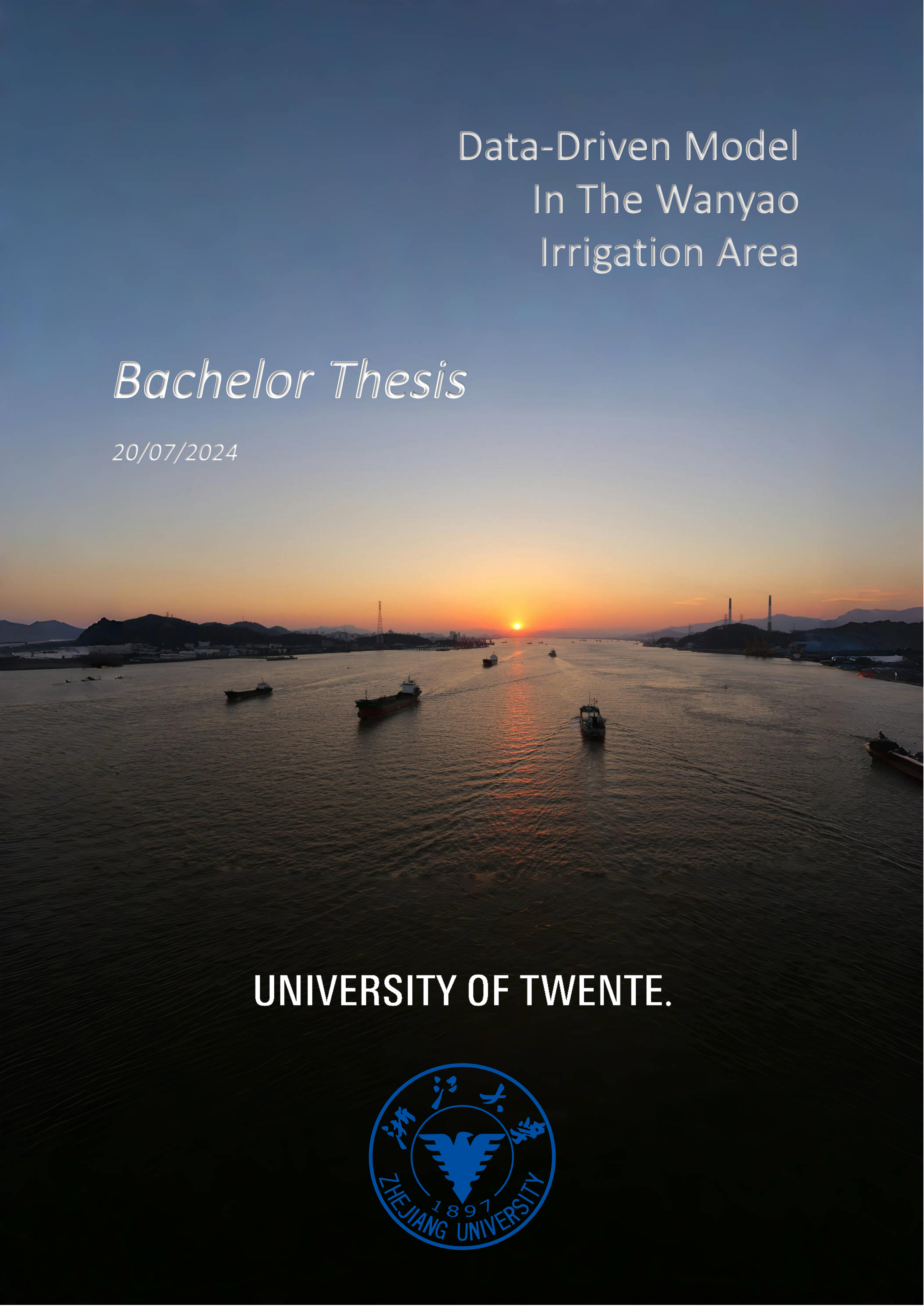
Data-Driven Model In The Wanyao Irrigation Area

Bachelor Thesis

20/07/2024



UNIVERSITY OF TWENTE.

浙江大学 ZHEJIANG UNIVERSITY 1897

Author	Yao Wang
Date	20/07/2024
Student Number	S3038130
Study	Bachelor Civil Engineering
Institution	University of Twente
Department	Department of Civil Engineering and Management
University of Twente Supervisor	Oleksandr Mialyk
External Supervisor	Yueping Xu
Cooperative Study Institution	Zhejiang University

Preface

This thesis is the final result of my undergraduate thesis research on "Data-driven model in the Wanyao irrigation area". This thesis was written when I graduated from the Bachelor of Civil Engineering at the University of Twente. To do so, I conducted a three-month research in the Hydrology and Water Resources Research Group of the Department of Civil Engineering in Zhejiang University. It is my hope that this research contributes to the advancement of water resource management practices in irrigation areas and inspires further research in the field of data-driven modelling.

First of all, I would like to express my deepest gratitude to my supervisor, Mr. Oleksandr Mialyk from the University of Twente, whose expertise, understanding, and patience greatly enriched my undergraduate project experience. I am particularly grateful for his support and guidance throughout the process. I would also like to thank him for the many hours spent reading and revising drafts, as well as for the many fruitful discussions. Without him, I would not have been able to complete this thesis.

Secondly, I would like to thank Professor Xu Yueping from Zhejiang University for giving me the greatest help during the course of my graduation assignment. When I encountered difficulties and challenges in my research, her reminders and advice always gave me new inspiration and direction. Her expertise and encouragement have been instrumental in shaping this study and its outcomes. Her insights and feedback have greatly improved the quality of my research work. Her commitment to excellence inspires me, and I am truly fortunate to have had the opportunity to work with her.

In addition, I would like to thank Dr. Martijn Booij for his efforts in finding my undergraduate graduation project, and Ms. Oosterzee-Nootenboom for helping me solve the problems I encountered during the paper submission process. I would also like to thank Zejian from the Zhejiang University research group for his help in building the model and Siwei for helping me solve the problems I encountered in my life while conducting project research at Zhejiang University. I am very grateful to every partner in the Zhejiang University research group. From the moment I started to the last day, they made me feel at home and spared no effort to help me.

Finally, thankful for the most ordinary myself, steadily progressing on the plain and continuous journey.

Table of contents

Table of figures

Table of tables

1 Introduction

Hydrological modelling is crucial for understanding and managing water resources, which are vital for human survival and ecosystem health. With climate change and increasing anthropogenic pressures, effective water resource management is becoming more challenging and essential. Accurate hydrological models can help predict water availability, manage flood risks, and ensure sustainable water use for agriculture, industry, and domestic needs.

In particular, this thesis addresses the need for improved hydrological modelling in the Wanyao irrigation area. Traditional models, while useful, often have limitations in accuracy and computational efficiency. This research aims to develop a more efficient and accurate hydrological model using advanced machine learning techniques.

Hydrological modelling is needed to simulate and predict the movement, distribution, and quality of water within natural and artificial systems. It helps in understanding the water cycle, assessing the impact of land use and climate changes, and managing water resources effectively. These models are essential tools for planning and decision-making in water resource management, flood forecasting, and environmental protection. For example, Soil and Water Assessment Tool (SWAT) is a widely used, semi-distributed hydrological model that simulates the effects of land management practices on water, sediment, and agricultural chemical yields in large, complex watersheds. It has been effectively used to predict the impact of land use changes and climate variability on water resources (Gassman et al., 2014).

- Several approaches have been used in previous studies to simplify hydrological modelling, including: Common AI Models in Hydrological Modelling: Machine learning and AI models such as Artificial Neural Networks (ANNs), Long Short-Term Memory networks (LSTMs), and Convolutional Neural Networks (CNNs) have been employed to improve the accuracy of hydrological predictions by learning complex patterns in data(Nourani et al., 2009).
- Transformer Models: Recently, transformer models have been introduced in hydrological modelling due to their ability to handle sequential data and capture long-range dependencies effectively. Transformers have shown promise in improving the efficiency and accuracy of hydrological predictions compared to traditional methods .

Despite these advancements, there are still gaps and challenges that need to be addressed. Existing hydrological models like SWAT require extensive calibration and high-quality data, which can be labour-intensive and time-consuming. They may also struggle with capturing the complex, nonlinear relationships inherent in hydrological systems. Additionally, common AI models, while powerful, often require significant computational resources and may not efficiently process large datasets .

This thesis proposes the development and application of a transformer-based model for hydrological prediction in the Wanyao irrigation area. The transformer model leverages attention mechanisms to improve prediction accuracy and computational efficiency. By reducing the need for extensive parameter calibration and efficiently processing large datasets, the transformer model aims to overcome the limitations of traditional and existing AI models.

The general structure of the thesis is as follows: The introduction section introduces the research topic and its importance in the Wanyao irrigation district, providing a background for the study. The theoretical background section lays the foundation to guide the study by introducing the research model and discussing the coefficients for evaluating the model performance. The methodology chapter details the methods used in the study, including describing the study area, the use of SWAT for water cycle modelling, and the overall approach adopted. Then the gaps are identified in existing models, outlines the research objectives, and introduces the Transformer model development process as a potential solution. The model analysis section then analyses the data, parameters, and techniques used in the study to evaluate the performance of the Transformer model in water cycle modelling. The results are explained in the discussion chapter, the implications of the research results are discussed, and an insight into the significance of the research results is provided. Finally, the main findings are summarized, the contributions of the study are emphasized, and future research directions are proposed to advance knowledge in this field.

2 Theoretical Background

2.1 Research model

Deep learning is a subfield of machine learning that focuses on algorithms inspired by the structure and function of the brain called artificial neural networks. Neural networks are computational models composed of interconnected nodes, called neurons, that work together to process complex information. In deep learning, neural networks with multiple layers (deep neural networks) are used to learn representations of data through a hierarchical learning process. These networks can automatically discover patterns and features in data, making them powerful tools for tasks such as image recognition, natural language processing, and speech recognition(Serrano, 2017).

