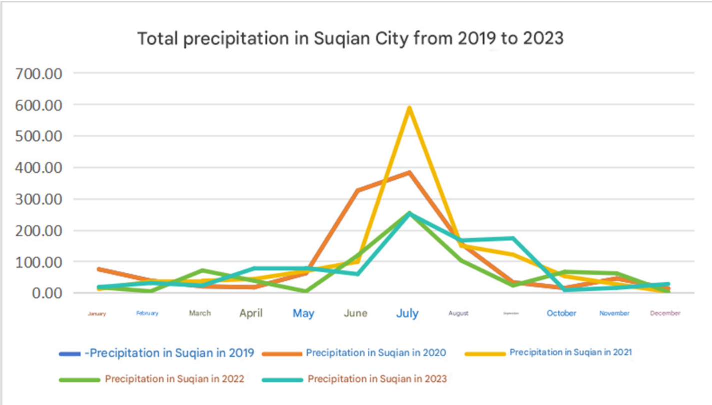

Fig 3. Monthly precipitation trends in Suqian City by year 2019-2023

According to Fig. 3, we can analyse that the monthly precipitation is concentrated in July-September, and its precipitation from October to June of the following year all shows a relatively flat state. The span between the rainy season and the non-rainy season is very large, so we will make separate predictions for the rainy season (July-September) and the non-rainy season (October-June) in the subsequent modelling, so that manually separating the seasonal fluctuations can greatly improve the accuracy of the prediction, and the prediction result is more gentle and stable.

4.2. Data Normalisation

We map each set of data linearly into the interval [0, 1], i.e., normalised, for all data as follows

For any set of data there is $X_i = (x_1, x_2, x_3, x_n, ..)$

$$x_i^{'} = \frac{x_i - \min x_i}{\max_{x_i} - \min_{x_i}} = 1, 2, 3..(1 \leq i \leq n) \tag{1}$$

Then inverse normalisation that is:

$$x_i = x_i^{'}(\max_{1\leq i\leq n} x_i - \min_{1\leq i\leq n} x_i) + \min_{1\leq i\leq n} x_i = 1, 2, 3... \tag{2}$$

4.3. Unit Root Test (ADF Test)

The unit root test is a statistical method used to test whether time series data has a unit root (i.e., whether it is non-stationary). In time series analysis, the unit root indicates that the series has a random walk characteristic, i.e., the series does not have a fixed mean and variance in time, and shows a constantly fluctuating trend. Therefore, we take the unit root test in data preprocessing to determine whether this set of data is a white noise series.

First make a raw sequence plot from the raw data:

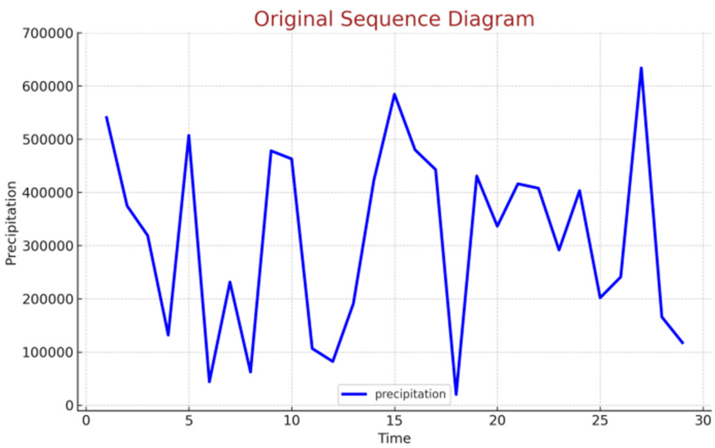

Fig 4. Original sequence diagram

The hypothesis is then formulated: the original hypothesis is that the series has a unit root, i.e., the series is non-stationary.

Selection of statistical tests: for this data set we chose the ADF test.

Calculate the test statistic: Perform a unit root test on the series and calculate the value of the test statistic.

Setting the level of significance: Determine the critical value for rejecting the original hypothesis based on the level of significance.

To determine whether to reject the original hypothesis: if the value of the test statistic is less than the critical value, the original hypothesis is rejected, indicating that the series is smooth; conversely, the original hypothesis is accepted, indicating that the series is non-smooth.[7]

Table 3 ADF test done according to the above steps.

Table 3. ADF inspection table

ADF Inspection Form							
variant	difference in order	t	P	AIC	threshold value		
					1%	5%	10%
measured quantity of rain	0	-1.398	0.583	563.354	-3.575	-2.924	-2.6
	1	-6.931	0.000***	550.374	-3.578	-2.925	-2.601
	2	-8.505	0.000**	555.489	-3.581	-2.927	-2.602

Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.

