

A Novel Deep Learning Algorithm for Groundwater Level Prediction based on Spatiotemporal Attention Mechanism

Chong Chen (✉ chenchong@cup.edu.cn)

China University of Petroleum-Beijing

Han Zhou

China University of Petroleum-Beijing

Hui Zhang

China University of Petroleum-Beijing

Lulu Chen

China University of Petroleum-Beijing

Zhu Yan

China University of Petroleum-Beijing

Huaqing Liang

China University of Petroleum-Beijing

Article

Keywords: Groundwater level, Spatiotemporal series, Neural network, Attention mechanism

Posted Date: August 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-59191/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A Novel Deep Learning Algorithm for Groundwater Level Prediction based on Spatiotemporal Attention Mechanism

Chong Chen^{1*}, Han Zhou¹, Hui Zhang¹, Lulu Chen¹, Zhu Yan¹, Huaqing Liang¹

1. College of Information Science and Engineering, China University of Petroleum-Beijing, Beijing, China

Abstract: Groundwater resources play a vital role in production, human life and economic development. Effective prediction of groundwater levels would support better water resources management. Although machine learning algorithms have been studied and applied in many domains with good enough results, the researches in hydrologic domains are not adequate. This paper proposes a novel deep learning algorithm for groundwater level prediction based on spatiotemporal attention mechanism. Short-term (one month ahead) and long-term (twelve months ahead) prediction of groundwater level are conducted with observed groundwater levels collected from several boreholes in the middle reaches of the Heihe River Basin in northwestern China. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to evaluate the performance of the proposed algorithm and several baseline models (i.e., SVR, Support Vector Regression; FNN, Feedforward Neural Networks; LSTM, Long Short-Term Memory neural network). The results show that the proposed model can effectively improve the prediction accuracy compared to the baseline models with MAE of 0.0754, RMSE of 0.0952 for short-term prediction and MAE of 0.0983, RMSE of 0.1215 for long-term prediction. This study provides a feasible and accurate approach for groundwater prediction which may facilitate decision making for water management.

Keywords: Groundwater level; Spatiotemporal series; Neural network; Attention mechanism

1 Introduction

Groundwater is an important source of drinking water, industrial water and agricultural irrigation water in many countries. Several unique qualities of groundwater (e.g. widespread and continuous availability, low development cost, drought reliability, etc.) has promoted the exploration of groundwater. In the past few decades, the effects of climate change and human activities have changed the distribution of groundwater resources seriously which led to shortages and over-exploitation of groundwater, as well as severe ecological and environmental problems in some arid and semi-arid areas ^{1,2}. Water resources management is a feasible solution for dealing with the abnormality in which the availability of groundwater resources is an important part. Groundwater level is an important indicator of groundwater resources which is highly nonlinear and nonstationary and depends on many heterogeneous environmental factors (e.g., aquifer properties, precipitation, surface water, exploration, evapotranspiration). The first procedure for groundwater management is the accurate simulation and prediction of groundwater level. Therefore, it is essential to explore and develop methods and models for groundwater level simulation.

Research on groundwater level simulation may date back to the nineteenth century. In 1856, the famous French water engineer Darcy proposed Darcy's law (also called the law of permeability) to describe the relationship between groundwater seepage and hydraulic gradient ³. In 1935, the Theis Equation was proposed to provide solution for calculating groundwater drawdown stimulated by the discharge of a well in transient groundwater hydraulics ⁴. At the end of the twentieth century, many hydrological models were developed (e.g., MODFLOW ⁵) which provided effective approaches to simulate and analyze the spatial-temporal variations in the

distribution of groundwater. Since then, the hydrological models have attracted the researchers attention to conduct groundwater level simulation. Furthermore, researchers tend to integrate different models for simulating the multidisciplinary nature of natural systems which lead to the development of agent-based model ⁶ and modeling environment (e.g., OMS ⁷, OpenMI ⁸). Cristian Guevara-Ochoa et al. coupled hydrological-hydrogeological model under different climate change scenarios to analyze the impact of climate change on water resources by quantifying the spatio-temporal dynamics of water balance and groundwater-surface water interactions ⁹. However, several issues of the traditional hydrological models should be addressed which include the relatively low accuracy, the high computation costs and the necessary requirements of expert knowledge.

With the development of computer science and hardware, the data-driven methods (e.g., regression analysis, statistics, gray theory, machine learning) have been thoroughly studied ¹⁰ and widely used in lots of areas with good-enough results ¹¹⁻¹³. In the fields of hydrological, French et al. developed a neural network to forecast rainfall intensity fields in space-time and presented the advantages of neural networks compared to numerical models ¹⁴. Since then, several researches have been conducted using data-driven models. Emery A et al. developed an Artificial Neural Networks (ANNs) model to predict groundwater level in a semi-confined glacial sand and gravel aquifer under variable state, pumping extraction and climate conditions with considerable accuracy ¹⁵. Uddameri constructed Feedforward Neural Network (FNN) models to forecast monthly and quarterly time-series water levels at a monitoring well with better results than the models developed by statistical regression techniques ¹⁶. Shiri et al. investigated the abilities of several methods (i.e., Gene Expression Programming, Adaptive Neuro-Fuzzy Inference System,

ANN, Support Vector Machine and Auto-Regressive Moving Average) for groundwater level prediction with the interval of 1-day up to 7-day ¹⁷. Gong et al. developed hybrid models by combining Ensemble Empirical Mode Decomposition (EEMD) with data-driven models to predict groundwater level fluctuations ¹⁸. Furthermore, Deep Learning (DL) algorithms have shown great potential in areas such as speech recognition, computer vision, natural language processing and intelligent recommendation ¹⁹. Yan et al. introduced a multivariable Long Short-Term Memory neural network (LSTM) to involve more information in the groundwater level prediction. A comparative study analyzed the pros and cons of machine learning and numerical models for simulating groundwater dynamics ¹². Although great progress has been made in the study of groundwater level prediction, some problems still exist. First of all, there are few researches on long-term groundwater level prediction researches which is essential for regional water resources management. Second, there is a lack of analysis and modeling of the spatial and temporal sequences of groundwater levels. Given the fact that many groundwater level observation boreholes are distributed in a watershed, certain spatial and temporal correlations among different boreholes exist which would be great help to improve the accuracy of groundwater prediction.