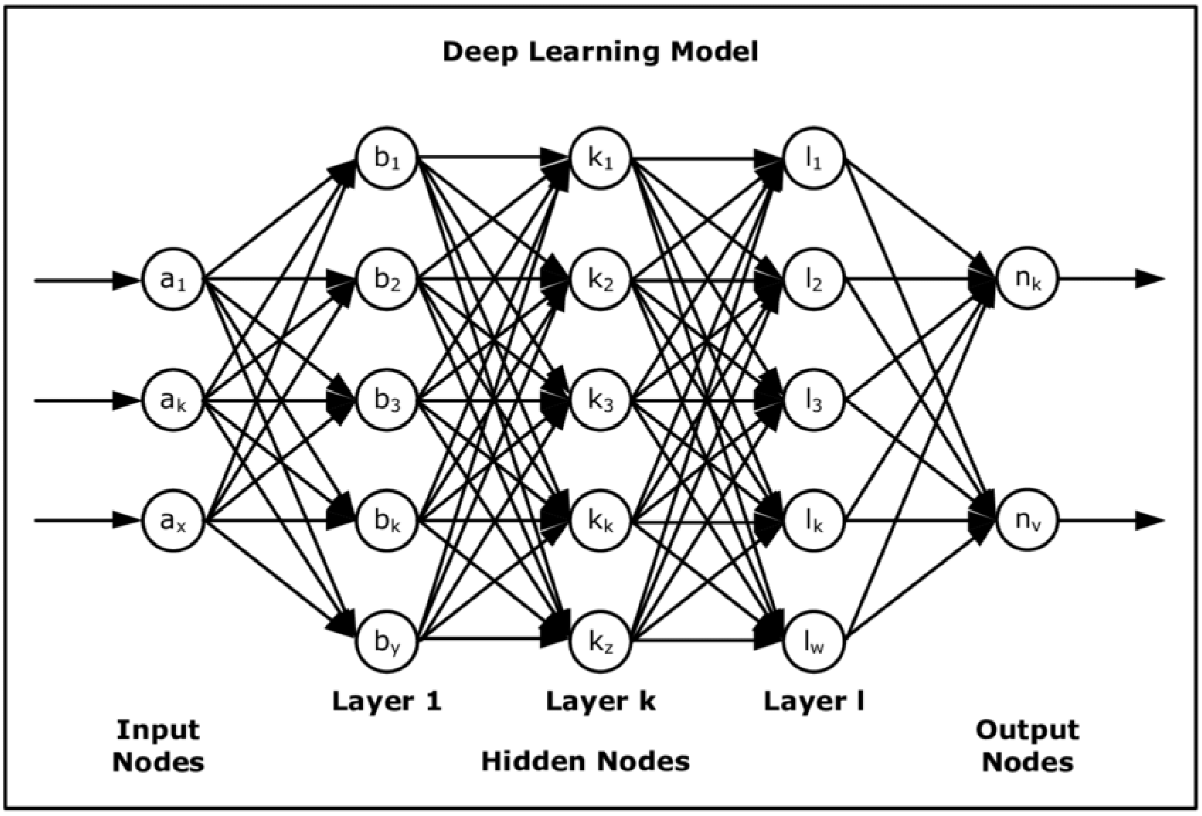


Figure 1: Artificial Neural Network-Deep Learning model(Serrano, 2017)

The Transformer model is a neural network architecture that relies entirely on attention mechanisms for sequence transduction tasks. It replaces recurrent layers commonly used in encoder-decoder architectures with multi-headed self-attention. The model consists of stacked self-attention and fully connected layers for both the encoder and decoder(Figure 2).

The encoder and decoder each have multiple identical layers, with each layer containing sub-layers for self-attention and feed-forward networks. The model allows for more parallelization, faster training, and has shown superior performance in tasks like machine translation(Vaswani et al., 2023).

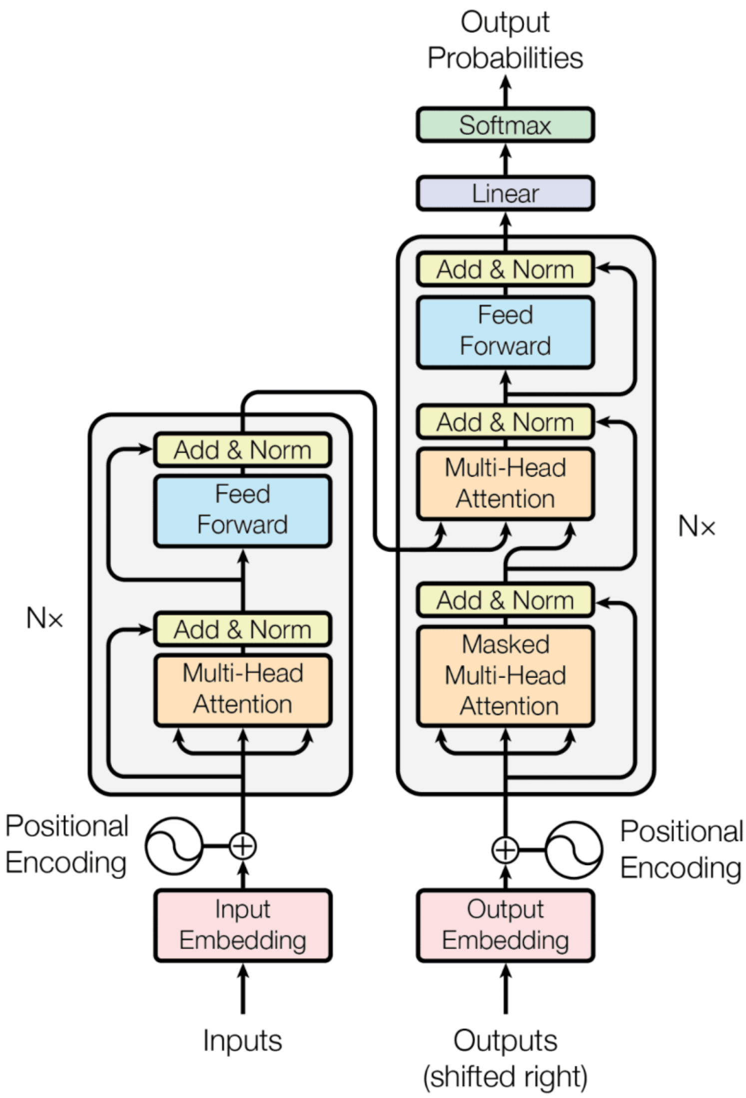


Figure 2: The Transformer - model architecture(Vaswani et al., 2023)

The Transformer model works by utilizing self-attention mechanisms to capture dependencies between input and output sequences without the need for recurrent or convolutional layers. Here is an overview of how the Transformer model operates(Vaswani et al., 2023) as shown in Figure 2:

- Encoder and Decoder Stacks: The Transformer consists of stacked encoder and decoder modules. The encoder stack comprises multiple identical layers, each containing a self-attention mechanism and a feed-forward network. The decoder stack also consists of identical layers but includes an additional attention mechanism over the encoder's output. Residual connections and layer normalization are applied around each sub-layer to facilitate learning.

- Self-Attention Mechanism: The core component of the Transformer is the self-attention mechanism. Self-attention allows the model to weigh the importance of different input positions when predicting the output at a particular position. It computes attention scores by comparing each input position with every other position, capturing global dependencies efficiently.
- Multi-Head Attention: To enhance the model's ability to focus on different parts of the input sequence, multi-head attention is employed. Multi-head attention involves projecting the input into multiple subspaces and performing attention in parallel, allowing the model to attend to different parts of the sequence simultaneously.
- Position-wise Feed-Forward Networks: In addition to self-attention, the Transformer includes position-wise feed-forward networks in each layer to further process the information captured by attention mechanisms.
- Masking in the Decoder: To ensure that the decoder does not peek at future information during training, masking is applied to prevent positions from attending to subsequent positions. This masking, combined with offsetting the output embeddings by one position, helps the model make predictions based only on previously generated outputs.
- Training and Inference: During training, the model is optimized to minimize a loss function that measures the dissimilarity between predicted and actual outputs. During inference, beam search is typically used to generate translations by considering multiple candidate sequences and selecting the most likely one based on the model's predictions.

Overall, the Transformer model's architecture allows for efficient parallelization, capturing long-range dependencies, and achieving state-of-the-art performance in various sequence transduction tasks like machine translation.

2.2 Model performance evaluation coefficient

In the evaluation of hydrological models, the Nash-Sutcliffe Efficiency (NSE) is generally used to measure the accuracy of model predictions. NSE provides a relative assessment of model performance by comparing the model prediction error with the mean error of the observed data.

The definition of NSE is as follows:

$$NSE = 1 - \sum_{i=1}^{n} \frac{(O_i - S_i)^2}{(O_i - \bar{O})^2}$$

Where:

O_i is the observed values at time i

S_i is the simulated values at time i

$\bar{O}$ is the average of the observed values

n is the number of time steps

NSE values range from $-\infty$ to 1 and have a clear interpretation. The closer the NSE value is to 1, the closer the model prediction is to the observed data and the better the performance. A value of 0 means that the model prediction ability is the same as using the observed mean as the prediction, and a negative value means that the model prediction performance is worse than using the observed mean(Nash & Sutcliffe, 1970).

NSE is highly sensitive to prediction errors, especially large errors. This means that it can effectively identify significant deviations from the model at certain times, providing a strong basis for adjusting and improving the model. Therefore, it is widely used in hydrological model evaluation and is recognized and used by many researchers and engineers. This standardized evaluation method helps to compare the research results of the two models, SWAT and Transformer.

Considering that the time scale is smaller (daily), the adjusted ratings in evaluating the ANN model performance developed by Kalin et al. (2010) were adapted in this study:

Very good: $E_{Nash} \geq 0.70$;

Good: $0.5 \leq E_{Nash} < 0.7$

Satisfactory: $0.3 \leq E_{Nash} < 0.5$

Unsatisfactory: $E_{Nash} < 0.3$

3 Methodology

3.1 Study area

The Wanyao irrigation area is located in the southwest of Zhejiang Province, at the western end of the Jinqu Basin, at the junction of Zhejiang, Fujian, and Jiangxi provinces, with geographical coordinates ranging from 118°22′29″ to 118°48′48″ east longitude and from 28°14′29″ to 28°53′24″ north latitude. The total land area of this area is 5,435 km². The irrigation scope mainly includes 12 towns and 2 streets in Jiangshan City, with a total population of 459,000.