Based on the above ADF test table, it is only necessary to make a first order difference series plot to show that in the case of unit root test (ADF) used to analyse whether the time series is smooth or not and based on the variable precipitation, the significance p-value is 0.000*** at the level of significance at the difference of the 1st order, which presents significance and rejects the original hypothesis that the series is a smooth time series[8]

Fig 5. First-order difference sequence diagram

5. Modelling Ideas

The modelling idea of this paper is as follows

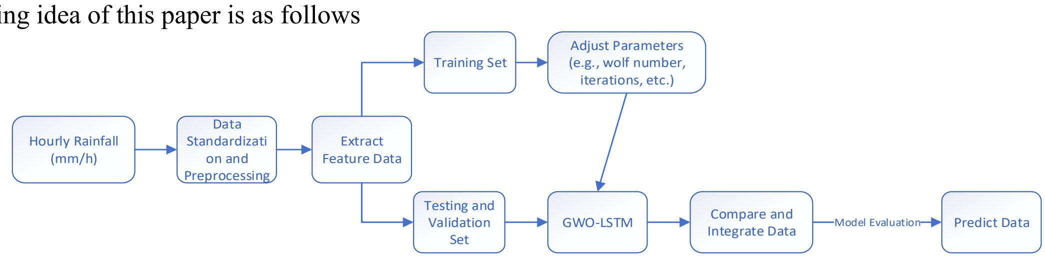

Fig 6. Modelling idea diagram

5.1. Grey Wolf Algorithm (GWO)

5.1.1. Algorithmic Thinking

In natural societies, wolves belong to packs and are at the top of the food chain. Grey wolves strictly adhere to a hierarchy of social dominance[1] The grey wolf strictly adheres to a hierarchy of social dominance.

Alpha wolf: The alpha wolf, responsible for making decisions and leading the hunt.

Beta Wolves: aids alpha wolves in decision making, maintains order in the pack, and can replace alpha wolves.

δ wolves: obey α wolves and β wolves, but rule over ω wolves.

ω Wolf: The lowest ranking wolf, responsible for obedience and giving information to the pack as a whole.

5.1.2. Algorithm Implementation Steps

Initialise group: randomly initialise the location of the grey wolf group.

Evaluating fitness: each grey wolf's fitness was evaluated according to the objective function and α, β and δ wolves were identified.

Update position: update the position of each grey wolf based on its behaviour in rounding up prey as follows:

(1) Rounding up prey:

$$\begin{cases} \boldsymbol{D} = \boldsymbol{C} \circ \mathbf{X}_p(t) - \mathbf{x}(t) \\ \mathbf{x}(t+1) = \mathbf{x}_p(t) - \boldsymbol{A} \circ \boldsymbol{D} \\ \boldsymbol{A} = 2\boldsymbol{a} \circ \boldsymbol{r}_1 - \boldsymbol{a} \\ \boldsymbol{C} = 2\boldsymbol{r}_2 \end{cases} \tag{3}$$

Where: t is the current iteration number. denotes the $hadamard$ multiplication operation; A and C are coefficient vectors; Xp denotes the position vector of the prey; X(t) denotes the current position vector of the grey wolf; a decreases linearly from 2 to 0 throughout the iteration; r1 and r2 are random vectors in [0, 1].[1] .

(2) Hunting:

Grey wolves have the ability to identify the location of potential prey (optimal solution), and the search process is mainly completed by the guidance of a ,β ,δ grey wolves. However, the solution space characteristics of many problems are unknown, and the grey wolf is unable to determine the exact location of the prey (optimal solution). In order to simulate the searching behaviour of grey wolves (candidate solutions), it is assumed that the grey wolves a ,β ,δ have a strong ability to identify the location of potential prey. Therefore, during each iteration, the best three grey wolves in the current population (a ,β ,δ) are retained, and then the positions of other search agents (including ω) are updated based on their position information.[8] The mathematical model of this behaviour can be expressed as follows The mathematical model of this behaviour can be represented as follows:

$$\begin{cases} D_\alpha = C_1 \circ \mathbf{X}_\alpha - \mathbf{X}, \quad D_\beta = C_2 \circ \mathbf{X}_\beta - \mathbf{X}, \quad D_\delta = C_3 \circ \mathbf{X}_\delta - \mathbf{X}, \\ \mathbf{X}_1 = \mathbf{X}_\alpha - A_1 \circ D_\alpha, \quad \mathbf{X}_2 = \mathbf{X}_\beta - A_2 \circ D_\beta, \quad \mathbf{X}_3 = \mathbf{X}_\delta - A_3 \circ D_\delta, \\ \mathbf{X}(t+1) = \dfrac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3}. \end{cases} \tag{4}$$

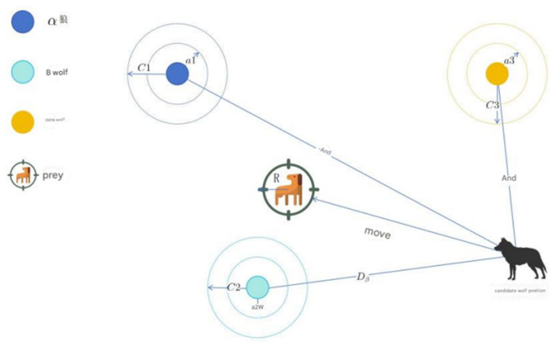

Fig 7. The grey wolf algorithm process

Updating the positions of α, β, and δ wolves: based on the new fitness values, the positions of α wolves, β wolves, and δ wolves are updated. And repeat the evaluation of fitness and updating fitness.

5.2. LSTM Model

In this paper, LSTM is selected as a prediction model, because the traditional Artificial Neural Network (ANN) cannot effectively utilise the time-series information, which is especially obvious in the meteorological field, as the meteorological data contains rich time-series information.[9] . Based on this problem, recurrent neural networks introduce recurrent connections, which allow neurons to maintain the information from the previous moment and pass it on to the subsequent time steps, thus effectively handling time-series information. RNNs perform well in handling time-series data and have been widely used in many fields.

