

# Research and application of deep learning techniques for forecasting groundwater levels in the Hanoi area

Ta Quang Chieu<sup>1\*</sup>, Nguyen Đac Phuong Thao<sup>1</sup>, Ngo Quang Minh<sup>1</sup>, Hoang Van Hung<sup>2</sup>

**Abstract:** Groundwater is essential for keeping the ecosystem in balance and meeting the water needs of homes, farms, and businesses, especially in big cities like Hanoi, which are strongly affected by climate change and overuse. In the context of climate change and overexploitation, accurate groundwater level forecasts not only help successful water resource management, but also serve as a foundation for long-term development. However, the intricacy, nonlinearity, and long-term dependency of groundwater levels present significant problems for standard approaches. Ours provides a novel approach for predicting groundwater levels at monitoring station Q64 in the Hanoi area using four deep learning models: RNN, LSTM, Transformer and Autoformer. We forecast short-term (48 hours), medium-term (120 hours) and long-term (360 hours). Experimental results reveal that Autoformer is clearly superior in forecasting situations and performs well in short-term, in the medium term forecasting. This demonstrates that the model based on an attention-based architecture can capture long-term properties of groundwater level time series. These findings support the use of deep learning in groundwater level forecasting, paving the way for the creation of intelligent forecasting systems, aiding decision-making in water resource management, and developing ways to adapt to climate change in large cities.

**Keywords:** Groundwater level forecasting, deep learning, transformer, autoformer.

## 1. Introduction

Groundwater, which is water that collects underground, is an important part of the hydrological cycle. Groundwater stays in the ground longer than surface water does, and it isn't as impacted by short-term changes in the weather. However, it is quite sensitive to long-term changes in the climate and human activities. The levels of groundwater can be affected by a number of things, such as the terrain, the amount of rain, the temperature, the permeability of the soil, the geology of the aquifer, the levels of surface water, the amount of groundwater abstraction, and human activities like farming and urbanization (Pawari & Gawande, 2015). Groundwater, which makes up over 97% of the world's freshwater that people can use (Zhao et al., 2024) is vital for both biodiversity and food and water security (Saccò et al., 2024). As freshwater shortages become more common, good groundwater management is important for bringing this resource back to life, making it more valuable, controlling extreme weather events, and increasing the amount of freshwater available (Vaux, 2011). MODFLOW (Pathak et al., 2018) and HBV-light are two examples of physical models that are often used to predict groundwater levels. However, they often need a lot of input data, are expensive to run, and have trouble simulating complicated nonlinear relationships, especially when the data is missing or has clear

periodicities (Seibert & Bergström, 2022). As a result, it is critical to provide credible forecasting tools to assist management agencies in assessing and monitoring groundwater level change patterns, allowing them to develop effective reaction strategies, optimize resource exploitation and allocation, and reduce risks and economic losses. Water supplies are one area where artificial intelligence is being used more and more. Machine learning systems have made a lot of progress in predicting groundwater levels, which helps make planning and managing water resources more efficient. There are two reliable ways to guess the amount of groundwater: support vector regression (SVR) (Mukherjee & Ramachandran, 2018) and random forest (RF) (Adnan et al., 2023). Furthermore, using the collected groundwater level dataset as an irregular time series, Mukherjee et al. used three models: linear regression model (LRM), support vector model (SVR), and artificial neural network (ANN) to overcome the limitations of traditional statistical methods (ARMA, ARIMA, etc.) and achieve positive results (Alsumaiei, 2020). Machine learning algorithms perform well with simple time series forecasting tasks, but their ability to extract complicated nonlinear or time-dependent correlations is restricted. So, deep learning models are becoming increasingly popular in hydrological research due to their capacity as powerful alternatives due to their capacity to automatically extract hierarchical temporal features, model complex nonlinear interactions, and learn long-range dependencies within hydrological time series and increase prediction accuracy. As a result, approaches to deep learning architectures for groundwater level forecasting problems are becoming a significant research topic. Since the 2000s, the study

<sup>1</sup>Department of Artificial Intelligence, Thuyloi University

<sup>2</sup>Center for Water Resources Technology and Data

\* Corresponding author

Received 30<sup>th</sup> Oct. 2025

Accepted 28<sup>th</sup> Nov. 2025

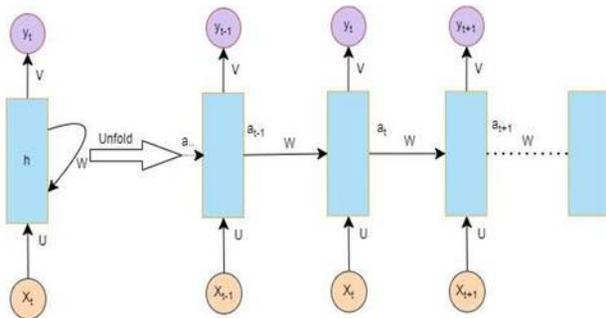
Publication date 31<sup>st</sup> Dec. 2025