The water system within the Wanyao irrigation area is part of both the Qiantang River system, including the Jiangshangang, Changshangang, Wuxi River, and the Yangtze River system, including the Xinjiang River basin of Poyang Lake. The primary research area mentioned in this proposal is the Qiantang River's Wanyao Irrigation Area. The Qiantang River basin shown in Figure 3 is located in the western part of Zhejiang Province and is the largest river in the province, with two main sources, the North and South. The total drainage area is 1809.1 km². The main tributaries include the main stream of Jiangshangang, Changtai Creek, Dahe Creek, as well as numerous smaller tributaries such as Sanqingkou Creek, Guangdu Creek, Sierdu Creek, Long Creek, Qingyangdian Creek, Sanqiao Creek, Fengmen Creek, Fengzu Creek, and Xielixi Creek.

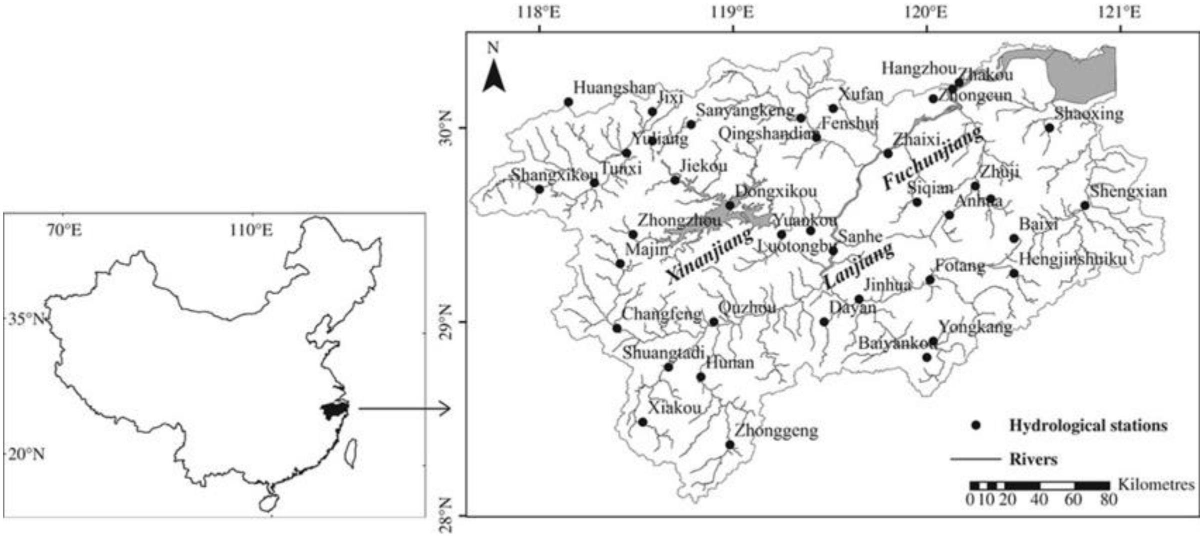


Figure 3: Qiantang River Basin(Ma et al., 2018)

UNIVERSITY OF TWENTE.

3.2 Water cycle modelling in irrigation area

The natural water circulation system in the Wanyao irrigation area and the artificial water circulation system, artificial water circulating subsystems (urban and rural water supply, farmland irrigation, hydropower and other systems) exist in mutual coupling and mutual assembly relationship.

Because the data input of the SWAT model is mostly distributed, the spatial distribution of meteorological factors such as precipitation and temperature can be fully considered to further determine the spatial distribution of runoff and water resources. Therefore, in accordance with the water circulation rules and characteristics of the irrigation area, the widely used distributed hydrological model in this irrigation area is SWAT.

3.2.1 Soil Water Assessment Tool (SWAT)

SWAT is a quasi-distributed watershed model simulating the movement of water, sediment, nutrient, crop growth, nutrient cycling, etc. in a watershed. It is a conceptual hydrologic model, operating at daily and sub-daily time steps (Arnold et al., 2012).

SWAT has widely been used for assessing water resources and nonpoint source pollution problems. Input information for each sub-watershed includes weather, soil properties, topography, and vegetation. The sub-watersheds are divided into hydrologic response units (HRUs) which are lumped land areas with unique land use, soil type and slope combinations (Noori & Kalin, 2016).

The climatic variables required by SWAT include precipitation, temperature. Depending on the potential evapotranspiration calculation method used, wind speed, solar radiation and relative humidity may be required too(Senent-Aparicio et al., 2019).

The SWAT model plays a vital role in estimating the daily flow of a river basin. In the study conducted by Zhou et al., the SWAT model was calibrated and validated after inputting data such as precipitation, temperature, and wind speed in the basin, and the output data such as runoff, soil ersion, and water quality of the basin were obtained (Zhou et al., 2021)(Figure 4).

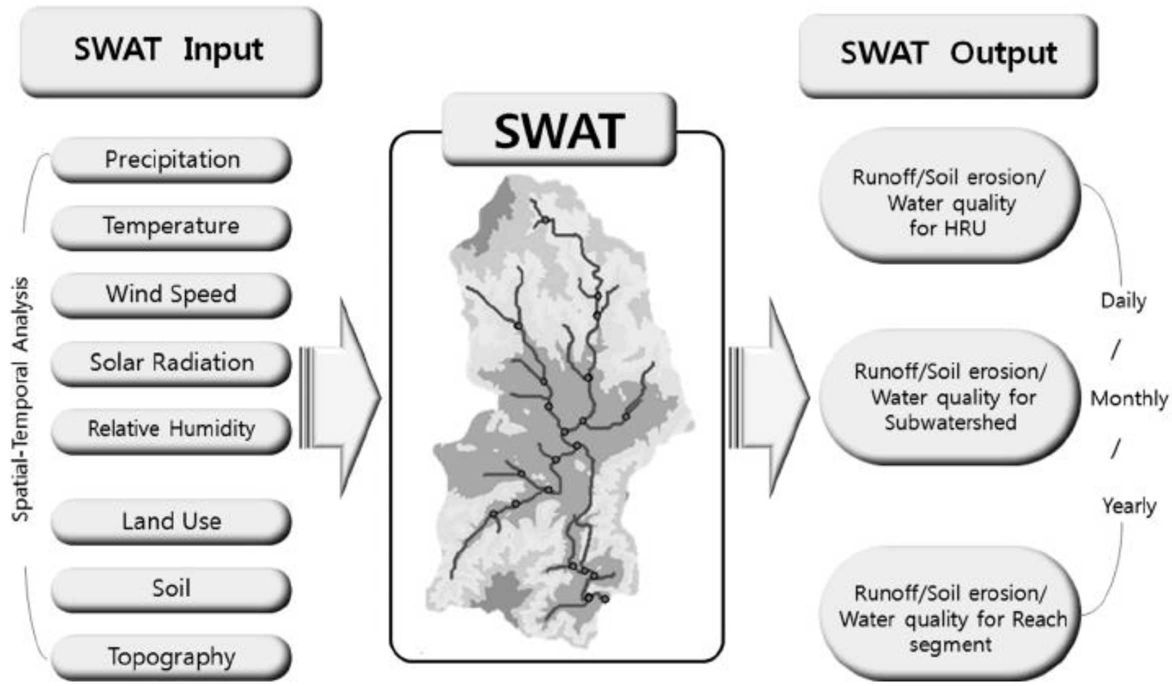


Figure 4: Overview of the swat model - model input/output parameters(Zhou et al., 2021)

3.2.2 SWAT model in Wanyao irrigation area

The SWAT model simulates agricultural water use in rice fields by defining the HRU where the rice fields in the irrigation area are located as depressions and presetting irrigation, water storage or water release operations. The agricultural water use simulation structure of the SWAT model is shown in Figure 5(Wang et al., 2020).

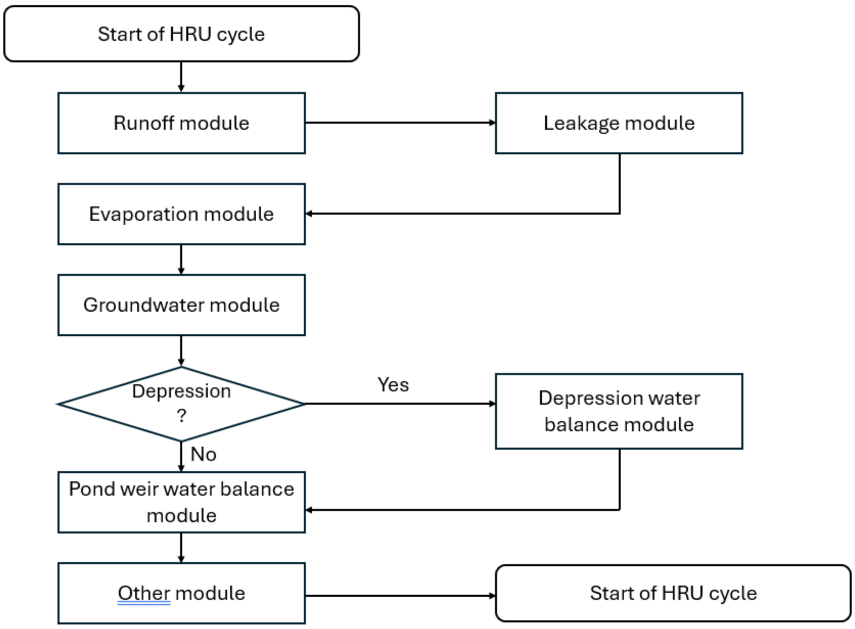


Figure 5: Structure diagram of irrigation water use in SWAT model

In the study of Wang et al., the SWAT model mainly includes the following aspects in the irrigation area research:

i. The SWAT model is used to simulate the water cycle process of the irrigation area, including precipitation, evaporation, runoff, groundwater recharge, etc., which helps to gain a deeper understanding of the recycling of water resources in the irrigation area.
ii. The SWAT model is used to count the irrigation water consumption of the irrigation area, including irrigation water sources, irrigation efficiency, irrigation operations, etc., providing important data support for water resource management and planning in the irrigation area.
iii. The SWAT model can simulate the changing laws of water volume for different purposes, including irrigation, power generation, urban water supply, ecological and environmental water use, etc., which helps to analyse the changing trends of various water demand.