The gradient vanishing problem has become particularly significant with the intensive study of recurrent neural networks (RNNs). This problem refers to the fact that as the network iterates through time, the information in the current situation is less affected by time, or even the effect of time on that time disappears, making it difficult for the network to learn long-term dependencies and retain long-term memories effectively. To solve this problem, researchers have proposed an improved RNN structure called Long Short-Term Memory (LSTM).LSTM introduces a gating mechanism, which includes a forgetting gate (forget-date), an input gate (input-date), and an output gate (output-date) three parts[10] to better control the information flow, thus effectively solving the gradient vanishing problem and enabling the network to learn and maintain long-term dependencies.

The inputs to the gates for long and short-term memory are all the current time step inputs X_t and the hidden state of the previous time step H_{t-1} , and the outputs are computed by a fully connected layer with an activation function that is a sigmoid function, and then all three gates have a value domain of [0,1].[9] .

Specifically, assume that the number of hidden cells is H , and given a small batch of inputs X_t at time step t and a hidden state H_{t-1} at the previous time step. The input gate I_t , the forget gate F_t and the output gate O_t of the time step t are computed as follows[11]:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \tag{5}$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \tag{6}$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \tag{7}$$

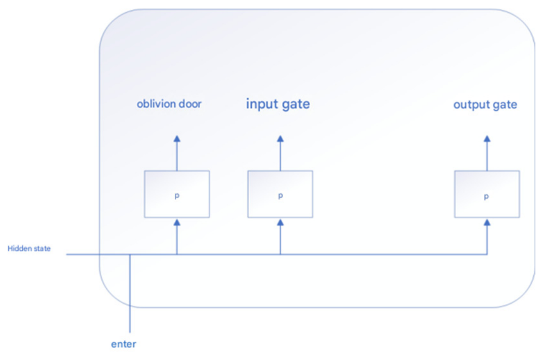

Fig 8. Computation of input, forgetting and output gates in long and short-term memory

The candidate memory cells C_t at time step t are calculated as

$$C_t = tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \quad (8)$$

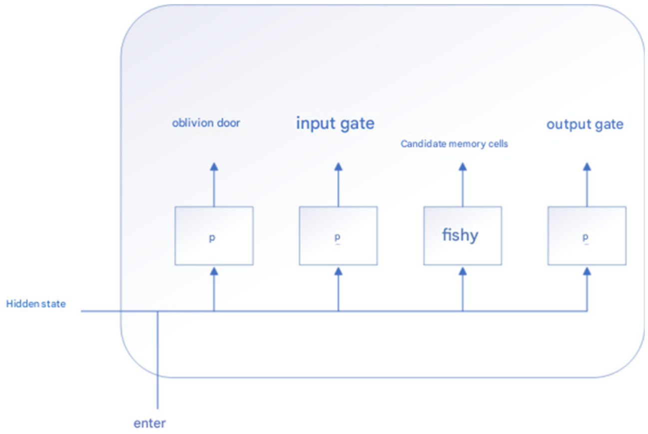

Fig 9. How to count candidate memory cells in long short-term memory (LSTM)

In the memory cell of a recurrent network, we can also introduce element value domains in the positive set of functions to control the flow of information in the hidden state by adding gate control cells, an approach that is also generally achieved by using per-element multiplication. This can effectively control the flow of information and help the network to better capture and utilise temporal information.

$$C_t = F_t \odot C_{t-1} + I_t \odot C_t \quad (9)$$

The forgetting gate controls whether the information in the memory cellC_{t-1} of the previous time step is passed to the current time step, while the input gate controls how the input X_t of the current time step flows through the candidate memory cellC_t to the memory cell of the current time step. If the forgetting gate is always approximated to 1 and the input gate is known to be approximated to 0, past memory cells will always be saved through time and passed to the current time step. This design can cope with the gradient decay problem in recurrent neural networks and better capture the dependency of large time step distances in time series[12] .

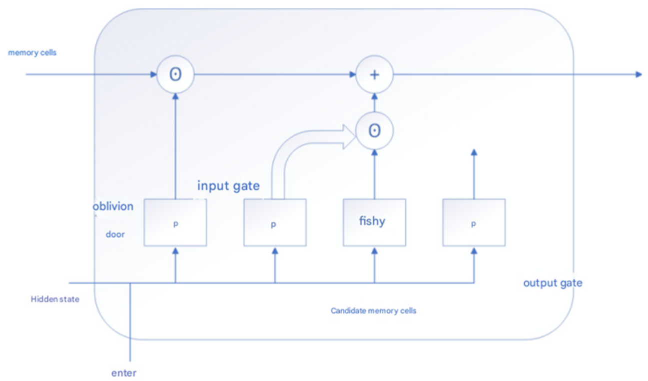

Fig 10. Memory cells in the LSTM model

With the memory cells in place, we can next control the flow of information from the memory cells to the hidden state H_t via output gates:

$$H_t = O_t \odot tanh(C_t) \quad (10)$$

The $tanh$ function here ensures that the hidden state element value is always between -1 and 1.

Special attention should be paid to the fact that when the value of the output gate tends to 1, the memory cell information at this time will be passed to the hidden state to provide the output layer to use; when the value of the output gate tends to 0, the memory cell information at this time will be retained by itself[12] .