In this paper, we propose a novel DL algorithm (ST-Att-LSTM) based on the spatiotemporal attention mechanism to forecast short-term and long-term groundwater level by combining information from multiple observation boreholes. The main contributions of this paper are: (1) the proposal of a novel deep learning algorithm based on spatiotemporal attention mechanism to incorporate multiple data sources; (2) the short-term and long-term prediction of groundwater level based on the proposed algorithm with more accurate results compared to baseline methods; (3) a potential way of knowledge discovery for domain experts. The rest of the article is organized

as follows: the study area and collected data are described in Section 2; The structure of the proposed algorithm based on spatiotemporal attention mechanism are depicted in Section 3. In, Section 4, the evaluation of the proposed algorithm are demonstrated in Section 4; the conclusion and future works are shown in Section 5.

2 Study area and data description

The study area is selected as the middle reaches of the Heihe River Basin (HRB) (38°38'N~39°53'N, 98°53'E~100°44'E; Figure 1) which located in the Hexi Corridor, northwestern China. The climate in this area is a typical continental one with limited rainfall (69~216 mm/a) but strong evaporation (1453~2351 mm/a). The study area which characterized by alluvial plain oasis is the primary water consumption area where the irrigated agriculture consumes most of the water supply ²⁰. The groundwater resource was overexploited for agricultural, industrial, and domestic use. The major rivers in the study area are the mainstream of the Heihe River and the Liyuan River. Groundwater level changes within one year are affected by precipitation, evapotranspiration, surface water, agricultural irrigation. The annual and interannual variations of groundwater level are summarized by ²¹.

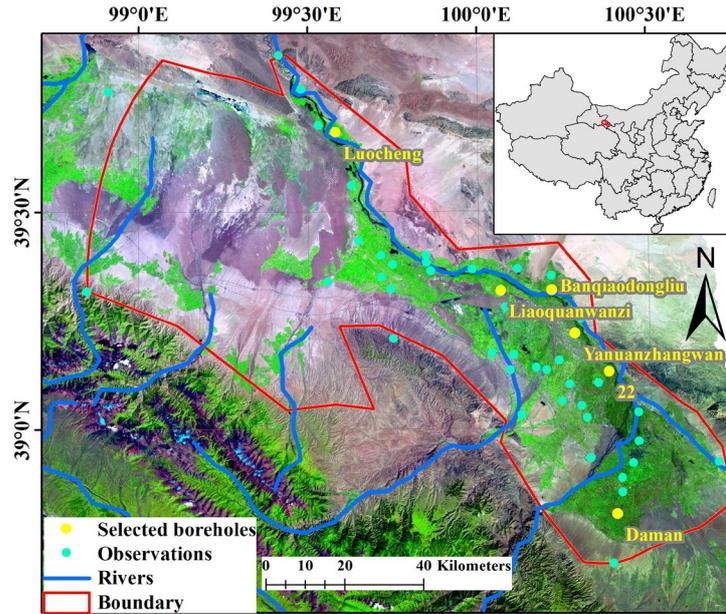


Figure 1. Location of the Heihe River Basin.

In the study area, the groundwater level observed at 42 observation boreholes (yellow and blue dots in Figure 1) were collected from 1986 to 2008. Groundwater level at observation borehole “22” was selected as the prediction target. Observation data from several spatially distributed boreholes near the river (i.e., “Daman”, “Yanuanzhangwan”, “Banqiaodongliu”, “Liaoquanwanzi” and “Luocheng”) (yellow dots in Figure 1) were selected to drive the algorithm. The spatial and geographical relationship between each observation well and the target observation borehole is differentiated by spatial attention mechanism. The data processing will be introduced in Section 4.

3 Model structure

The proposal of the novel DL algorithm for groundwater level prediction is based on LSTM, Sequence-to-sequence model (seq2seq) and attention mechanism. Therefore, the proposed algorithm would be depicted after brief descriptions on LSTM, seq2seq and attention mechanism in this section.

3.1 LSTM

The traditional Feedforward Neural Network (FNN) with back propagation establishes weight connections between layers and the outputs at the current time step which only depends on the input at the current time step. Recurrent Neural Network (RNN) was then proposed by ²² to overcome this limitation. In RNN, the hidden layer's information is able to accept its own information at the previous time step by establishing interconnection between hidden layers. However, the problem of long-term dependencies prevents RNN from large scale applications. To solve this problem, Hochreiter and Schmidhuber ²³ introduced Long Short-Term Memory neural networks. Figure 2 shows the structure of LSTM in which x_t refers to the input vector of time step t , h_t is the output of hidden vector. From this structure, we can derive that h_t contains information from h_{t-1} along with the input vector x_t . the information is passed to the next time step $t+1$ through the interconnection to ensure the memorization of information.

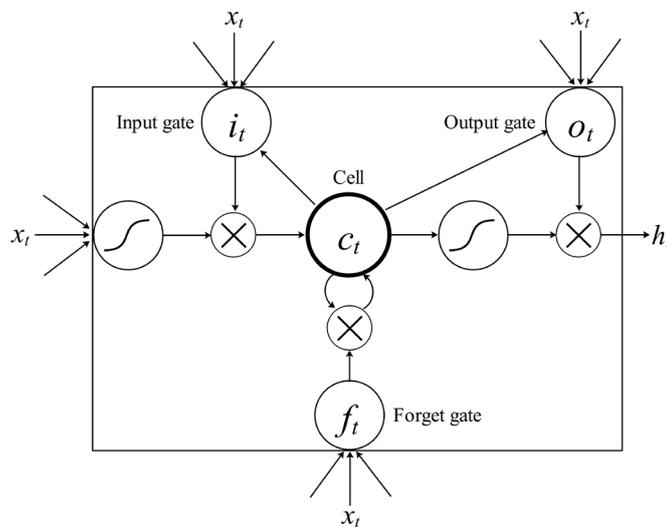


Figure 2. Structure of LSTM

The difference between LSTM and RNN is that it adds a “processor” to the algorithm to judge whether the information is useful or not. The structure of this processor is called cell. Three

gates are contained in a cell which named input gate, forget gate and output gate. The information which enters into LSTM can be judged according to the rules. Only information that complies with the algorithm's certification will be kept. The information that does not match will be discarded (forgotten) through the forget gate ²⁴. Equations (1) ~ (6) describe how the LSTM cell maps the input vector sequence x to the hidden vector sequence h in detail. In these equations, f_t , i_t , o_t , and C_t represent the forget gate, input gate, output gate and the memory cell vector, respectively; W_f , W_i , W_o , and W_C are matrices of weighting parameter; σ and \tanh are activation functions given by Equations (7) and (8).