(Gers et al., 2000) that used Transformer and LSTM to guess shallow groundwater found that 5-day predictions were correlated with a  $R^2$  value of 0.523 . In the research (Liu et al., 2019) about Autoformer, an advanced deep learning architecture, achieved outstanding success in predicting the flow of the Yellow River. Its  $R^2$  coefficient was as high as 0.89, which indicates that it can learn long-term time series data very well. In 2021, Deo & Lee (Seo & Lee, 2021) using multi-satellite data and combining GRACE satellite data with the CNN-LSTM model correctly forecast changes in groundwater reserves. In 2023, Chen et al. (Chen et al., 2023) demonstrated that a combination of Generative Adversarial Networks (GANs), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNNs) improved the accuracy of groundwater level forecasts. In 2024, the experiment of Ali et al. (Ali et al., 2024) to help stop flooding used the LSTM and RNN models to predict groundwater levels, with a metric RMSE of only 0.03 m. Based on the existing research, our research team suggested and implemented RNN, LSTM, Transformer, and Autoformer models for predicting groundwater levels in the Hanoi area. Deep learning algorithms can capture complicated relationships in time series and automatically extract important features, thus improving groundwater level forecasting accuracy in the short, medium, and long term.

**2. Theoretical foundation**

**2.1. Recurrent Neural Network (RNN)**

RNN (Gers et al., 2000) is a type of neural network that works with data that comes in a certain order. It has a feedback system that lets it store and process information from earlier steps in the sequence. This



**Figure 1.** RNN network architecture

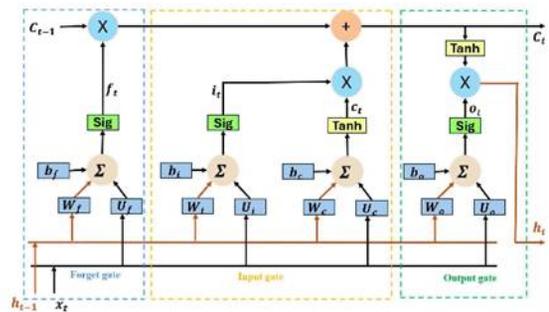
**2.3. Transformer**

Transformer (Xiong et al., 2023) is a powerful deep learning architecture designed for a variety of NLP and voice challenges, including machine translation, text generation, categorization, and speech recognition. Unlike RNN, Transformer evaluates the full input sequence in parallel using a self-attention mechanism, allowing it to capture the relationship between any items in the sequence while

helps the model understand the context or relationship between the elements in the sequence . Because of this, the model can handle tasks like analyzing text, recognizing audio, and predicting time series. At each time step, RNN obtains an input and modifies the hidden state by combining the new information with the old state. This hidden state affects the output at each step, which helps RNN learn and show how the parts in the sequence are connected. We use the back propagation algorithm over time (BPTT) to figure out the mistake and change the network's weights during the training process.

**2.2. Long Short-Term Memory (LSTM)**

LSTM (Liu et al., 2019) is a better version of RNN that was made to fix the problems with classic RNN, especially its ability to remember, use, and manage dependencies in long-term datasets. LSTM is made to control the flow of information well by using gates and cell states. The memory cell is the central component, acting as a long-term memory, allowing information to be stored beyond time. The maintenance and removal of information in the memory cell are coordinated by three types of gates, each of which has a distinct regulatory function. The forget gate determines which information in the memory cell should be erased, allowing the model to discard unnecessary data. The input gate controls how new information gets into the memory cell, which lets the model learn new things with each time step. Lastly, the output gate tells the memory cell which part of its information will be used to make the output at this stage. This structure allows the LSTM to learn long-term dependencies in data sequences while remaining stable throughout training.



**Figure 2.** LSTM Network Architecture

also effectively utilizing GPU/TPU technology to reduce training time. The model has two main parts: the encoder and the decoder. The encoder is made up of many self-attention layers, and the decoder is made up of many other layers. Multi-Head Self-Attention is one of Transformer's most important features. It lets the model learn many connections between words in a statement at the same time.