In conclusion, the SWAT model's ability to reproduce the average daily flow in the Wanyao Irrigation District was evaluated, and it was found that the SWAT model provided acceptable performance indicators during the validation period, including a modified Nash-Sutcliffe efficiency of 0.64 and a good consistency index.

This shows that the SWAT model is suitable for estimating flow on a daily time scale in the Wanyao Irrigation District. Through the SWAT model, the water resource utilization of the irrigation district can be simulated and evaluated, providing a scientific basis for water resource management and planning in the irrigation district(Wang et al., 2020).

4 Research Questions and Aim

4.1 Gaps

4.1.1 Shortcomings of SWAT model in simulating water cycle in Wanyao irrigation area

i. When simulating the water cycle in the irrigation area, the SWAT model only considers the rice field as part of an independent HRU, and does not simulate it as an independent HRU, which makes the simulation process of agricultural water use in the rice field incomplete.

ii. The calculation method of the SWAT model for water balance elements (such as size, surface runoff, evapotranspiration, and soil water evaporation from the rice field) does not conform to the actual situation of the irrigation area.

iii. The SWAT model simulates rice field irrigation in the same way as dry land. Its automatic irrigation operation also uses the plant water stress threshold or soil water shortage threshold judgment method, which cannot accurately reflect the actual irrigation operation of rice fields. In addition, the SWAT model does not consider the canal water loss when calculating the irrigation water volume, which is inconsistent with the actual irrigation area.

iv. The SWAT model was originally developed for the long-term prediction of the impact of land management measures on the hydrology, sediment and agricultural chemical production in complex watersheds. Therefore, the model does not have simulation modules for non-agricultural water use processes such as urban water supply, hydropower generation, and ecological replenishment.

4.1.2 Shortcomings in constructing the SWAT model

SWAT is sensitive to the scale of application. The model performs poorly at very large scales due to limitations in representing spatial heterogeneity.

Due to uncertainty in input data, model parameters, and inherent model structure, SWAT predictions carry uncertainties that need to be quantified and communicated.

In addition, the structure of the SWAT model is complex, with many processes and interactions, making it difficult for users to fully understand and interpret model behavior.

Running SWAT over large watersheds or for long periods of time can be computationally intensive, requiring significant processing power and time. Learning and building a SWAT model often takes one to two months or even longer.

4.2 Research objectives

In this study, I will try to train and validate the input and output through the data-driven Transformer model to see if it can replace a very complex physical SWAT model.

4.2.1 Transformer model development process

This study aims to develop a Transformer model to replace SWAT for flow prediction. The daily data of Wanyao Irrigation District from 1986 to 2006 were selected as the research object. The input data of the Transformer model are exactly the same as those of the SWAT model, namely the temperature, precipitation, wind speed and relative humidity of the day. The output is the daily flow of the total drainage outlet of Wanyao Irrigation District.

The original data collection has been completed by the local hydrological station. After that, all daily data are divided into training data, validation data and test data in a ratio of 6:2:2. Since the training is carried out in batches, the process is iterated over multiple epochs to enable the model to learn from the entire dataset.

It is important to choose a suitable loss function for the research task. For continuous objectives such as flow prediction, mean square error (MSE) or mean absolute error (MAE) is usually used. Then, by observing the image of the Loss function, techniques such as dropout, early stopping and learning rate decay are implemented to prevent overfitting.

Finally, the NSE of the data training results and test results are printed out to observe the simulation effect of the Transformer model and compare it with the SWAT model.

4.2.2 Validation

Validation of Transformer models, especially when intended to replace complex hydrological models like SWAT, requires the use of a variety of validation indicators.

The following are commonly used indicators for hydrological models (besides the NSE mentioned above):

- Mean Squared Error (MSE): the average of the squares of the errors, which are the differences between the predicted values and the actual values. Lower MSE indicates better model performance, as it means that the predictions are closer to the actual values(Khan et al., 2023).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where:

n is the number of observations

y_i is the actual value for the i-th observation

$\hat{y}_i$ is the predicted value for the i-th observation

- Root mean square error (RMSE): the measure of the differences between values that are predicted by a model and values that are actually observed. RMSE ranges from 0 to ∞, where 0 indicates a perfect fit(Sharma et al., 2022).

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

Where:

n is the number of observations

y_i is the actual value for the i-th observation

$\hat{y}_i$ is the predicted value for the i-th observation

- Mean absolute error (MAE): the average absolute difference between the observed (actual) values and the values predicted by the model. A lower MAE indicates better model performance, as it signifies that the model's predictions are closer to the actual values(Sharma et al., 2022).

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

Where:

n is the number of observations

y_i is the actual value for the i-th observation

$\hat{y}_i$ is the predicted value for the i-th observation

- R² (coefficient of determination): the proportion of the variation in the dependent variable that is predictable from the independent variables.

$$\text{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Where:

$y_i$are the observed values.

$\hat{y}_i$ are the predicted values.

$\bar{y}$ is the mean of the observed values.

$\sum(y_i - \hat{y}_i)^2$ is the sum of the squared residuals (also known as the sum of squared errors or SSE).

$\sum(y_i - \bar{y})^2$ is the total sum of squares (TSS), representing the total variation in the observed data.

4.2.3 Advantages compared to the SWAT model

Because the Transformer model can learn complex patterns and relationships directly from large datasets, it is possible to capture nonlinear interactions that traditional models (such as SWAT) may miss.

And the Transformer has shown high accuracy in a variety of applications, and it continues to learn and adapt as new data emerges, which has the potential to improve its accuracy over time, which can more accurately predict water flow. Most importantly, the transformer model can efficiently process high-dimensional datasets and integrate various types of data (such as weather, soil properties, land use) without a lot of preprocessing.

The Transformer model automatically extracts relevant features from the input data, which may reduce the complexity and time of model setup. And as mentioned above, compared to SWAT, which takes a lot of time for professional researchers to build and calculate, the Transformer model can be highly parallelized and can efficiently process large-scale datasets and simulations using modern hardware (such as GPUs).

In addition, unlike SWAT, which requires a lot of calibration of a large number of parameters, machine learning models such as the Transformer model can learn end-to-end mapping from input data to output, thereby reducing the need for professionals to manually calibrate with the expertise of traditional hydrological models.

5 Model Analysis

5.1 Basic parameters

After the Transformer model framework (see Appendix 9.1) is built, the first step is to tune the hyperparameters. Tuning basic parameters (hyperparameters) for Transformer models is crucial for optimizing their performance, generalization, and efficiency. Proper hyperparameter tuning can significantly improve model accuracy, reduce training time, and ensure the model does not overfit or underfit the data(He et al., 2023).

Here are some hyperparameters of the Transformer model:

Table 1: Hyperparameters of the Transformer model

Basic parameters	Definition	Effect	Tuning methods
Sequence Length	Maximum length of input sequences	The speed and stability of convergence	Adjust based on the specific application and data characteristics. Longer sequences require more memory and computational power.
Batch Size	Number of training examples used in one iteration	Model stability and GPU memory usage	Start with a smaller batch size (e.g., 32) and increase it gradually.
Epochs	One complete pass through the entire training dataset	Through multiple epochs, the model can gradually adjust its parameters to minimize the loss and improve its performance.	Plot learning curves of training and validation loss over epochs to visualize the model's learning progress and identify the point where additional epochs do not yield significant improvements
Random State	The setting or seed used to initialize the random number generator (RNG) for any operation that relies on randomness	Facilitates tracking down issues by ensuring the same sequence of random events in every run	Choosing a number randomly, but 42 generally gives the best results(Islam et al., 2023).
Learning Rate	Controls the size of the steps the optimizer takes during training	The speed and stability of convergence	Test a range of values (e.g., 0.00001 to 0.001)
Number of Layers	Depth of the Transformer model, typically for both the encoder and decoder	Model capacity and ability to learn complex patterns	Experiment with a range of layers (e.g., 4 to 12)

Dropout Rate	Regularization technique to prevent overfitting	Regularizes the model to prevent overfitting	Adjust based on the level of overfitting observed during validation, common values are 0.1 to 0.5

The parameter tuning results are shown in the following table:

Table 2: Tuning results

Basic parameters	Results
Sequence Length	unknown
Batch Size	32
Epochs	unknown
Random State	42
Learning Rate	0.0001
Number of Layers	2
Dropout Rate	0.02

5.2 Processing of input data

As shown in the table 3 below, the training data selected by the model at the beginning is all daily meteorological data from January 1, 1986 to March 31, 2008 as input.

Table 3: Input (Daily meteorological data)

Date	Average Temperature (°)	Maximum temperature (°)	Minimum temperature (°)	Average wind speed (m/s)	Sunshine hours (h)	Relative humidity (%)	Precipitation (mm)
1986/1/1	3.9	9.4	1.5	1.5	7.1	79	0
1986/1/2	1.6	7.3	-1.2	3.8	6.3	83	0
1986/1/3	4.6	8.2	2.0	0.8	0.0	77	0
1986/1/4	3.9	10.7	-1.0	1.5	7.2	65	0
1986/1/5	-0.2	5.2	-4.4	1.3	9.3	67	0
1986/1/6	2.4	8.8	-2.0	2.3	8.6	55	0
1986/1/7	4.3	13.1	-2.4	3.0	9.1	57	0

1986/1/8	3.0	13.1	-3.6	1.0	8.8	72	0
1986/1/9	2.9	13.9	-1.7	0.5	8.6	74	0
1986/1/10	3.0	12.1	-4.0	2.5	9.0	72	0
1986/1/11	4.7	10.0	1.5	4.5	6.8	67	0
1986/1/12	6.2	16.1	-0.9	0.0	7.9	78	0
……							
2008/12/31	5.7	10.1	3.8	1.9	5.6	69	0

The output used to train the Transformer model is the actual daily flow at the irrigation area outlet. When the first training is performed according to the parameters determined in Section 5.1, the results are shown in the following table (Table 4).