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

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

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

3.2 seq2seq model

RNN requires fixed dimensions of inputs and outputs. In order to solve this problem, the seq2seq model is proposed. The most distinguished elements in seq2seq model is the inclusion of an encoder and a decoder ²⁵ with RNN processing the sequence information. The basic idea of seq2seq model is to use the encoder to compress the input sequence into a vector of a specified length which can be used as an implicit representation of the original sequence. The most direct

way to obtain the representation of the hidden vector is to use the hidden state at the last time step of the RNN. It is also possible to transform all hidden states of the input sequence to obtain hidden variables. The role of the decoder is to generate the specified result based on the input of the hidden vector C at the current time step. The common used calculation method is to input the hidden vector C obtained by the encoder as the initial state into the RNN structure of the decoder. The output will be used as the input of the current time step. The hidden vector C is used and only used as the initial state to participate in the operation.

3.3 Attention mechanism

The attention mechanism has recently been demonstrated success in a wide range of tasks ²⁶. LSTM typically uses the last hidden state or the mean of all hidden states as an output, while the attention mechanism allows for a more direct dependency between the states of the model at different points through time. According to the definition of ²⁷ which given a model that produces a hidden state h_t at each time step, the attention-based model calculates a "context" vector c_t as a weighted average of the state sequence h . c_t is calculated as:

$$c_t = \sum_{j=1}^T \alpha_{ij} h_j \quad (9)$$

where T is the total number of time steps in the input sequence and α_{ij} is weight computed at each time step t for each state h_j . These context vectors are then used to calculate a new state sequence s , where s_t depends on s_{t-1} , c_t and the model outputs at $t-1$. The weights α_{ij} are then computed by Equations (10) and (11).

$$e_{ij} = a(s_{t-1}, h_j) \quad (10)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^T \exp(e_{ik})} \quad (11)$$

3.4 Spatiotemporal attention model for groundwater level prediction

The spatial and temporal dynamics of the groundwater level are reflected in the spatial distribution of different observation wells. Each observation well records the groundwater level in time series. The groundwater level of one observation borehole may be closely related to the previous time of another observation well. To this end, we propose to use the attention mechanism to model this spatiotemporal dynamic relationship (refer to ST-Att-LSTM hereafter). The goal of the spatial attention is to dynamically change the influence of other observation boreholes on the target observation borehole at the current time step. The goal of the temporal attention is to combine the input sequence with different weights corresponding to the prediction results at each time step.

Spatial attention refers to the inputs of each observation borehole. A certain attention weight is given to each observation borehole according to the influence factor. This weight α_t^k measures the importance of the groundwater level of the k -th observation borehole at time t . The overall inputs after weighting are:

$$\tilde{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n) \quad (12)$$

The weight α_t^k can be calculated according to the hidden state and cell state of the encoder at the last time step as:

$$e_t^k = v_e^t \tanh(W_e[h_{t-1}; s_{t-1}] + U_e x_k) \quad (13)$$

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{j=1}^N \exp(e_t^j)} \quad (14)$$

where v_e^t , W_e and U_e are trainable parameters. The structure of the spatial attention is shown in Figure 3.

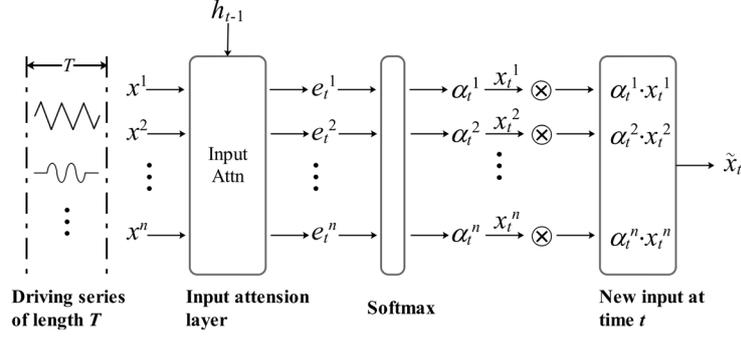


Figure 3. Structure of spatial attention mechanism

After the spatial attention mechanism, the inputs are updated by multiplying the attention weights. The encoder based on the LSTM is used for time-series encoding. Different types of decoding strategies are designed based on the characteristics of short-term and long-term prediction. For short-term prediction, the output of each time step of LSTM passes through the attention layer, The final outputs derive from the weighting of the output of each time step. In Equations (15)~(17), *LSTM* represents the calculation process of a hidden layer unit; *ATT* represents the calculation process of the attention mechanism; c_t represents the resulting context vector; *FC* represents the fully connected model, and y is the final prediction output.

$$h_t = LSTM(x_t, h_{t-1}) \quad (15)$$

$$c_t = ATT(h_t, e_t) \quad (16)$$

$$y = FC(c_t) \quad (17)$$

For long-term prediction, the seq2seq model structure is used with LSTM being the decoder. Each time step of the decoding process obtains different encoding vectors through the time attention mechanism. In Equations (18)~(20), *decoder* represents the calculation process of a hidden layer unit of the encoder; c_t represents the context vector in the current decoding state.

$$s_t = decoder(y_{t-1}, s_{t-1}) \quad (18)$$

$$\hat{s}_t = \tanh(W_c[s_t; c_t]) \quad (19)$$

$$y_t = FC(\hat{s}_t) \quad (20)$$

The structure of spatiotemporal attention mechanism is shown in Figure 4.