The encoder slowly pulls out semantic features at

different levels, while the decoder uses hidden self-attention to create a series of outputs. The encoder and decoder's attention is then combined to combine the input data.

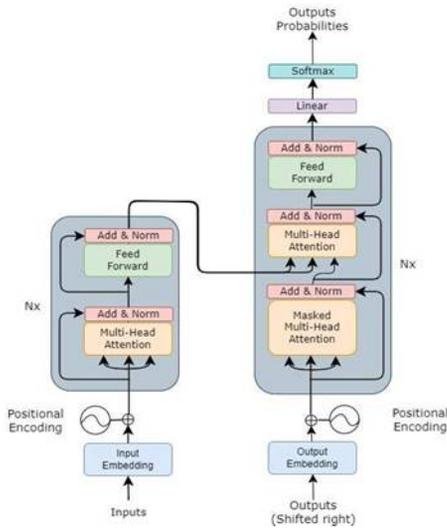


Figure 3. Transformer network architecture

#### 2.4. Autoformer

Autoformer (Wu et al., 2021) is an enhanced version of Transformer that was made for predicting long-term time series. The model includes two major components: the Series Decomposition Block, which decomposes the series into long-term trend and periodic components, reducing noise and improving the ability to learn periodic patterns; and the Auto-Correlation Mechanism, which replaces traditional dot-product attention with cross-correlation, detecting past segments that are strongly correlated with the present and exploiting the data's periodicity. Keeping the encoder-decoder architecture, Autoformer inserts breakdown blocks between layers and extracts features using auto-correlation, while the decoder generates forecasts based on the decomposed components. This significantly increases the model's capacity to detect long-term and periodic correlations in time series.

#### 2.5. Evaluation Metrics

In this section, the metrics used to evaluate the quality of the regression model are: Coefficient of determination ( $R^2$ ), Mean absolute error (MAE), Mean square error (RMSE). The mathematical formulas of the metrics are as follows:

$$R^2 = \left[ \frac{\sum_{i=1}^n (H_i^{tt} - H^{tt})(H_i^{dd} - H^{dd})}{\sqrt{\sum_{i=1}^n (H_i^{tt} - H^{tt})^2 \sum_{i=1}^n (H_i^{dd} - H^{dd})^2}} \right]^2 \quad (1)$$

$$MAE = \frac{\sum_{i=1}^n |H_i^{tt} - H_i^{dd}|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (H_i^{tt} - H_i^{dd})^2}{n}} \quad (3)$$

Where:

-  $H_i^{tt}$  and  $H_i^{dd}$  are the real and predicted amounts of water in the soil at time  $i$ .

-  $H^{tt}$  and  $H^{dd}$  are the average amounts of water in the soil at time  $i$ .

The  $R^2$  index typically ranges from 0 to 1, with a higher value indicating a closer fit between the simulation model and observed data. MAE measures the average absolute error between forecast and reality, whereas RMSE is more sensitive to big errors due to the usage of squares in the calculation. A model is deemed to have strong forecast quality when MAE and RMSE decline and  $R^2$  approaches one. Combining these three variables allows for a more comprehensive evaluation of the model, including accuracy and dependability, which aids in the selection of a forecasting approach appropriate for the research area's hydrological features. By combining these three variables, we can get a more complete understanding of the model's accuracy and dependability. This helps us choose a forecasting method that works best for the hydrological features of the research area.

### 3. Results and experiments

#### 3.1. Implementation Process

Figure 4 illustrates the development and evaluation of the groundwater level prediction. (1) Data collection: Water level data, which will be collected continuously every two hours from January 1 of 2014, to June 30 of 2024, at three monitoring stations: Q.64 in Trung Tu ward, Dong Da district; Q.69 in Phu Lam ward, Ha Dong district; and Q.66 in Ngu Hiep ward, Thanh Tri district. (2) Data preprocessing: Visualize the acquired data to discover important fields, then, unnecessary data will be deleted before the next steps. (3) Model training: We will use deep learning models, respectively, RNN, LSTM, Transformer, and Autoformer, to predict groundwater levels. (4) Evaluation and conclusion: The study will compare various models and parameter sets to determine which model produces the best forecast quality.