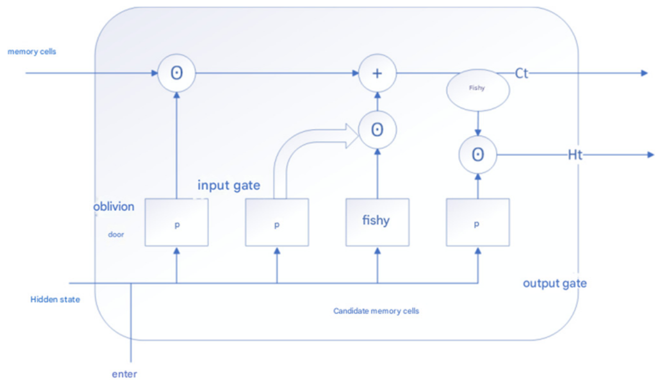

Fig 11. LSTM model hidden states

Then the LSTM modelling process in this paper is as follows:

Calculate the output values of the Long Short-Term Memory (LSTM) neural network:

Based on the inputs and the state of the memory cell at the current moment, the output value at the current moment is calculated.

Calculating an error term for the back propagation of memory cells in time and up one level of the network hierarchy[5] :

Based on the output value at the current moment and the desired output value, the error term is calculated. The error term is then back-propagated back to the upper level of the network and error propagation in the time series is considered.

Gradient process based on error term-by-error term calculation of weights:

Using the error terms, the gradients of each parameter (e.g., weights and biases) in the LSTM network are calculated one by one. These gradients reflect how much the parameters affect the error and can be used to adjust the parameters to reduce the error.

Based on the calculated gradient, a specific gradient optimisation algorithm is used to update the parameters in the LSTM network to minimise the error. This process is repeated until a certain number of training iterations is reached or the error reaches a certain level of convergence.

5.3. GWO-LSTM Optimisation Model

The GWO used in this paper, optimises the Long Short-Term Memory neural network (LSTM) for global search[1] . The specific steps are as follows:

5.3.1. Model Parameters

Given the number of data sets in this paper, we are given a range of [10, 100] for the number of neurons in the hidden layer, [0.001, 0.01] for the learning rate, [10, 100] for the number of trainings, and between [10,100] iterations.

5.3.2. Adaptation Function Adjustment

The selection of the fitness function plays a key role in the performance of the GWO-optimised network, which directly affects the performance of the model and hence the effectiveness of precipitation extraction. In this paper, GWO and LSTM are integrated to construct a monthly precipitation analysis and prediction model for Suqian City based on the GWO-LSTM model. The parameters are optimised to obtain a new learning rate and the number of neurons in the hidden layer. We continuously increase the number of training times in the expectation of obtaining the optimal combination, which further enhances the model's ability to deal with nonlinear problems. The optimised parameters are then brought into the LSTM model and combined with it to predict the future precipitation of Suqian using the model[13] .

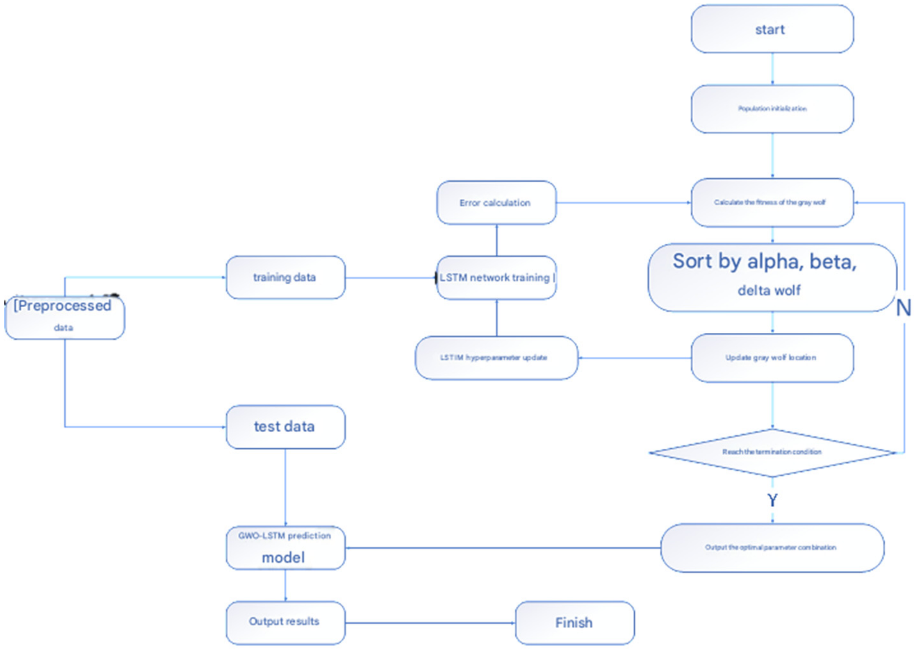

Fig 12. GWO-LSTM optimisation process

6. Analysis of Monthly Precipitation in Suqian City based on GWO-LSTM Model

This study is dedicated to exploring the way to combine the coyote optimisation algorithm (GWO) and the long short-term memory (LSTM) network to build a GWO-LSTM model that can be applied to the analysis of monthly precipitation in Suqian City. Firstly, the hyperparameters of the LSTM network are optimised by GWO, and the optimal learning rate, the number of neurons in the hidden layer, and the number of training times are specified to improve the nonlinear mapping ability of the model. Secondly, the optimised parameters are used to construct a GWO-LSTM model, which is used as a nonlinear transformation function of the monthly precipitation in Suqian City. Thirdly, the GWO-LSTM model is used as a nonlinear transformation function to predict the future monthly precipitation of Suqian City, and the optimal prediction results are obtained.