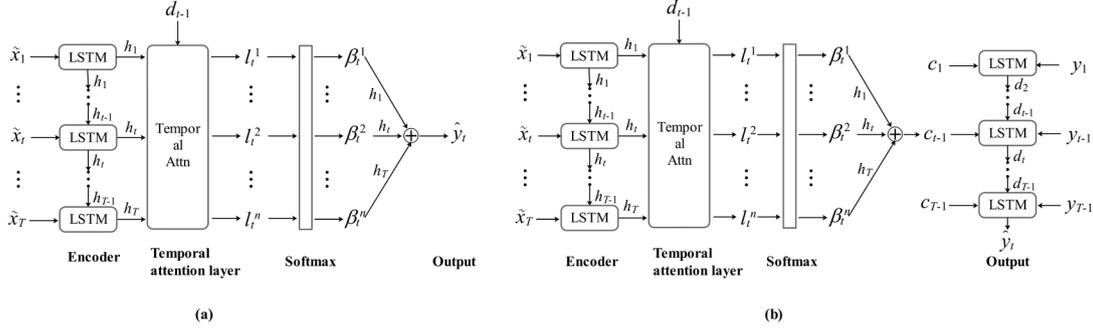


Figure 4. Structure of spatiotemporal attention mechanism ((a) the short-term prediction model; (b) the long-term prediction model)

4 Experiment and Results

4.1 Raw data processing

The information of the middle reaches of HRB where the data was collected from are introduced in Section 2. The location of the target observation boreholes and auxiliary boreholes selected in our experiment is shown in Figure 1. The dataset of historical groundwater levels from January 1986 to December 2008 were collected and processed with monthly interval. The dataset was divided into training set and validation set with proportion of 90% and 10%. The experiment is designed to predict the groundwater level of the target observation borehole with the groundwater level sequences of the auxiliary boreholes. The groundwater level at previous time steps (refer to time window hereafter) are used to obtain the groundwater level at the current/future time steps. The size of time window for the input is $T=12$ which means the input sequence is $\{x_t, x_{t+1}, x_{t+2}, \dots, x_{t+11}\}$. The output is x_{t+12} (with the size of time window being 1) for short-term prediction and $\{x_{t+11+1}, x_{t+11+2}, \dots, x_{t+11+\tau}\}$ (with the size of time window being τ) for the long-term prediction where τ is the length of the output sequence.

Because of the existence of the characteristics of different dimension eigenvalues and target eigenvalue changes, the data was scaled to $[0,1]$ using Equation (21).

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (21)$$

where x_i is the normalized data, x is the raw data, x_{min} and x_{max} are the minimum value and maximum value of x .

4.2 Parameters and performance evaluation

The environments (software and hardware) for conducting the experiment are shown in Table 1. Backpropagation with Adaptive moment estimation (Adam) ²⁸ is used to train the proposed algorithm. The training is to minimize a loss function with regard to the parameters. A mini-batch gradient descent optimization method which is simple, efficient and computationally inexpensive is used to train the proposed model. It is suitable for non-stationary objective functions. The hyper-parameters of the proposed algorithm and baseline models will be described in the relevant sections.

Table 1. The environments of the experiment

System	Ubuntu 18.04 64-bit
Software	Python 3.6 PyCharm Community Edition 2018.2.3 Keras 2.1.5
Hardware	Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz 16 GB RAM NVIDIA GEFORCE RTX 2080 Ti

The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used to evaluate the performance of the algorithm. RMSE finds large errors and results in a scale which is in the same units as the forecast data. MAE can better reflect the error between the observations and the predictions. The indexes are calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (22)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (23)$$

In the equations, n stands for the number of values, y_i refers to the observed groundwater

level, \hat{y}_i refers to the calculated groundwater level. From the definition, one can conclude that better results are reflected by RMSE and MAE values near zero.

4.3 Short-term prediction

To evaluate the proposed neural network with spatiotemporal attention mechanism (ST-Att-LSTM), Support Vector machine Regression (SVR), Feedforward Neural Network (FNN), LSTM, and LSTM with spatial attention mechanism (S-Att-LSTM) are used as baseline models. The hyper-parameters of all the baseline models and the proposed algorithms are optimized and shown in Table 2. The results shown in Figure 5 indicate reasonable match for all the models. The results from LSTM are better than those from SVR and FNN because of the inclusion of time dependencies in the time sequence. However, S-Att-LSTM performs better which may benefit from the features extracted from different time sequences distributed spatially in the study area. The best match is obtained by ST-Att-LSTM because of the spatiotemporal attention mechanism which would consider both the spatial and temporal relations with several certain attention weights. A quantitative comparison between different methods is conducted using the performance criteria (i.e., RMSE and MAE). The results shown in Table 3 indicate the best match from ST-LSTM which is in accordance with Figure 5. It is apparent that better results are obtained from the LSTM-based algorithms than those obtained from SVR and FNN which may be attributed to that LSTM-based algorithms consider the information through time. The S-Att-LSTM algorithm with only spatial attention mechanism performs better than the traditional LSTM without attention mechanism. The ST-Att-LSTM algorithm with spatiotemporal attention mechanisms performs best among three LSTM-based algorithms. Comparisons between LSTM-based algorithms indicates that the attention mechanism further improve the performance of the algorithms.

Table 2. Hyper-parameters for baseline models (SVR, FNN, LSTM, S-Att-LSTM) and ST-Att-LSTM in short-term prediction

Name	Number of hidden layers	Number of units	Learning rates	Number of epochs	Number of parameters
SVR		C = 1.0, epsilon = 0.01			-
FNN	2	128/64	0.01	100	17665
LSTM	1	128	0.01	100	69249
S-Att-LSTM	2	128	0.01	100	70146
ST-Att-LSTM	2	128	0.01	100	71386

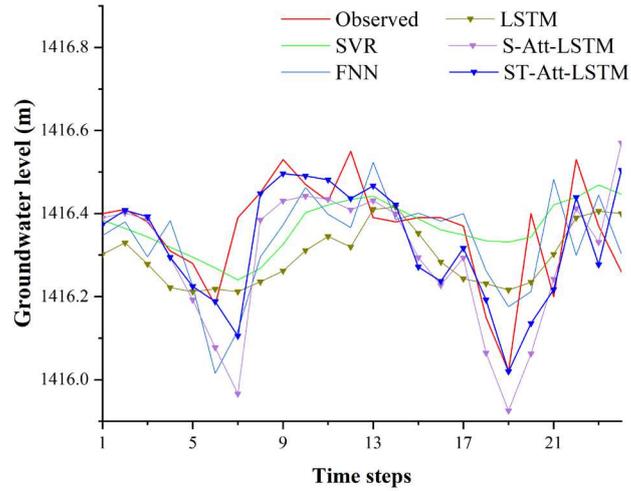


Figure 5. Short-term prediction of groundwater level

Table 3. Evaluation of short-term prediction results

Model	MAE	RMSE
SVR	0.0928	0.1238
FNN	0.1058	0.1346
LSTM	0.0854	0.1275
S-Att-LSTM	0.0837	0.1156
ST-Att-LSTM	0.0754	0.0952