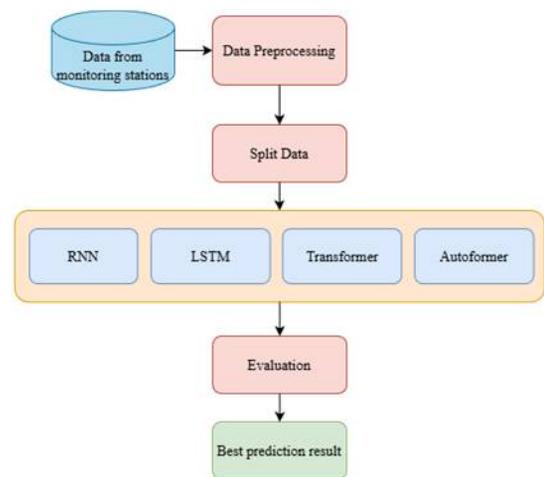


Figure 4. Problem processing flowchart

To test the hypothesis of the association between the monitoring stations, the study team ran a

correlation analysis on each pair of stations. We computed pairwise correlations among the stations; the Q.64–Q.66, Q.66–Q.69, and Q.64–Q.69 pairs had correlation coefficients of 0.17, 0.35, and 0.36, respectively, indicating weak associations. Based on

geological analysis (being in the same aquifer) and time series correlation, the team chose data from Q.64 to feed into the deep learning model for forecasting groundwater levels at station Q.64.

### 3.2. Training Hyperparameter Selection

**Table 1.** Training Hyperparameter selection for the models

Models	Learning rate	Epochs	Units	Batch size	Dropout
RNN					
LSTM					
Transformer	[0.001, 0.01]	[20,30]	[32, 64]	[32]	[0.2,0.3]
Autoformer					

Table 1 presents the hyperparameter search space used during model training. The listed ranges (learning rate, epochs, units, batch size, dropout) are applied consistently across all models to ensure a fair comparison.

To ensure objectivity, each hyperparameter combination was trained and evaluated using the cross-validation approach. Various configurations were tested on four models to determine the best option in terms of predictive performance. GridSearch was used to find the optimal parameter set for each model design. Table 1 shows the specific optimal hyperparameters used for the models.

### 3.3. Experimental Results

Experimental findings of deep learning models such as RNN, LSTM, Transformer, and Autoformer in groundwater level forecasting at three time points: short-term 48 hours, medium-term 120 hours, and long-term 360 hours. Model performance is assessed using the coefficient of determination  $R^2$ , MAE and RMSE indexes. Furthermore, the results are assessed for each time period and compared to practical requirements in the research area, which has a

consistent water supply demand and is influenced by urbanization and climate change. In this table 3 below, the top-row labels (48h, 120h, 360h) indicate input window sizes, whereas the 48h, 120h, and 360h values under each block denote the forecasting horizons.

*Evaluation with 48-Hour Input Sequences:* With 48-hour input sequences, Autoformer demonstrates the most consistent and superior performance across all forecasting horizons, achieving the highest  $R^2$  values (0.9872, 0.9738, 0.9341) and the lowest MAE and RMSE. This confirms its strong capacity to capture both short- and long-term dependencies from limited data. LSTM follows closely, performing reliably in short- to medium-term forecasts due to its effective handling of temporal dependencies. In contrast, RNN rapidly loses accuracy over longer horizons because of its restricted memory depth, while Transformer delivers moderate but less stable results, likely affected by shorter sequence inputs. Overall, Table 2 highlights the clear advantage of attention-based architectures, particularly Autoformer in achieving high forecasting accuracy even with short input lengths.

**Table 2.** Forecast results at 3 time points with 48h, 120h, 360h data input

Models	Metrics	48h			120h			360h		
		48 hours	120 hours	360 hours	48 hours	120 hours	360 hours	48 hours	120 hours	360 hours
RNN	R2	0.9539	0.8894	0.4084	0.9704	0.8527	0.3759	<b>0.9750</b>	0.8905	0.3119
	MAE	0.0848	0.1327	0.3473	0.0683	0.1587	0.3625	<b>0.0629</b>	0.1358	0.3760
	RMSE	0.1208	0.1874	0.4341	0.0968	0.2163	0.4459	<b>0.0890</b>	0.1863	0.4683
LSTM	R2	0.9706	0.9535	0.6966	0.9249	0.9213	0.6750	0.9313	<b>0.9603</b>	0.7462
	MAE	0.0728	0.0838	0.2455	0.1284	0.1170	0.2538	0.1261	<b>0.0832</b>	0.2258
	RMSE	0.0965	0.1215	0.3105	0.1543	0.1581	0.3217	0.1475	<b>0.1121</b>	0.2844
Transformer	R2	0.8909	0.9058	0.7829	0.9261	0.8982	0.8066	0.8456	0.8523	0.6351
	MAE	0.2139	0.1979	0.3188	0.1799	0.2056	0.3055	0.2598	0.2504	0.4168
	RMSE	0.2846	0.2660	0.4052	0.2353	0.2765	0.3824	0.3405	0.3331	0.5253
Autoformer	R2	<b>0.9872</b>	<b>0.9738</b>	<b>0.9341</b>	<b>0.9880</b>	<b>0.9697</b>	<b>0.9149</b>	0.9727	0.9473	<b>0.8801</b>
	MAE	<b>0.0697</b>	<b>0.0997</b>	<b>0.1653</b>	<b>0.0661</b>	<b>0.1100</b>	<b>0.1946</b>	0.1035	0.1442	<b>0.2281</b>
	RMSE	<b>0.0974</b>	<b>0.1400</b>	<b>0.2231</b>	<b>0.0943</b>	<b>0.1506</b>	<b>0.2535</b>	0.1428	0.1987	<b>0.3011</b>