6.1. Specific Operational Procedures for the Model

In this paper, a GWO-LSTM model is developed with the aim of optimising the parameters of the Long Short-Term Memory Network (LSTM) by the Grey Wolf Optimisation (GWO) algorithm in order to improve its performance in dealing with nonlinear problems. The specific steps are as follows:

1. Selection of a suitable training dataset

In order to ensure the accuracy and generalisation ability of the model, a suitable training dataset is first selected. In this paper, the monthly precipitation data of Suqian City is selected, which has significant time series characteristics and nonlinear variations.

2. Optimising LSTM network parameters using the GWO model

In order to enhance the LSTM model to handle nonlinear problems, we optimise the hyperparameters of LSTM with the help of Grey Wolf Optimisation (GWO) algorithm. The optimised parameters include the learning rate, the number of hidden layer neurons and the number of training cycles. The optimal combination of these parameters can greatly improve the prediction accuracy and stability of the model.

3. Apply the best parameters derived from the model to the LSTM

The best combination of parameters for the LSTM model was obtained through the GWO optimisation process. These parameters include:

Learning Rate, Number of Hidden Neurons and Number of Training Epochs, these optimal parameters are applied to the LSTM model for training.

4. Training models

The LSTM model is trained using the optimised combination of parameters. During training, the model adjusts the internal weights according to the input data to minimise the prediction error. Through multiple iterations of training, the model gradually converges to an optimal state.

5. Extraction of forecasts of monthly precipitation in Cebu

After the training is completed, the trained LSTM model is used to predict the monthly precipitation in Suqian City. The performance and accuracy of the model can be evaluated by comparing the prediction results with the actual data.

The above is the optimisation process of LSTM network by GWO in this paper. By introducing GWO to optimise the parameters of the LSTM model, we effectively improve the accuracy and stability of the model in predicting the monthly precipitation in Suqian City.

6.2. Model Summary

In this chapter, the algorithms used in this study and their optimisation methods are described in detail. The LSTM optimised by the Grey Wolf optimisation algorithm used in this paper, we selected the optimisation capability of the Grey Wolf optimisation algorithm and the temporal processing of the long and short term memory model and continuously optimised it for precipitation seasonal forecasting. The computational process and advantages of these methods are described in detail in this paper.

Various algorithms have been investigated to better predict the precipitation season. At the beginning, we introduce the Grey Wolf Optimization Algorithm, a mathematical model based on grey wolves, which live in packs in nature, and whose predatory abilities enable them to find the optimal solution efficiently. Secondly, we introduce the long and short-term memory neural network, which is a special kind of recurrent neural network with strong memory ability, which will make our processing of temporal sequences more convenient and more suitable for processing time series data. Finally, we apply the grey wolf optimisation algorithm to optimise the hyperparameters of the LSTM network, which improves the prediction performance of the model. Through the study in this chapter, we delve into the implementation details and advantages of these algorithms, which provide important theoretical support and methodological guidance for precipitation season prediction.

7. Precipitation Analysis of Suqian City, Jiangsu Province based on LSTM Model

7.1. Training and Validation Data Sets

Firstly, we divided the year December into two parts rainy season (July-September) and non-rainy season (October-June) based on the fact that the dataset has significant seasonal fluctuations.

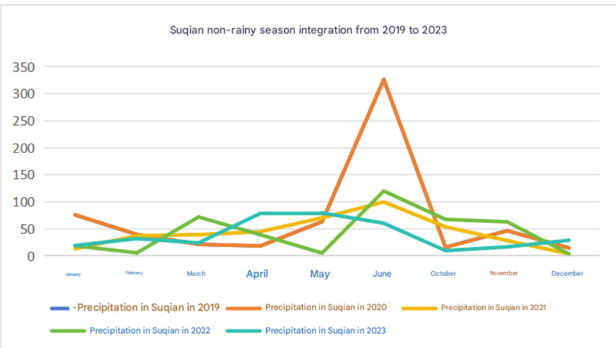

Fig 13. Cebu City 2019-2023 non-rainy season precipitation

LSTM models have a strong memory capability to better capture and exploit long-term dependencies in sequential data, and therefore have a significant advantage in solving nonlinear problems.

Meanwhile, the model performs well on the training set. However, in the rainy season, the precipitation in the region is in July and August, reaching the highest value, due to the small dataset collected in this time, it is not possible to give a more accurate value, the value given by this model after training is more conservative. And in the non-rainy season is relatively smooth, the region is in the subtropical monsoon zone to the temperate monsoon zone transition zone, its influence by the monsoon precipitation has uncertainty, such as the model prediction value and the real value, there is a large error.