Table 4: First training results

Results	Training period	Validation period
MSE	4539.4877	15877.2565
RMSE	67.3757	126.0050
MAE	31.6614	48.8029
NSE	0.1016	0.0652
R^2	0.2173	0.1550

Both NSE values (0.1 for training and 0.06 for validation) are quite low. These values indicate that the model is only marginally better than a naive prediction using the mean of the observed data. Therefore, the result is not considered good by typical standards in hydrological modelling.

The small difference between training and validation NSE values suggests that the Transformer model is not overfitting, but it also indicates that the model is not effectively capturing the underlying patterns in the data.

Potential reasons for the above problems are as follows:

1. The input to the Transformer model may be too complex.

2. The quality or quantity of the training data may be problematic.
3. The features used for modelling may not contain enough information.

In the following chapters, I will modify the above issues one by one to observe whether the simulation results are improved.

5.2.1 Reducing model input

Sunlight increases evaporation and plant transpiration, reducing soil moisture content and potentially reducing the amount of water available to flow to the surface or groundwater. Wind speed can increase evapotranspiration rates, slightly increasing water loss from soil and plant surfaces. High relative humidity reduces evaporation rates, while low relative humidity increases them. But the effects of all three factors are more pronounced over longer timescales and larger spatial scales, affecting regional climate patterns and long-term water availability rather than immediate, short-term outflows from local irrigation areas.

So, although sunshine hours, wind speed, and relative humidity are required input data for the SWAT model, they are not the main factors that directly affect the outflow of the system. The outflow of water from the Wanyao irrigation area is generally controlled by factors that are more directly related to the hydrological and hydraulic characteristics of the area. Therefore, in this data training, these three columns of data were removed to ensure that the model input is simple enough and reduce interference with the main input (such as precipitation).

After removing the three inputs of sunshine time, wind speed, and relative humidity, the training results of the second time are as follows:

Table 5: Second training results

Results	**Training period**	**Validation period**
MSE	132.0124	12306.0042
RMSE	11.4897	110.9324
MAE	5.7005	50.2522
NSE	0.9739	0.2754
R^2	0.9754	0.2841

The NSE during training is 0.9739 indicates that the model performs very well on the training data, while the NSE during validation is only 0.2754, which indicates that the model performs poorly on the validation data (as shown in Figure 6).

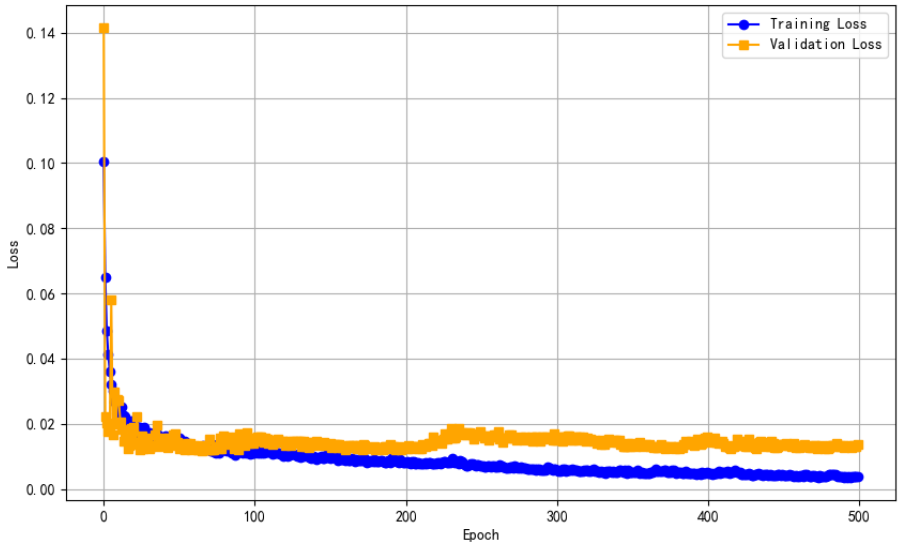


Figure 6: Training loss and validation loss of the second time

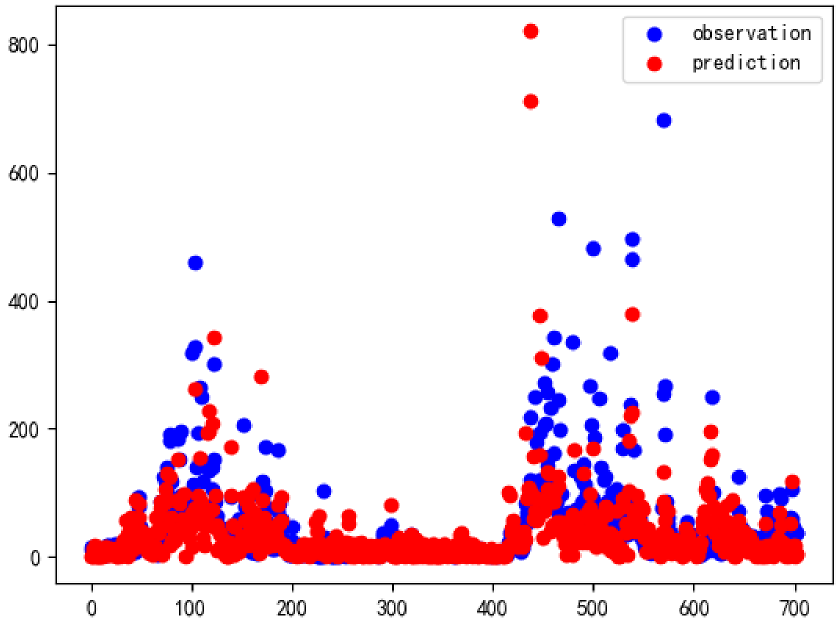


Figure 7: Scatter plot of the predictions and observations in the second time

The NSE dropped significantly from 0.97 during training to 0.21 during validation, indicating that the model does not generalize well (Figure 7). A good transformer model should perform well on both the training and validation datasets. This poor generalization indicates that the model may be too complex or that the training process still needs to be adjusted.

Therefore, the input data needs to be further simplified.

Evaporation is a key process in the hydrological cycle. There are several methods to calculate evaporation, ranging from empirical formulas to complex physical models. Among them, the ***Penman-Monteith equation*** is a widely used method for estimating evaporation, especially for reference evapotranspiration (ET_0). It combines both energy balance and aerodynamic concepts(McNaughton & Jarvis, 1984).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

Where:

ET_0 = Reference evapotranspiration (mm/day)

R_n = Net radiation at the crop surface (MJ/m²/day)

G = Soil heat flux density (MJ/m²/day)

T = Mean daily air temperature (°C)

u_2 = Wind speed at 2 meters height (m/s)

e_s = Saturation vapor pressure (kPa)

e_a = Actual vapor pressure (kPa)

$e_s - e_a$ = Saturation vapor pressure deficit (kPa)

Δ = Slope of the vapor pressure curve (kPa/°C)

γ = Psychrometric constant (kPa/°C)

According to the ***Penman-Monteith equation***, it is inferred that the daily maximum and minimum temperatures are of little significance in the current model training. Therefore, only the daily average temperature and daily precipitation are retained as input for the third training.

Table 6: Third training results

Results	Training period	Validation period
MSE	126.1746	11807.9845

RMSE	11.2886	108.6646
MAE	5.3847	45.4368
NSE	0.9848	0.3048
R^2	0.9836	0.3445

The training results are obviously better than the second time. However, the problem still exists that the NSE in the training period is much higher than that in the validation period (as shown in Figure 8).

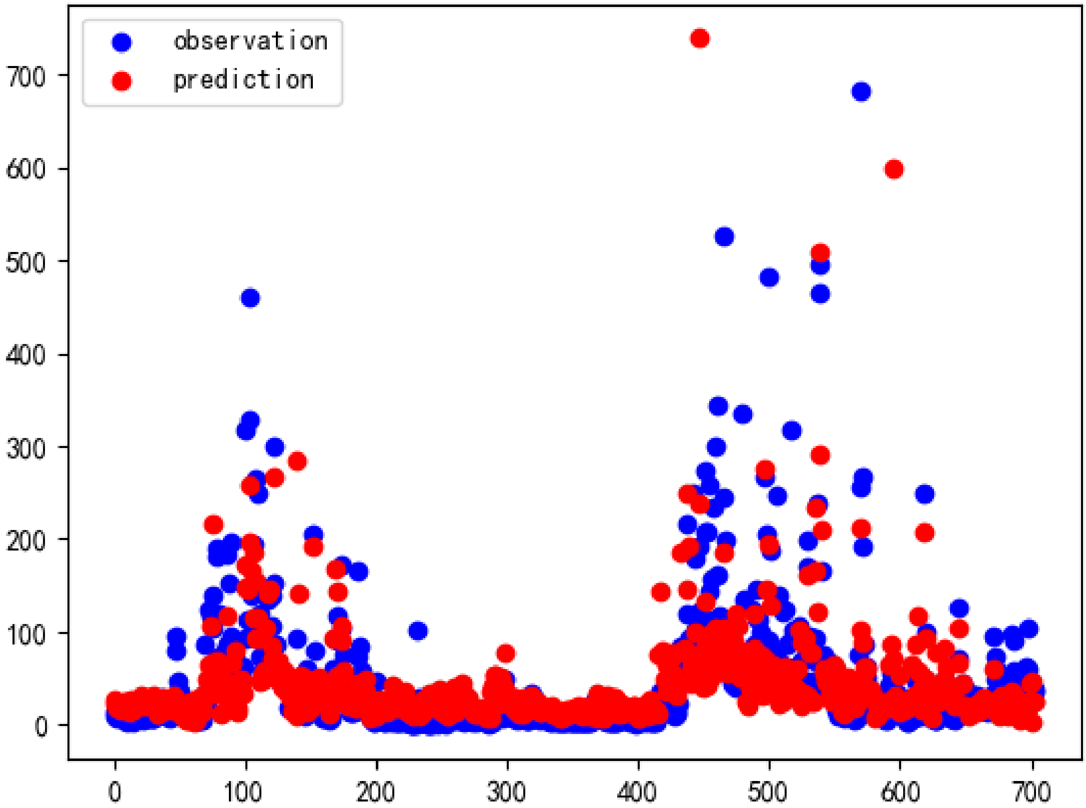


Figure 8: Scatter plot of the predictions and observations in the third time

As shown in Figure 8, a new problem has been discovered. When 500 is used as the number of epochs, the model shows signs of overfitting: when the training loss continues to decrease, it reaches a lower value compared to the validation loss, which indicates that there may be overfitting. Therefore, it shows that the model performs exceptionally well on the training data, but does not generalize well on the validation data. In addition, it can be seen from Figure 8 that the validation loss stabilizes after about 50 epochs, indicating that the performance of the model on unseen data will not improve significantly with additional training. This can be used as a heuristic method to decide to stop early to prevent overfitting and reduce training time.