*Evaluation with 120-Hour Input Sequences:* As the input sequence extends to 120 hours, all models

demonstrate improved utilization of temporal dependencies. Among them, Autoformer consistently

delivers the most accurate and stable predictions, achieving the highest  $R^2$  values (0.9880, 0.9697, 0.9149) and the lowest error metrics, thereby confirming its robustness across multiple forecasting horizons. LSTM remains a strong competitor for medium-term predictions ( $R^2 = 0.9213$ , MAE = 0.1170), reflecting its effectiveness in modeling moderate-range temporal dynamics. RNN exhibits marginal improvement in the short-term but deteriorates significantly for longer horizons, reinforcing its limitation in representing extended dependencies. Transformer produces relatively stable yet inferior outcomes, likely constrained by its sensitivity to longer input series and absence of explicit decomposition mechanisms. Overall, Table 4 demonstrates that longer input sequences enhance model stability and predictive reliability, with Autoformer and LSTM emerging as the most effective architectures for medium-term groundwater level forecasting.

*Forecast results over 360 hours:* When trained with a 360-hour input sequence, Autoformer exhibits clear dominance in long-term groundwater level forecasting, achieving the highest coefficient of determination ( $R^2 = 0.8801$ ) and the lowest error metrics (MAE = 0.2281, RMSE = 0.3011). These results substantiate its superior capacity to capture extended temporal dependencies and periodic dynamics inherent in hydrological time series. LSTM continues to deliver competitive performance at the medium-term horizon ( $R^2 = 0.9603$ ), reflecting its effectiveness in modeling intermediate temporal fluctuations. In contrast, RNN maintains acceptable short-term accuracy but experiences pronounced degradation over longer forecast horizons due to its limited contextual memory, while Transformer remains less competitive, hindered by its weaker ability to represent long-range dependencies. Collectively, the outcomes in Table 5 reinforce Autoformer's robustness and reliability as the most effective architecture for long-term groundwater forecasting, whereas LSTM and RNN are better suited for medium- and short-term prediction tasks, respectively.

#### 4. Conclusion and future work

This study presents a comprehensive investigation into the application of deep learning architectures RNN, LSTM, Transformer, and Autoformer for groundwater level forecasting in the Hanoi metropolitan area. Using continuous monitoring data from station Q64, the models were evaluated across short (48 h), medium (120 h), and long-term (360 h) forecasting horizons.

The experimental findings demonstrate that Autoformer consistently outperforms the other models, achieving the highest predictive accuracy and stability across all temporal scales, particularly in long-term forecasting scenarios ( $R^2 = 0.8801$ , MAE = 0.2281,

RMSE = 0.3011). This confirms the superiority of the attention-based and autocorrelation mechanisms in capturing extended temporal dependencies and periodic fluctuations inherent in hydrological time-series data. LSTM exhibits strong performance for medium-term predictions due to its ability to retain intermediate temporal features, while RNN remains suitable for short-term forecasting where local temporal continuity dominates.

Collectively, these results substantiate the potential of deep learning, especially attention-driven architectures, as powerful tools for groundwater level prediction. The outcomes contribute a methodological foundation for the development of intelligent forecasting systems to support evidence-based decision-making in urban water-resource management and climate-adaptation strategies.