Furthermore, the attention weights of ST-Att-LSTM are examined to verify that the effectiveness of spatiotemporal attention mechanism on combining the effects from different observation boreholes and different time steps (shown Figure 6). It can be seen that the groundwater level of observation borehole “22” at the current time step is most related to the previous two-time steps and the initial time step (Figure 6 (a)). Due to the influence of the routine

groundwater pumping, agricultural irrigation and upstream runoff, certain annual pattern exists in the observation of groundwater level. Therefore, the groundwater level at the predicted time step correlates the most with the groundwater level at the same month and the adjacent time steps. At the same time, it can be seen from the Figure 6 (b) which depicts the spatial attention weights that the groundwater level of the prediction target (observation borehole “22”), besides having the highest correlation with its own historical data, is also highly relevant to the groundwater level of Banqiaodongliu observation wells due to the strong hydraulic connections. The distributions of attention weights indicate the effectiveness of spatiotemporal attention mechanism which can significantly extract the spatiotemporal dynamic relationships between different observation boreholes and time steps which would benefit the prediction accuracy.

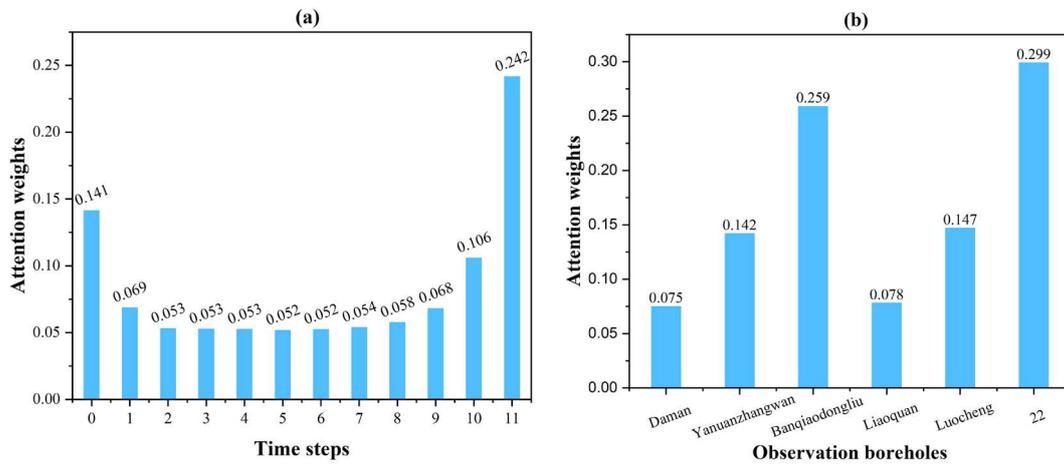


Figure 6. Spatiotemporal attention weight: (a) attention weights among different time steps; (b) attention weights among different observation boreholes

4.4 Long-term prediction

To evaluate the proposed algorithm (ST-Att-LSTM), different baseline algorithms (ARIMA, LSTM, seq2seq, seq2seq with time attention (seq2seq-Att)) are selected deliberately. The size of predicted time window is set to 12 ($\tau=12$ in Section 4.1) which denotes the length of the output sequence. The same methods for pre-processing data and optimizing neural networks are used as

in the experiment for short-term prediction are used in this experiment. The hyper-parameters of all the baseline models and the proposed algorithms are optimized and shown in Table 4. The results of the algorithms are shown in Figure 7 which indicate that the performance of traditional methods (ARIMA, LSTM) for long-term prediction are worse than those from the advanced algorithms (seq2seq, seq2seq-Att, ST-Att-LSTM). Parts of the reason may be contributed to the larger number of parameters in the advanced algorithms which brings more degree of freedom to approximate the data. It is noted that a local minimum value around the 9th time step is mismatched by all the algorithms which may be accessible considering the following two reasons. First, the purpose of the long-term prediction is to capture the trend of the data (which has already been captured well) rather than some special value in case of overfitting. Second, the groundwater level at time step 9 could be considered as an abnormal variation in this scale. A qualitative comparison between different algorithms is conducted by calculating the MAE and RMSE (Table 5). The performance of the proposed ST-Att-LSTM algorithm ranks the best in all the algorithms which is in accord with Figure 7.

Table 4. Hyper-parameters for baseline models (ARIMA, LSTM, seq2seq, seq2seq-Att and ST-Att-LSTM in long-term prediction

Name	Number of hidden layers	Number of units	Learning rates	Number of epochs	Number of parameters
ARIMA	(p, d, q) = (1, 1, 1), (P, D, Q) = (1, 0, 1), s = 12				-
LSTM	1	256	0.01	100	272396
seq2seq	1	256	0.005	100	794881
seq2seq-Att	1	256	0.005	100	795394
ST-Att-LSTM	1	256	0.005	100	810376

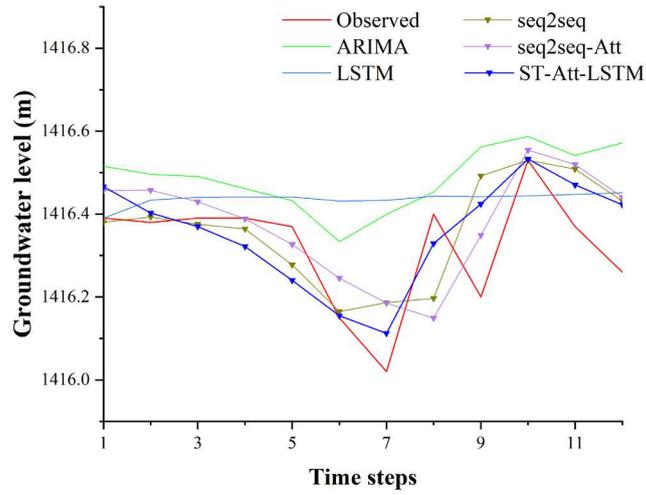


Figure 7. Long-term prediction of groundwater level

Table 5. Evaluation of long-term prediction results

Model	MAE	RMSE
ARIMA	0.1664	0.2021
LSTM	0.1533	0.2086
seq2seq	0.1165	0.1447
seq2seq-Att	0.1029	0.1367
ST-Att-LSTM	0.0983	0.1215