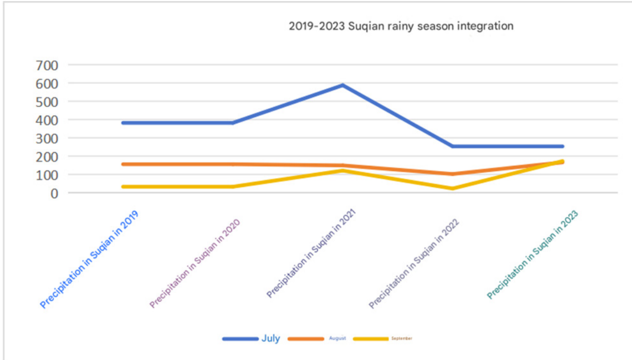

Fig 14. Rainy season precipitation in Cebu City, 2019-2023

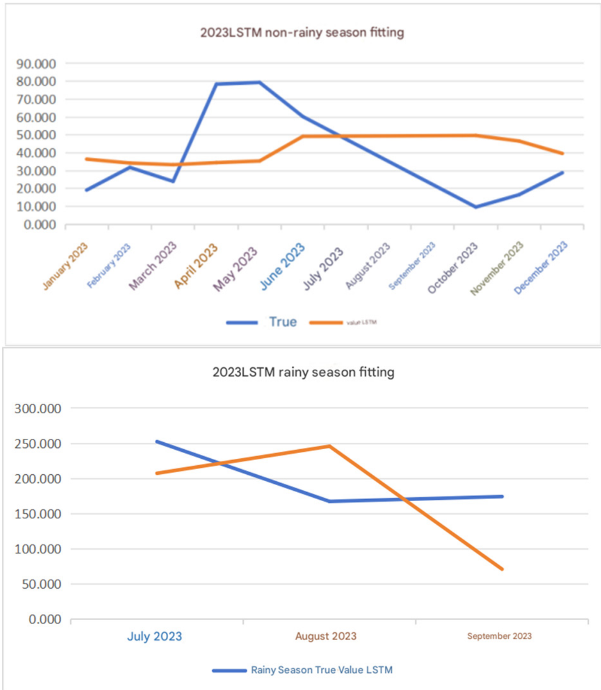

Fig 15. LSTM predictive value training fit

7.2. Performance Evaluation Indicators

Two models are constructed in this paper, the Long Short Term Memory Neural Network (LSTM) model and the Grey Wolf Optimisation Algorithm based Long Short-Term Memory Neural Network (GWO-LSTM) model. These two models were used for adaptation operations on training data and test data respectively, and the predictive performance of the models was measured by comparing their results against each other. This paper employs the following means to determine the degree of model fitness.

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

Mean Absolute Percentage Error (Mean Absolute Error)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{12}$$

Table 4. Error values for LSTM rainy and non-rainy seasons

	RMSE	MAE
LSTM rainy season fitting	79.28649858	61.39871118
LSTM non-rainy season fitting	27.90861496	5.538797209

8. Combining the GWO-LSTM Model

8.1. Empirical Analysis of the Model

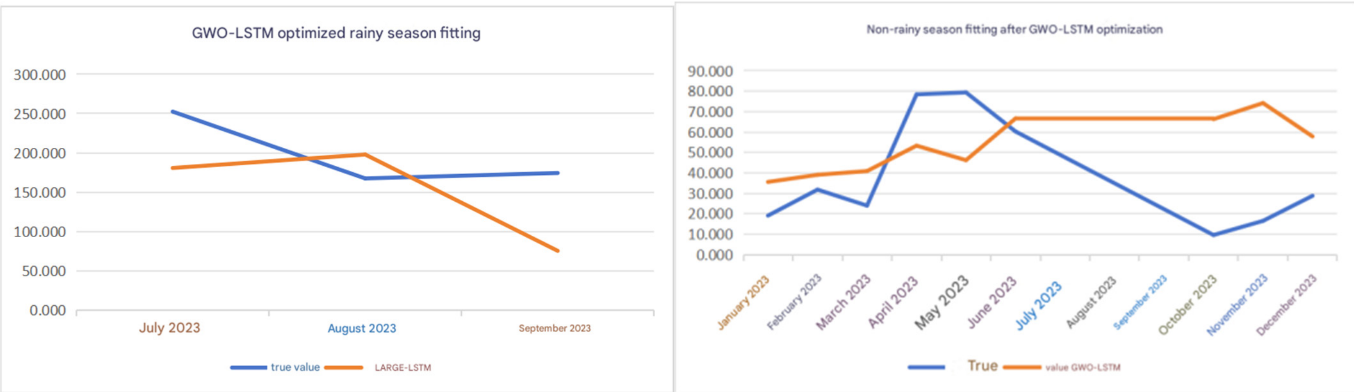

Fig 16. GWO-LSTM predicted values

The optimised LSTM model after the GWO model effectively compensates for the shortcomings of this dataset and gives a more accurate rainy season prediction, with trends and precipitation values roughly similar to the true values. Its error is significantly improved compared to the pre-optimisation period. However, given the small dataset this time, the predictions for the non-rainy season, have a greater sense of fluctuation from the true values, such as February, April, and November, and it is obvious from the figure that there is a significant discrepancy in this prediction for the non-rainy season.

8.2. Analysis of Model Results

Table 5. Values after GWO optimisation

	RMSE	MAE	Optimised RMSE	Optimised MAE
GWO-LSTM rainy season fitting	79.28649858	61.39871118	72.80451213	46.82695069
GWO-LSTM non-rainy season fitting	27.90861496	5.538797209	19.62558965	5.37258257

Detailed error assessment and optimisation tuning were carried out during the fitting of rainy and non-rainy season data using the GWO-LSTM model. The initial evaluation showed that the root mean square error (RMSE) for the rainy season data was 79.28649858 and the mean absolute error (MAE) was 61.39871118. This indicates that the model has relatively low prediction accuracy on the rainy season data. Similarly, the initial RMSE for non-rainy season data was 27.90861496 and MAE was 5.538797209, which showed that the model performed better on non-rainy season data, but there is still room for improvement.