Future research will focus on extending the modeling framework to a regional multi-station network encompassing the entire Hanoi aquifer system. Integrating exogenous variables such as rainfall, temperature, evapotranspiration, and anthropogenic extraction rates will be prioritized to enhance spatiotemporal generalization and forecasting precision. Additionally, coupling deep learning models with hydrological simulation frameworks and IoT-based real-time monitoring systems is envisioned to enable dynamic, adaptive management of groundwater resources under the pressures of climate variability and urban expansion.

#### Acknowledgements

The authors acknowledge the support of the Thuyloi University to the research team and the financial support of IRD (Across), which enabled the completion of this study.

#### References

- Adnan, R. M., Dai, H. L., Mostafa, R. R., Islam, A. R. M. T., Kisi, O., Heddad, S., & Zounemat-Kermani, M. (2023). *Modelling groundwater level fluctuations by ELM merged advanced metaheuristic algorithms using hydroclimatic data*. Geocarto International, 38(1). <https://doi.org/10.1080/10106049.2022.2158951>
- Alsumaiei, A. A. (2020). *A Nonlinear Autoregressive Modeling Approach for Forecasting Groundwater Level Fluctuation in Urban Aquifers*. Water 2020, Vol. 12, Page 820, 12(3), 820. <https://doi.org/10.3390/W12030820>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). *Learning to forget: Continual prediction with LSTM*. Neural Computation, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Liu, Y., Gong, C., Yang, L., & Chen, Y. (2019). *DSTP-RNN: a dual-stage two-phase attention-based*

- recurrent neural networks for long-term and multivariate time series prediction. *Expert Systems with Applications*, 143. <https://doi.org/10.1016/j.eswa.2019.113082>
- Mukherjee, A., & Ramachandran, P. (2018). *Prediction of GWL with the help of GRACE TWS for unevenly spaced time series data in India: Analysis of comparative performances of SVR, ANN and LRM*. *Journal of Hydrology*, 558, 647–658. <https://doi.org/10.1016/J.JHYDROL.2018.02.005>
- Pathak, R., Awasthi, M. K., Sharma, S. K., Hardaha, M. K., & Nema, R. K. (2018). *Ground Water Flow Modelling Using MODFLOW –A Review*. *International Journal of Current Microbiology and Applied Sciences*, 7(2), 83–88. <https://doi.org/10.20546/IJCMAS.2018.702.011>
- Pawari, M. J., & Gawande, S. (2015). *Ground Water Pollution & Its Consequences*. *International Journal of Engineering Research and General Science*, 3(4).
- Saccò, M., Mammola, S., Altermatt, F., Alther, R., Bolpagni, R., Brancelj, A., Brankovits, D., Fišer, C., Gerovasileiou, V., Griebler, C., Guareschi, S., Hose, G. C., Korbel, K., Lictevout, E., Malard, F., Martínez, A., Niemiller, M. L., Robertson, A., Tanalgo, K. C., ... Reinecke, R. (2024). *Groundwater is a hidden global keystone ecosystem*. *Global Change Biology*, 30(1), e17066. <https://doi.org/10.1111/GCB.17066>
- Seibert, J., & Bergström, S. (2022). *A retrospective on hydrological catchment modelling based on half a century with the HBV model*. *Hydrology and Earth System Sciences*, 26(5), 1371–1388. <https://doi.org/10.5194/HESS-26-1371-2022>
- Vaux, H. (2011). *Groundwater under stress: The importance of management*. *Environmental Earth Sciences*, 62(1), 19–23. <https://doi.org/10.1007/S12665-010-0490-X/METRICS>
- Wada, Y. (2016). *Modeling Groundwater Depletion at Regional and Global Scales: Present State and Future Prospects*. 229–261. [https://doi.org/10.1007/978-3-319-32449-4\\_10](https://doi.org/10.1007/978-3-319-32449-4_10)
- Wu, H., Xu, J., Wang, J., & Long, M. (2021). *Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting*. *Advances in neural information processing systems*, 34, 22419–22430. <https://doi.org/10.48550/arXiv.2106.13008>
- Xiong, Z., Zhang, Z., Xin, Q., Xiong, Z., Zhang, Z., & Xin, Q. (2023). *A transformer-based model to predict surface soil moisture using meteorological variables for ecosystem process model*. *AGUFM*, 2023(2195), B31I-2195. <https://ui.adsabs.harvard.edu/abs/2023AGUFM.B31I2195X/abstract>
- Zhao, Y., Zhang, M., Liu, Z., Ma, J., Yang, F., Guo, H., & Fu, Q. (2024). *How Human Activities Affect Groundwater Storage*. *Research*, 7. <https://doi.org/10.34133/RESEARCH.0369>