5 Conclusion

In this paper, a groundwater level prediction algorithm based on the spatiotemporal attention mechanism was proposed to explore the potential information in spatial and temporal distribution. The groundwater level from five spatial distributed observation boreholes (i.e., “Daman”, “Yanuanzhangwan”, “Banqiaodongliu”, “Liaoquanwanzi”, “Luocheng” and “22”) in the middle reaches of the HRB were collected and used to train and validate the proposed algorithm both in the short-term prediction and long-term prediction. The MAE and RMSE values were used to evaluate the performance by comparing with several baseline algorithms. The results showed that the inclusion of spatiotemporal information by attention mechanisms was able to improve the performance of deep learning algorithms significantly. Furthermore, the attention weights among

different spatial-distributed observation boreholes were examined which may provide information for potential hydraulic connections for hydrologist. Therefore, the proposed algorithm provides (1) a potential application of spatiotemporal data; (2) an accurate prediction for groundwater level; and (3) a potential way of discovering knowledge for domain experts. Future research should focus on exploring the solution for combining domain expert knowledge into the deep learning algorithms to obtain accurate results and hence develop explainable deep learning algorithms.

Acknowledgements

This work was supported by Science Foundation of China University of Petroleum-Beijing under grant No. 2462018YJRC007, 2462020YXZZ025. Gratitude is expressed to the Cold and Arid Regions Science Data Center at Lanzhou (<http://westdc.westgis.ac.cn>) for providing data.

Author Contributions

C.C. initiated and designed this work. C.C. and H. Z. designed and implemented the experiments with inputs from L. L. C. Z. Y. processed the data and the results. H. Q. L supervised the work and organized the resources dedicated to the project. C.C. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

References

- 1 Grafton, R. Q., Ward, M. B., To, H. & Kompas, T. Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. *Water Resour. Res.* **47**, doi:10.1029/2010wr009685 (2011).
- 2 Barzegar, R., Moghaddam, A. A., Tziritis, E., Fakhri, M. S. & Soltani, S. Identification of hydrogeochemical processes and pollution sources of groundwater resources in the Marand plain, northwest of Iran. *Environ. Earth Sci.* **76**, 297, doi:10.1007/s12665-017-6612-y (2017).
- 3 Swain, E. D. & Wexler, E. J. A coupled surface-water and ground-water flow model (MODBRANCH) for simulation of stream-aquifer interaction. Report No. 06-A6, (1996).

- 4 Chang, F.-J., Chang, L.-C., Huang, C.-W. & Kao, I. F. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *Journal of Hydrology* **541**, 965-976, doi:<https://doi.org/10.1016/j.jhydrol.2016.08.006> (2016).
- 5 Hughes, J. D., Langevin, C. D. & Banta, E. R. Documentation for the MODFLOW 6 framework. Report No. 6-A57, (Reston, VA, 2017).
- 6 Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7280-7287, doi:10.1073/pnas.082080899 (2002).
- 7 David, O. *et al.* Object Modeling System v3.0: Developer and User Handbook. (2010).
- 8 Gregersen, J. B., Gijsbers, P. J. A. & Westen, S. J. P. OpenMI: Open modelling interface. *J Hydroinform* **9**, 175-191, doi:DOI 10.2166/hydro.2007.023 (2007).
- 9 Guevara-Ochoa, C., Medina-Sierra, A. & Vives, L. Spatio-temporal effect of climate change on water balance and interactions between groundwater and surface water in plains. *Sci. Total Environ.* **722**, 137886, doi:<https://doi.org/10.1016/j.scitotenv.2020.137886> (2020).
- 10 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature*. **521**, 436-444, doi:10.1038/nature14539 (2015).
- 11 Zhou, H., Chen, C., Liu, H., Qin, F. & Liang, H. in *CCF International Conference on Natural Language Processing and Chinese Computing*. 724-735 (Springer).
- 12 Chen, C., He, W., Zhou, H., Xue, Y. & Zhu, M. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. *Scientific Reports* **10**, 3904, doi:10.1038/s41598-020-60698-9 (2020).
- 13 DeVries, P. M. R., Viégas, F., Wattenberg, M. & Meade, B. J. Deep learning of aftershock patterns following large earthquakes. *Nature* **560**, 632-634, doi:10.1038/s41586-018-0438-y (2018).
- 14 French, M. N., Krajewski, W. F. & Cuykendall, R. R. Rainfall forecasting in space and time using a neural network. *Journal of Hydrology* **137**, 1-31, doi:[https://doi.org/10.1016/0022-1694\(92\)90046-X](https://doi.org/10.1016/0022-1694(92)90046-X) (1992).
- 15 Coppola, E. A., Jr., Rana, A. J., Poulton, M. M., Szidarovszky, F. & Uhl, V. W. A neural network model for predicting aquifer water level elevations. *Ground water* **43**, 231-241, doi:10.1111/j.1745-6584.2005.0003.x (2005).
- 16 Uddameri, V. Using statistical and artificial neural network models to forecast potentiometric levels at a deep well in South Texas. *Environmental Geology* **51**, 885-895, doi:10.1007/s00254-006-0452-5 (2007).
- 17 Shiri, J., Kisi, O., Yoon, H., Lee, K.-K. & Hossein Nazemi, A. Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. *Computers & Geosciences* **56**, 32-44, doi:<https://doi.org/10.1016/j.cageo.2013.01.007> (2013).
- 18 Gong, Y. C., Wang, Z. J., Xu, G. Y. & Zhang, Z. X. A Comparative Study of Groundwater Level Forecasting Using Data-Driven Models Based on Ensemble Empirical Mode Decomposition. *Water* **10**, 20, doi:10.3390/w10060730 (2018).
- 19 闫佰忠, 孙剑, 王昕洲, 韩娜 & 刘博. 基于多变量 LSTM 神经网络的地下水水位预测. *吉林大学学报(地球科学版)* **50**, 208-216 (2020).
- 20 Wu, F., Zhan, J. & Güneralp, İ. Present and future of urban water balance in the rapidly urbanizing Heihe River Basin, Northwest China. *Ecol. Model.* **318**, 254-264,