To improve the prediction performance of the model, we optimised the LSTM model using the GWO model. The optimised results show a significant reduction in RMSE to 72.80451213 and MAE to 46.82695069 for rainy season data.This indicates a significant improvement in the prediction accuracy of the model on rainy season data. For non-rainy season data, the optimised RMSE was reduced to 19.62558965 and MAE was slightly reduced to 5.37258257, which further improved the accuracy of the model.

Overall, through the optimisation, we significantly improved the prediction performance of the LSTM model on both rainy and non-rainy season data, reducing the error and increasing the reliability and applicability of the model.

9. Comparison of Models

9.1. Grey Prediction GM (1,1)

Grey prediction is a prediction method based on grey system theory, which is suitable for systems with incomplete information, small sample data or lack of clear laws. It is mainly used for short-term forecasting and trend prediction, and is often applied in the fields of economy, management, environment and so on.

We will divide into rainy and non-rainy seasons for their respective levels of data prediction, and the predicted values will be analysed for error with the actual monthly precipitation:

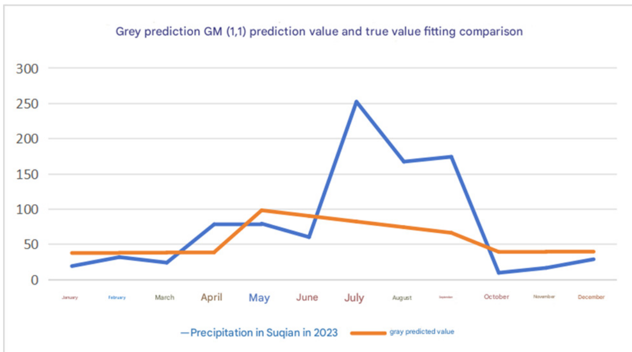

Fig 17. Grey predicted fitted values

The grey predicted GM (1,1) value has an error of 260.1% from the true value, which is too large and implies a poor model fit.

9.2. Seasonal ARMIA Modelling

The Seasonal Autoregressive Moving Average Model (Seasonal ARIMA) is a statistical model for time series analysis and forecasting. It is an extension of the Autoregressive Integral Moving Average Model (ARIMA) and is specifically designed to deal with time series data with seasonal characteristics.[14][15] It is an extension of the autoregressive integral moving average model (ARIMA).

The ARIMA model is a commonly used time series modelling method that takes into account the autoregressive, difference and moving average components of the series. However, for data with significant seasonal characteristics, the traditional ARIMA model may not be sufficient to capture the effects of seasonal variations. This is where the use of a seasonal ARIMA model comes into play.[14]

The construction of a seasonal ARIMA model usually involves seasonal decomposition of the time series data to determine the appropriate seasonal order (P, D, Q, s). The parameters of the model can then be estimated using maximum likelihood estimation or other methods, leading to forecasting and analysis.

We brought the data to the seasonal ARMIA model to get the projected monthly precipitation for 2023.

The model results are shown as SARIMAX (0, 0, 0) × (2, 0, 0, 12). According to the variable precipitation, the results of the residual Q statistic were analysed and it was found that Q6 was not significant at the level, and the hypothesis that the model's residuals were white noise series could not be rejected, so the model basically met the requirements; however, the model's goodness-of-fit, R^2, was 0.495, which was a poor performance of the model. [15]

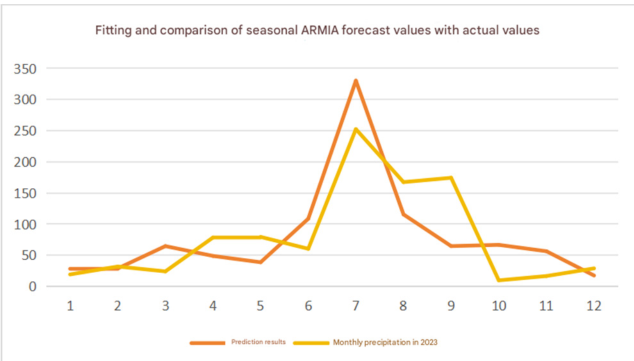

Fig 18. Seasonal ARIAM predictions vs. actual fit

10. Conclusion

10.1.Summary

1.This study uses the precipitation data provided by the Xixi and Energy Meteorological Data Platform for Suqian City from 2019 to 2023 for short-term precipitation forecasting, and the GWO-LSTM model proposed in this paper can be used to improve the accuracy of the LSTM with its strong memory capability as a means of predicting the precipitation in Suqian City for the next 12 months. The short-term precipitation forecast for Suqian City provides a more accurate model method, which is not only conducive to the safety of citizens travelling and playing pleasure, but also provides great convenience in agricultural production and transportation, which helps Suqian City to flourish in the basic industries and accelerate the development of the city to the intelligent city.

2.Compared with the conventional LSTM model, the approach mentioned in this paper has a better feature selection process and has good results. However, it is difficult to analyse these nonlinear factors because precipitation may be affected by many linear and nonlinear factors, some of which are potential. This prediction analyses the precipitation in Suqian City, Jiangsu Province, by using the LSTM model and the GWO-based long and short-term memory neural network, which has significant advantages in dealing with nonlinear problems.

3.By comparing the grey prediction GM with the seasonal ARMIA model, the error of the model is larger and the model performance is poorer, this study can provide some reference for subsequent researchers to further improve the prediction accuracy.