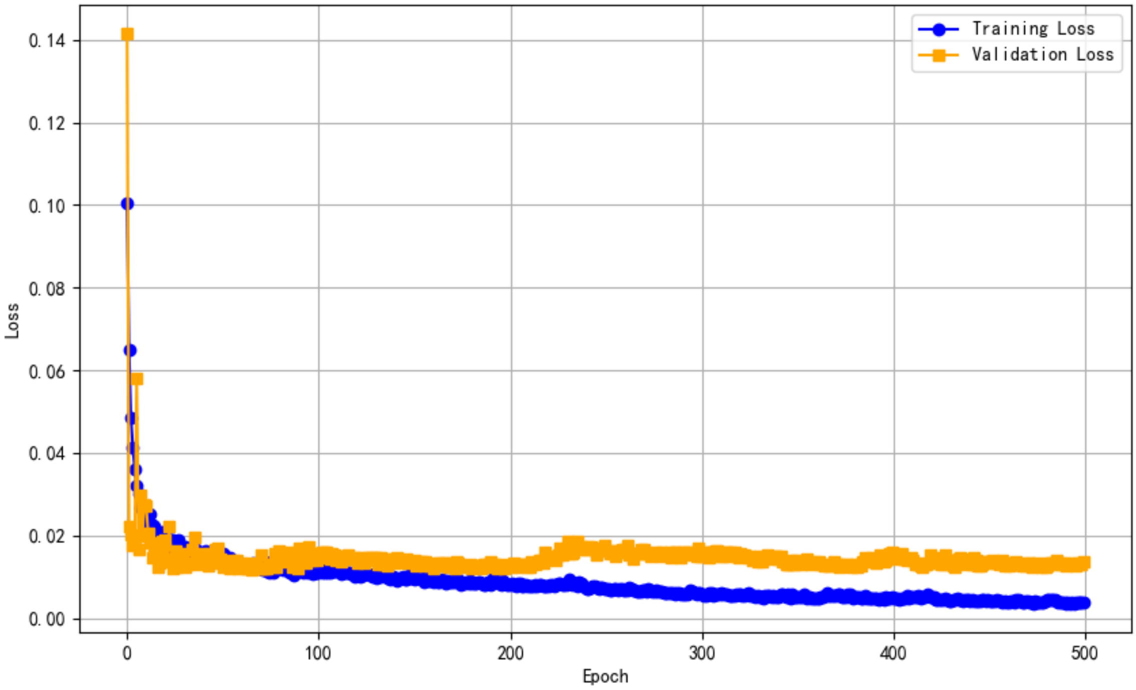


Figure 9: Training loss and validation loss of the third time

5.2.2 Early stopping

To address potential overfitting issues and improve model performance, early stopping is implemented to stop training when the validation loss stops improving. This prevents the model from overfitting the training data. An appropriate patience parameter is chosen to allow for slight fluctuations in validation loss before stopping.

The code of early stopping is shown in Appendix 9.3.

5.2.3 Precipitation delay

In hydrological modelling, especially when predicting river flow or streamflow, it is essential to account for precipitation delay because the relationship between precipitation and flow is not instantaneous. Several hydrological processes occur between the time when precipitation falls on the catchment and when it contributes to river or streamflow. Ignoring these delays can result in inaccurate predictions(Smith, 2003).

It takes time for precipitation to travel from the point where it falls to a river or stream. This delay can vary depending on distance, topography, and surface features of the catchment area. Some precipitation penetrates the soil and percolates down to replenish groundwater. It takes time for this water to contribute to runoff through baseflow(Beven & Hornberger, 1982).

In order to understand the residence time between precipitation and runoff, a line graph needed to be drawn for specific observation. Since the amount of data is too large, several years of flood season were selected as typical ones for mapping.

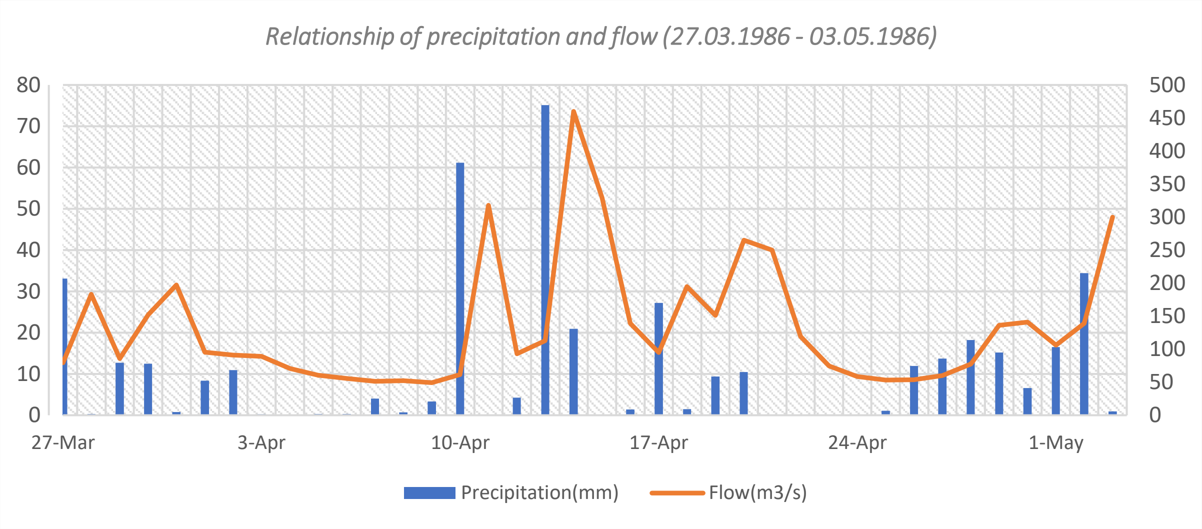


Figure 10: Relationship of precipitation and flow between 27-Mar-1986 to 03-May-1986

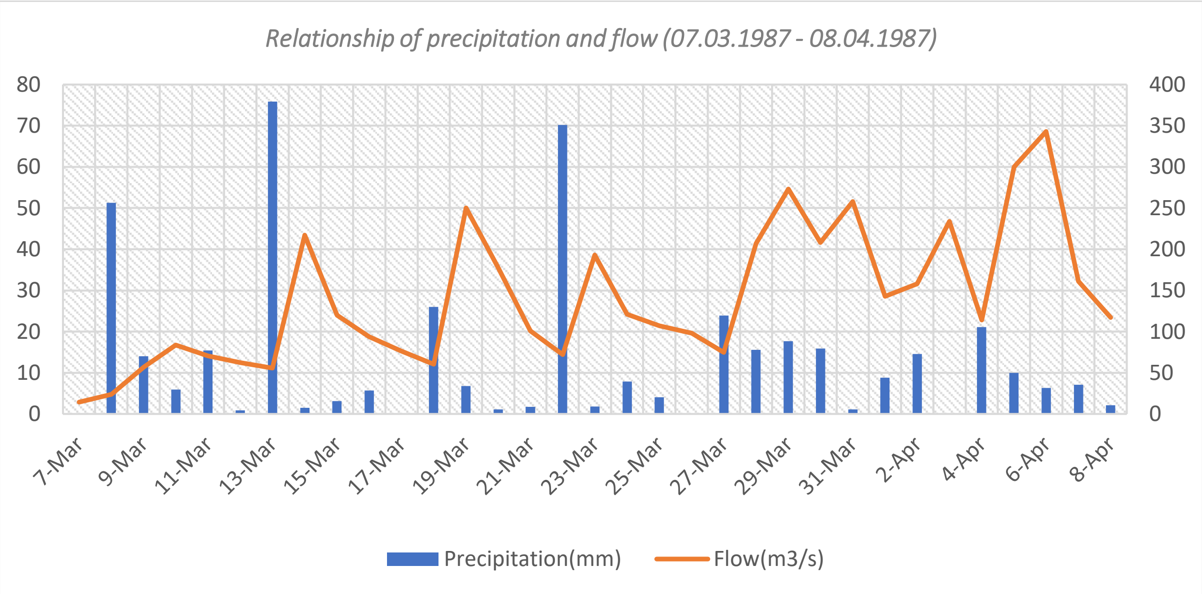


Figure 11: Relationship of precipitation and flow between 07-Mar-1987 to 08-Apr-1987