- doi:<https://doi.org/10.1016/j.ecolmodel.2014.11.032> (2015).
- 21 米丽娜 *et al.* 1985-2013 年黑河中游流域地下水位动态变化特征. *冰川冻土* **37**, 461-469 (2015).
- 22 Hiji, S. E. & Bengio, Y. in *neural information processing systems*. 493-499.
- 23 Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Computation* **9**, 1735-1780 (1997).
- 24 Ke, J., Zheng, H., Yang, H. & Chen, X. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies* **85**, 591-608, doi:<https://doi.org/10.1016/j.trc.2017.10.016> (2017).
- 25 Zaytar, A. & Amrani, C. E. Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. *International Journal of Computer Applications* **143**, 7-11 (2016).
- 26 Su, H., Li, X., Yang, B. & Wen, Z. Wavelet support vector machine-based prediction model of dam deformation. *Mechanical Systems and Signal Processing* **110**, 412-427, doi:<https://doi.org/10.1016/j.ymssp.2018.03.022> (2018).
- 27 Bahdanau, D., Cho, K. & Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR* **abs/1409.0473** (2015).
- 28 赵小强 & 宋昭漾. Adam 优化的 CNN 超分辨率重建. *计算机科学与探索* **13**, 858-865 (2019).

Figures

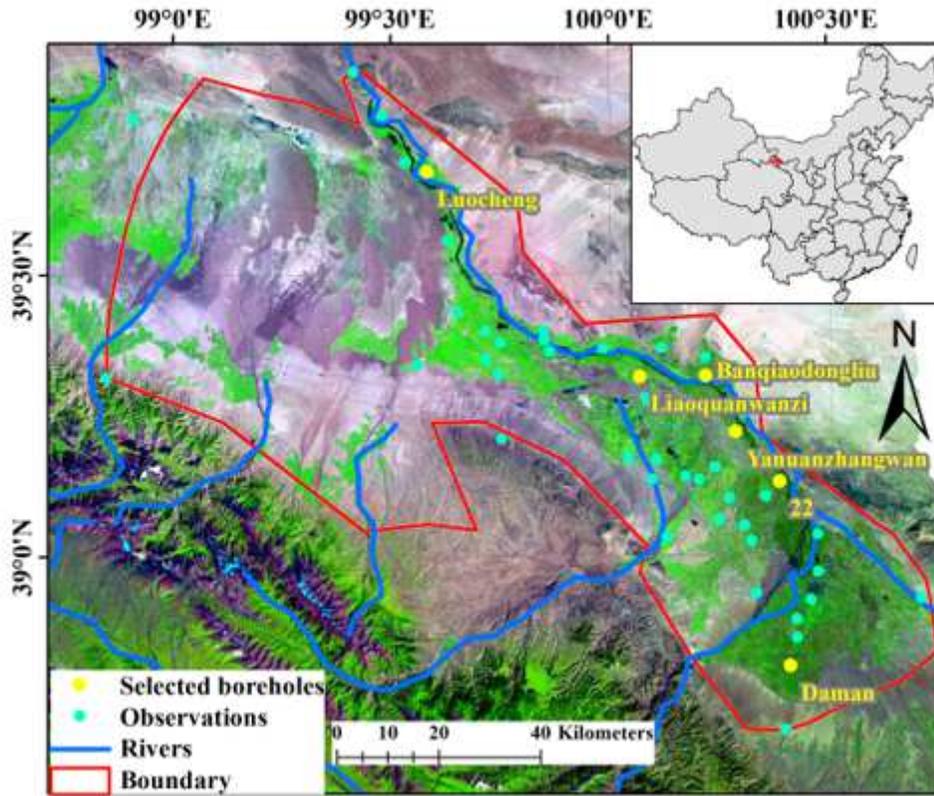


Figure 1

Location of the Heihe River Basin. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