4.Precipitation data for the rainy season in Suqian is highly variable, with maximum monthly precipitation of up to 700mm, indicating a high probability of extreme weather in Suqian and erratic precipitation, making prediction very challenging. Precipitation in the rainy season is quite high compared to the non-rainy season. Precipitation in the non-rainy season is more stable, with less than 100mm. The rainy season precipitation can reach up to 600mm in a single month.

5.Since the literature on GWO-long and short-term memory neural network models for precipitation prediction is relatively scarce, this paper is informative in related aspects.

10.2.Recommendations

Due to the less precipitation data collected in this paper, after optimisation using the GWO model, although there is a significant improvement, but the prediction of the non-rainy season our model gives a greater volatility of precipitation in the non-rainy season, followed by the rainy season is more conservative, later we need to collect more meteorological precipitation data, and constantly train to improve the model.

Strengthen the construction and improvement of water conservancy projects and increase investment in water conservancy projects and facilities within Suqian City to ensure that urban traffic is not affected by rainfall and to improve flood and drought prevention.

Table 6. Precipitation for the next 12 months given by the GWO-LSTM model

months	January	second month (of the lunar year)	March	April	May	June
measured quantity of rain	40.37	30.67	35.50	39.71	56.22	186.41
months	July	August	September	October	November	December
measured quantity of rain	372.70	147.16	77.76	32.47	40.05	13.005

Acknowledgments

The authors are supported by the University Students' Innovation and Entrepreneurship Training Program of Suqian University (No.202414160103Y).

References

[1] Lu Chao. Research on multi-objective shop floor scheduling problem with controllable machining time [D]. Huazhong University of Science and Technology, 2017.

[2] Hong J. Research on domain concept extraction method based on Deep Learning [D]. East China Normal University, 2014.

[3] Cao, Longchun. Research on prediction algorithm for long-term multi-step call volume [D]. Zhejiang University of Technology, 2020. doi: 10.27463/d.cnki.gzgyu.2020.001423 .

[4] MA Xiaoyi, FAN Ke, YANG Hongqing. Causes and predictability of the inter-monthly evolution of extreme precipitation in eastern China during the summer of 2021 [J/OL]. Journal of Atmospheric Sciences, 1-21[2024-05-02].

[5] ZHANG Xing, GUAN Yue, DAN Xinxin, et al. Research on monthly precipitation prediction based on genetic algorithm-long and short-term memory neural network[C]// Chinese Society of Statistical Education, Steering Committee for Teaching Statistics in Higher Education Institutions of the Ministry of Education, National Steering Committee for Postgraduate Education of Professional Degree in Applied Statistics. 2021 (The Seventh) National Student Statistical Modelling Competition Awarded Papers Collection (II). Guangxi University of Science and Technology, 2021: 41. DOI:10.26914/c.cnkihy.2021.044013.

[6] MA Xiaolei, REN Linghui. Impact of seasonal snowpack changes on runoff in the Tarim River Basin[J]. Water Resources Science and Economy, 2024, 30 (04): 70-74.

[7] Shen Xiafen, Cai Qiang, Wu Lifang,et al. Food safety risk factor identification and prediction method based on PCA and ARIMA [J]. Food Industry, 2021, 42 (10): 332-337.

[8] Zhang Wanlin. Analysis of logistics demand forecasting in Ningbo port based on GM (1,1)-ARIMA combined model [J]. Logistics Engineering and Management, 2023, 45 (01): 32-35.

[9] Gu Jianfang. Comparison of applications of ARIMA and LSTM algorithms [J]. Digital Technology and Applications, 2022, 40 (01): 58-60. DOI:10.19695/j.cnki.cn12-1369. 2022. 01. 19

[10] Xu Rongnan. Research on multimodal indoor positioning method with weak correlation of location privacy [D]. Yanshan University, 2023. doi: 10.27440/d.cnki.gysdu.2023.000933.

[11] Yu Daben. Research on operation optimisation of urban car sharing system based on spatio-temporal big data [D]. Southwest Jiaotong University, 2021. DOI:10.27414/ d.cnki. gxnju. 2021.000037.

[12] Zhu Honggen, Tian Liqin, Wei Jun, et al. LSTM combinatorial prediction model approach and its application [J]. Software, 2020, 41 (11): 79-84.

[13] YIN Wenwen, JIN Qiu. Tourism planning of Tianjin central city based on grey wolf optimisation algorithm [J]. Digital Technology and Application, 2023, 41 (06): 134-136+172. DOI: 10.19695/j.cnki.cn12-1369.2023.06.42.

[14] XIAO Ying, XIAO Xiangwei, DUAN Zhuang, et al. Flow field prediction of centrifugal pumps using intrinsic orthogonal decomposition and long- and short-term memory network models [J/OL]. Journal of Xi'an Jiaotong University, 1-11 [2024-07-12]. http://kns. cnki.net/kcms/ detail/61.1069. T. 20240703. 1631.002.html.

[15] YU Mu, YIN Weiqin,YANG Le. Prediction of carbapenem-resistant Acinetobacter baumannii prevalence trends based on ARIMA product seasonal model [J/OL]. Chinese Journal of Hospital Infection, 2024, (17): 2575-2579 [2024-07-12]. http://kns.cnki.net/kcms/detail/11.3436.R.20240703.1923.006 .html.