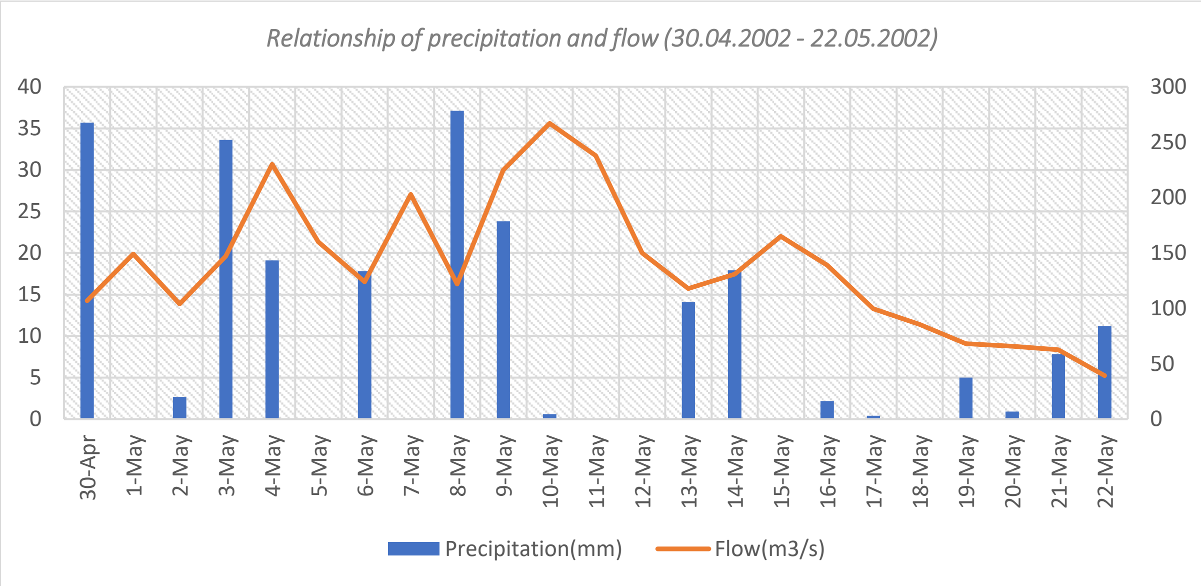


Figure 12: Relationship of precipitation and flow between 30-Apr-2002 to 22-May-2002

From Figures 9, 10 and 11 above, it can be found that the delay time between precipitation and flow is about one day. If the delay between precipitation and flow is approximately one day, setting the sequence length to 2 means that each input sequence to the model will include data from two consecutive days.

In addition, the daily precipitation in the output was changed to the accumulated precipitation over two days. After making the above changes, the fourth training results are as follows (Table 7):

Table 7: Fourth training results

Results	Training period	Validation period
MSE	703.4587	1276.4663
RMSE	26.5228	46.7356
MAE	17.0434	38.2272
NSE	0.8753	0.6538
R^2	0.8997	0.6947

The simulation results of the model are shown in the scatter plot below (Figure 13 and Figure 14):

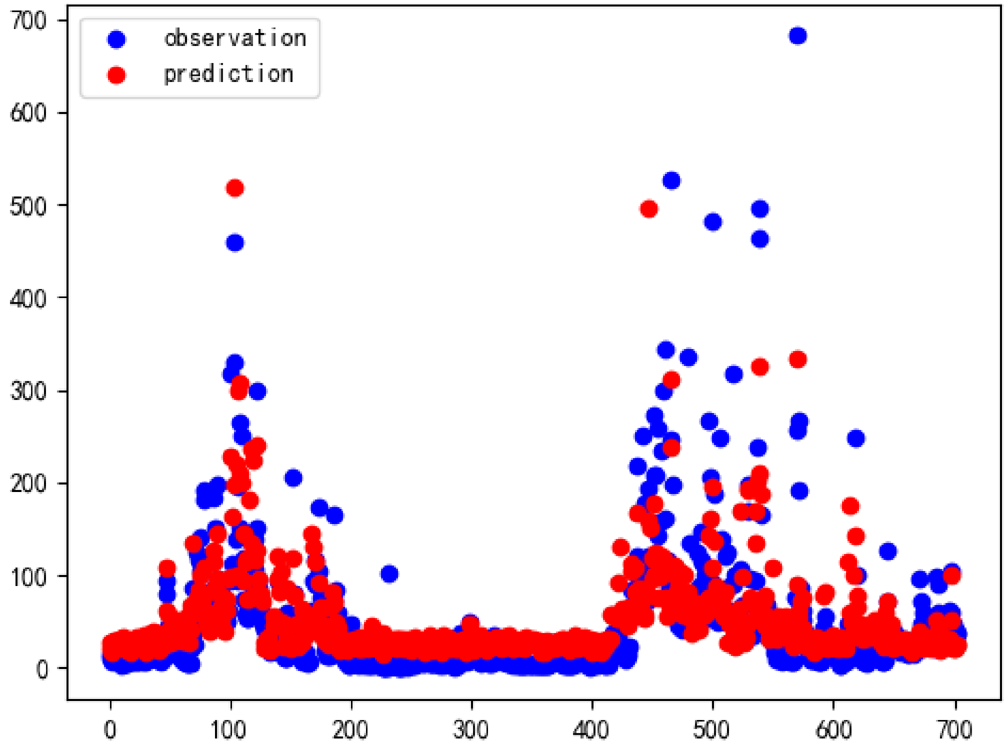


Figure 13: Scatter plot of the predictions and observations in the fourth time

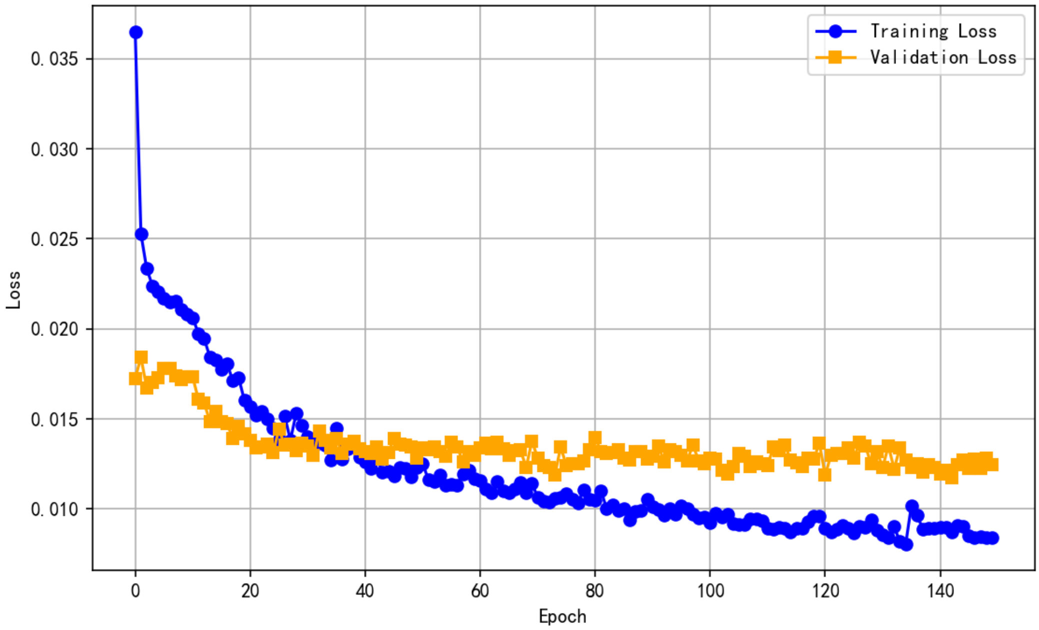


Figure 14: Training loss and validation loss of the fourth time

6 Discussions

6.1 Transformer model evaluation

When evaluating the performance of a Transformer model (or any hydrological model), the NSE value can be interpreted as follows:

- When $NSE > 0.75$, the model is considered to be performing well. The model predictions are very close to the observed data.
- When $0.65 < NSE \leq 0.75$, the model is considered to be performing well. The model predictions are reasonably accurate and can be used for many applications.
- When $0.50 < NSE \leq 0.65$, the model is performing satisfactorily. The model predictions are acceptable but may not be adequate for all purposes.
- If $0.00 < NSE \leq 0.50$, the model is not performing satisfactorily. The model predictions are better than random or average predictions, but there is still much room for improvement.

So for practical applications, especially in hydrology, NSE above 0.65 is generally considered a good indicator of good model performance. However, the specific threshold of "acceptable" performance may vary depending on the specific requirements of the application, the complexity of the watershed, and the quality of data available.

As mentioned in 2.2 above, since the input of the model is daily data (the time scale is very small), the model performance is considered good when the NSE during the validation period is between 0.5 and 0.7(Noori & Kalin, 2016).

In general, when the SWAT model is used to predict runoff flow using the same daily data as input, the NSE ranges from 0.474 to 0.898, with an average of 0.685(Chen et al., 2023). This study shows that the final validation period NSE obtained by using the Transformer model as a hydrological model is about 0.65.

6.2 Advantages of using Transformer model

Although the simulation results of the Transformer are slightly inferior to those of the SWAT model, the following reasons prove that it is still feasible to use the Transformer model for hydrological modelling.

First, the Transformer model is simpler to build than the SWAT model, taking only about one-tenth of its time. And the Transformer has parallel processing capabilities that can efficiently process large data sets. After training, the Transformer model can predict results faster than traditional hydrological models such as SWAT that need to solve complex physical equations.

Second, SWAT requires detailed physical parameterization of the basin, which may require a lot of labour and rely on high-quality field data. On the other hand, Transformer relies more on data-driven methods, reducing the need for a large number of parameter calibrations.

Another point worth considering is that Transformers are good at capturing complex nonlinear relationships in the data, which is very useful in hydrological modeling because such relationships are common. And Transformers are good at capturing time dependencies in sequential data, so they are suitable for modelling the temporal dynamics of rainfall-runoff processes.

It is also worth mentioning that the precipitation data provided in this study is not very accurate due to its age, which also leads to the poor effect of the Transformer model after data training. Combining the above points, in the future, if accurate historical data on precipitation, temperature, land use, and runoff for a watershed are known. It is still feasible to implement the Transformer model as a hydrological model. Because after training, the Transformer can quickly predict runoff, which is critical for real-time flood forecasting and water resource management. And the Transformer can easily integrate new data types, such as satellite imagery or real-time sensor data, to improve prediction accuracy. As new data becomes available, the model can be updated frequently to ensure accuracy under changing climate conditions.

6.3 Future researches

Future researches could focus on developing the Transformer model and exploring its advantages compared to traditional models like SWAT, paving the way for more accurate and reliable water cycle simulations in similar irrigation areas.

Many studies that have been conducted and are in progress also prove that the application of Transformer models in hydrology and water cycle has unlimited potential. For example, Transformers can improve the accuracy of hydrological models by effectively capturing complex, nonlinear relationships in data. This is particularly useful in the following areas:

- Flood forecasting: providing more accurate and timely forecasts to improve disaster preparedness and response capabilities, e.g., predicting the water level of a river one day ahead, by using the past water levels of its upstream branches as predictors (Castangia et al., 2023).
- Streamflow forecasting: improving forecast accuracy for better management of water resources.
- Rainfall-runoff modelling: improving understanding of how rainfall is converted to runoff, which is critical for managing watersheds and designing hydraulic structures, e.g., proposing a novel rainfall-runoff model named RR-Former based on the Transformer, which is entirely composed of attention mechanisms (Yin et al., 2022).

Transformers can also help assess and predict the impacts of climate change on water resources in the following ways:

- Extreme event modelling: predicting the frequency and intensity of extreme weather events such as droughts and floods, e.g., developing a Transformer-based model to forecast urban river discharges and predict flood peaks in the hydrological study(Li et al., 2024).
- Scenario analysis: evaluating different climate scenarios to guide policy and infrastructure planning.
- Trend analysis: detecting long-term trends in hydrological data to understand the impact of climate change on water availability and quality , e.g. using the transformer deep learning model to forecast hydrological drought, with a benchmark comparison with the long short-term memory (LSTM) model(Amanambu et al., 2022).

Transformer models have a promising future in hydrology and water resources, with the potential to revolutionize the prediction, management and understanding of water-related systems. By addressing current challenges and promoting interdisciplinary collaboration, these models can make a significant contribution to sustainable water resource management and resilience to climate change.

7 Conclusion

The study successfully evaluated the performance of water cycle modelling in the irrigation area using the Transformer model. Through the analysis of basic parameters and input data processing techniques, valuable insights were gained into the water resource management in the Wanyao Irrigation Area. The research findings indicate the potential of the Transformer model to replace the complex physical SWAT model, offering a more efficient and data-driven approach to water cycle simulation.

The findings of this study underscore the importance of integrating advanced data-driven techniques in water resource management, particularly in complex irrigation systems. The results indicate that the transformer model not only enhances predictive capabilities but also reduces the labour-intensive processes typically associated with model calibration.

Furthermore, this research contributes to the broader field of hydrology by providing a framework for future studies to build upon. The insights gained from the application of the transformer model pave the way for further exploration into its potential applications across different hydrological contexts.

Overall, this thesis not only highlights the transformative potential of data-driven modelling in hydrology but also emphasizes the need for continued innovation in the field to address the pressing challenges of water resource management in an increasingly complex and variable climate.

8 References

Amanambu, A. C., Mossa, J., & Chen, Y.-H. (2022). Hydrological Drought Forecasting Using a Deep Transformer Model. *Water*, *14*(22), Article 22. https://doi.org/10.3390/w14223611

Beven, K. J., & Hornberger, G. M. (1982). Assessing the Effect of Spatial Pattern of Precipitation in Modeling Stream Flow Hydrographs. *JAWRA Journal of the American Water Resources Association*, *18*(5), 823–829. https://doi.org/10.1111/j.1752-1688.1982.tb00078.x

Castangia, M., Grajales, L. M. M., Aliberti, A., Rossi, C., Macii, A., Macii, E., & Patti, E. (2023). Transformer neural networks for interpretable flood forecasting. *Environmental Modelling & Software*, *160*, 105581. https://doi.org/10.1016/j.envsoft.2022.105581

Chen, S., Huang, J., & Huang, J.-C. (2023). Improving daily streamflow simulations for data-scarce watersheds using the coupled SWAT-LSTM approach. *Journal of Hydrology*, *622*, 129734. https://doi.org/10.1016/j.jhydrol.2023.129734

Gassman, P. W., Sadeghi, A. M., & Srinivasan, R. (2014). Applications of the SWAT Model Special Section: Overview and Insights. *Journal of Environmental Quality*, *43*(1), 1–8. https://doi.org/10.2134/jeq2013.11.0466

He, X., Li, C., Zhang, P., Yang, J., & Wang, X. E. (2023). Parameter-Efficient Model Adaptation for Vision Transformers. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(1), Article 1. https://doi.org/10.1609/aaai.v37i1.25160

Islam, M. M., Hasan, M., Athrey, K. S., Braskich, T., & Bertasius, G. (2023). *Efficient Movie Scene Detection Using State-Space Transformers*. 18749–18758. https://openaccess.thecvf.com/content/CVPR2023/html/Islam_Efficient_Movie_Scene_Detection_Using_State-Space_Transformers_CVPR_2023_paper.html

Khan, N. A., Qureshi, M. A., Akber, S., & Hameed, T. (2023). Chapter 8 - Accruement of nonlinear dynamical system and its dynamics: Electronics and cryptographic engineering. In D.

Baleanu, V. E. Balas, & P. Agarwal (Eds.), *Fractional Order Systems and Applications in Engineering* (pp. 133–156). Academic Press. https://doi.org/10.1016/B978-0-32-390953-2.00015-3

Li, X., Sun, Q.-L., Zhang, Y., Sha, J., & Zhang, M. (2024). Enhancing hydrological extremes prediction accuracy: Integrating diverse loss functions in Transformer models. *Environmental Modelling & Software*, *177*, 106042. https://doi.org/10.1016/j.envsoft.2024.106042

Ma, D., Wang, T., Gao, C., Pan, S., Sun, Z., & Xu, Y.-P. (2018). Potential evapotranspiration changes in Lancang River Basin and Yarlung Zangbo River Basin, southwest China. *Hydrological Sciences Journal*, *63*(11), 1653–1668. https://doi.org/10.1080/02626667.2018.1524147

McNaughton, K. G., & Jarvis, P. G. (1984). Using the Penman-Monteith equation predictively. *Agricultural Water Management*, *8*(1), 263–278. https://doi.org/10.1016/0378-3774(84)90057-X

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Noori, N., & Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, *533*, 141–151. https://doi.org/10.1016/j.jhydrol.2015.11.050

Nourani, V., Komasi, M., & Mano, A. (2009). A Multivariate ANN-Wavelet Approach for Rainfall–Runoff Modeling. *Water Resources Management*, *23*(14), 2877–2894. https://doi.org/10.1007/s11269-009-9414-5

Senent-Aparicio, J., Jimeno-Sáez, P., Bueno-Crespo, A., Pérez-Sánchez, J., & Pulido-Velázquez, D. (2019). Coupling machine-learning techniques with SWAT model for instantaneous peak flow prediction. *Biosystems Engineering*, *177*, 67–77. https://doi.org/10.1016/j.biosystemseng.2018.04.022

Serrano, W. (2017). Smart Internet Search with Random Neural Networks. *European Review*, *25*, 1–13. https://doi.org/10.1017/S1062798716000594

Sharma, D. K., Chatterjee, M., Kaur, G., & Vavilala, S. (2022). 3—Deep learning applications for disease diagnosis. In D. Gupta, U. Kose, A. Khanna, & V. E. Balas (Eds.), *Deep Learning for Medical Applications with Unique Data* (pp. 31–51). Academic Press. https://doi.org/10.1016/B978-0-12-824145-5.00005-8

Smith, R. B. (2003). A linear upslope-time-delay model for orographic precipitation. *Journal of Hydrology*, *282*(1), 2–9. https://doi.org/10.1016/S0022-1694(03)00248-8

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Wang, H., Wang, S., Wen, J., Wang, S., & Zhu, M. (2020). Monitoring and statistical methods of irrigation water consumption of Wanyao Irrigation Area. *IOP Conference Series: Earth and Environmental Science*, *510*(3), 032025. https://doi.org/10.1088/1755-1315/510/3/032025

Yin, H., Guo, Z., Zhang, X., Chen, J., & Zhang, Y. (2022). RR-Former: Rainfall-runoff modeling based on Transformer. *Journal of Hydrology*, *609*, 127781. https://doi.org/10.1016/j.jhydrol.2022.127781

Zhou, M., Deng, J., Lin, Y., Zhang, L., He, S., & Yang, W. (2021). Evaluating combined effects of socio-economic development and ecological conservation policies on sediment retention service in the Qiantang River Basin, China. *Journal of Cleaner Production*, *286*, 124961. https://doi.org/10.1016/j.jclepro.2020.124961

9 Appendix

Due to the length of the thesis, most of the code and the processed data used for training mentioned before are saved in the form of text or tables in Google Drive. The relevant links are attached below. If interested, clicking to view the details.

9.1 Transformer Architecture

https://docs.google.com/document/d/1eMM4o287v3wxz7fpYIKwFSvau83C3t-kBu4Mmg9LwaI/edit?usp=drive_link

9.2 Data

- First training data:

  https://docs.google.com/spreadsheets/d/1RUPiP8norqnnhY3VRrBJOILxcW1bYO9NnqrFvoaC_7Y/edit?usp=drive_link

- Second training data:

  https://docs.google.com/spreadsheets/d/1a3Q9xJDROfY36Vt1ZAZX1sTK0hax3WNAXpZtjxMX4qI/edit?usp=drive_link

- Third training data:

  https://docs.google.com/spreadsheets/d/1KD9Zu74tOmRSj1L3C3YMv7MQsLfnNkw0KM4OUMG8Lu8/edit?usp=drive_link

- Fourth training data:

  https://docs.google.com/spreadsheets/d/1aodYI5WC1O2wmW3D32HGXwuYUo_fOYSoo38e2DRTUfc/edit?usp=drive_link

9.3 Early stopping

```
from tensorflow.keras.callbacks import EarlyStopping

# Define early stopping callback
early_stopping = EarlyStopping(monitor='val_loss', patience=10,
restore_best_weights=True)

# Fit the model with early stopping
history = model.fit(X_train, y_train,
                    epochs=500,
                    batch_size=32,
                    validation_data=(X_val, y_val),
                    callbacks=[early_stopping])

# Plot training and validation loss
```

```python
import matplotlib.pyplot as plt

plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.title('Training and Validation Loss')
plt.show()
```