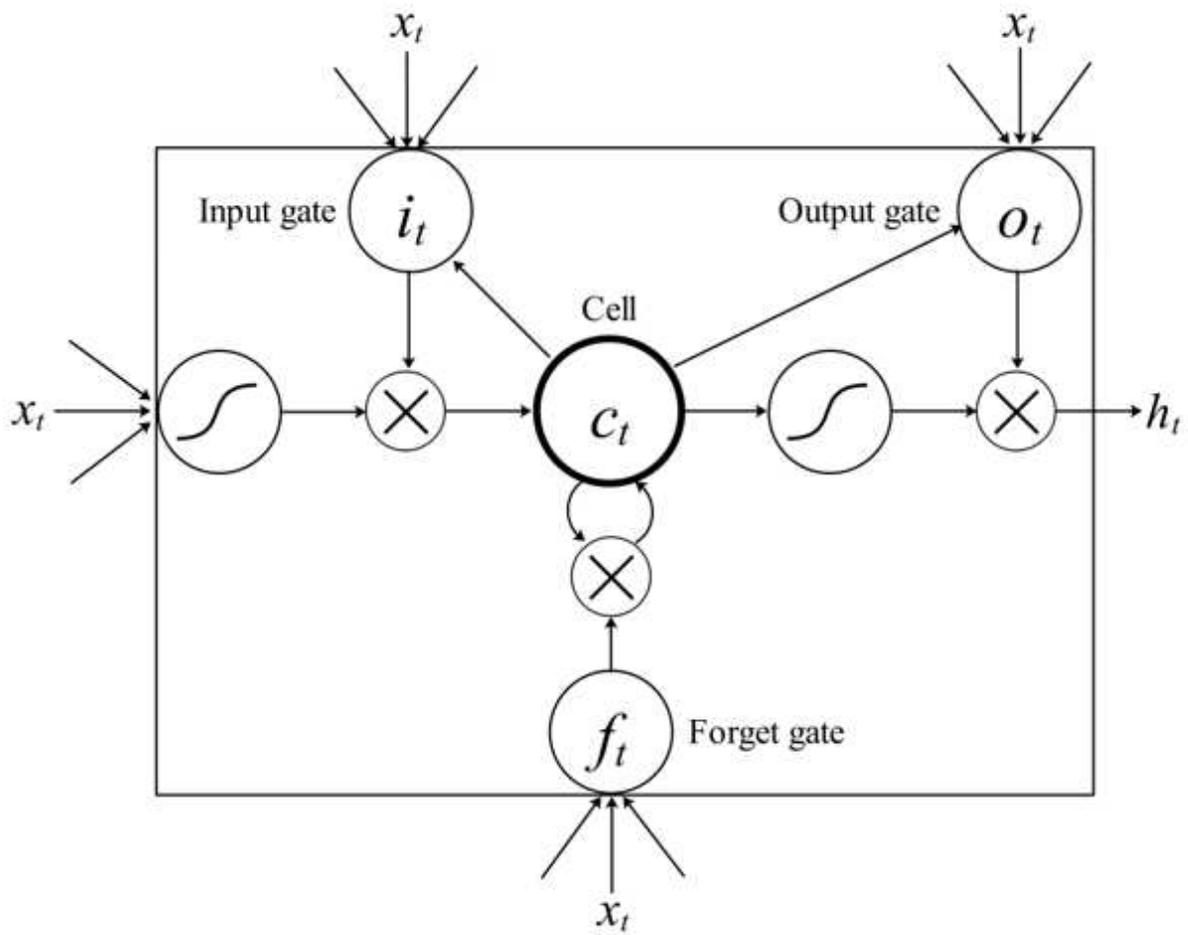


Figure 2

Structure of LSTM

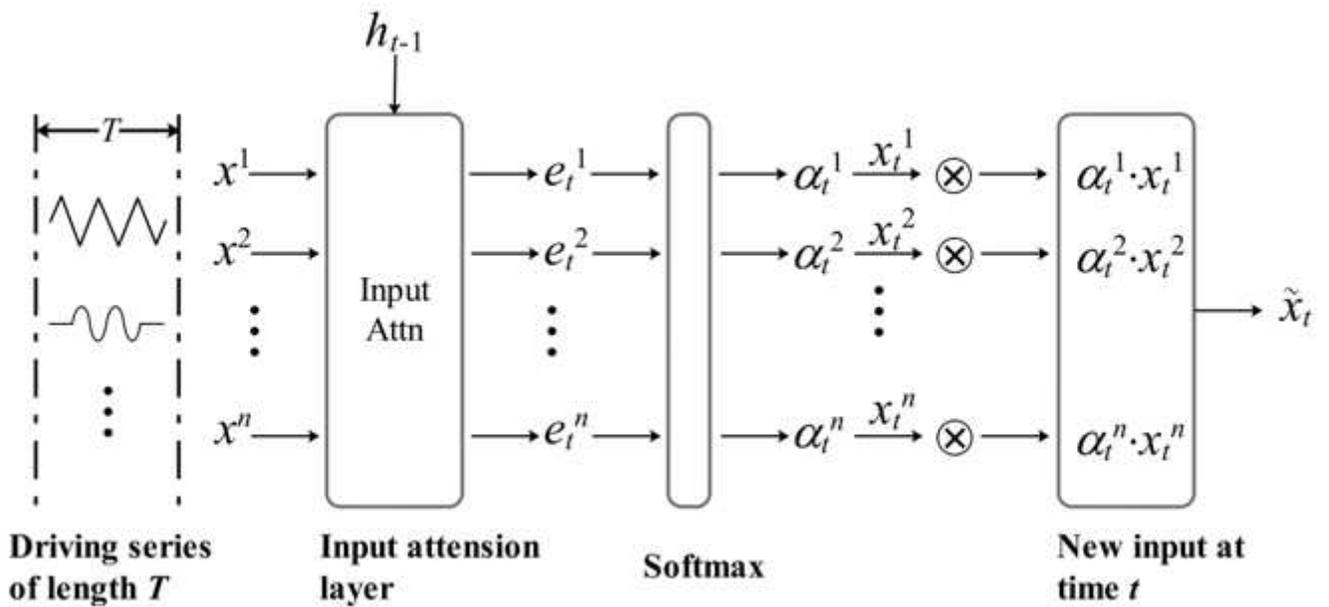


Figure 3

Structure of spatial attention mechanism

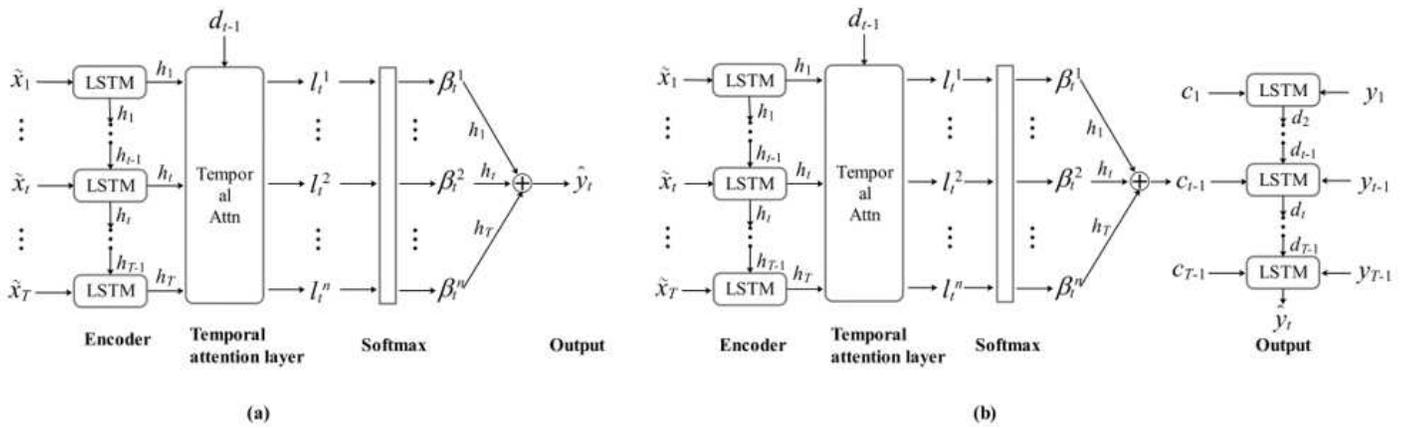


Figure 4

Structure of spatiotemporal attention mechanism ((a) the short-term prediction model; (b) the long-term prediction model)

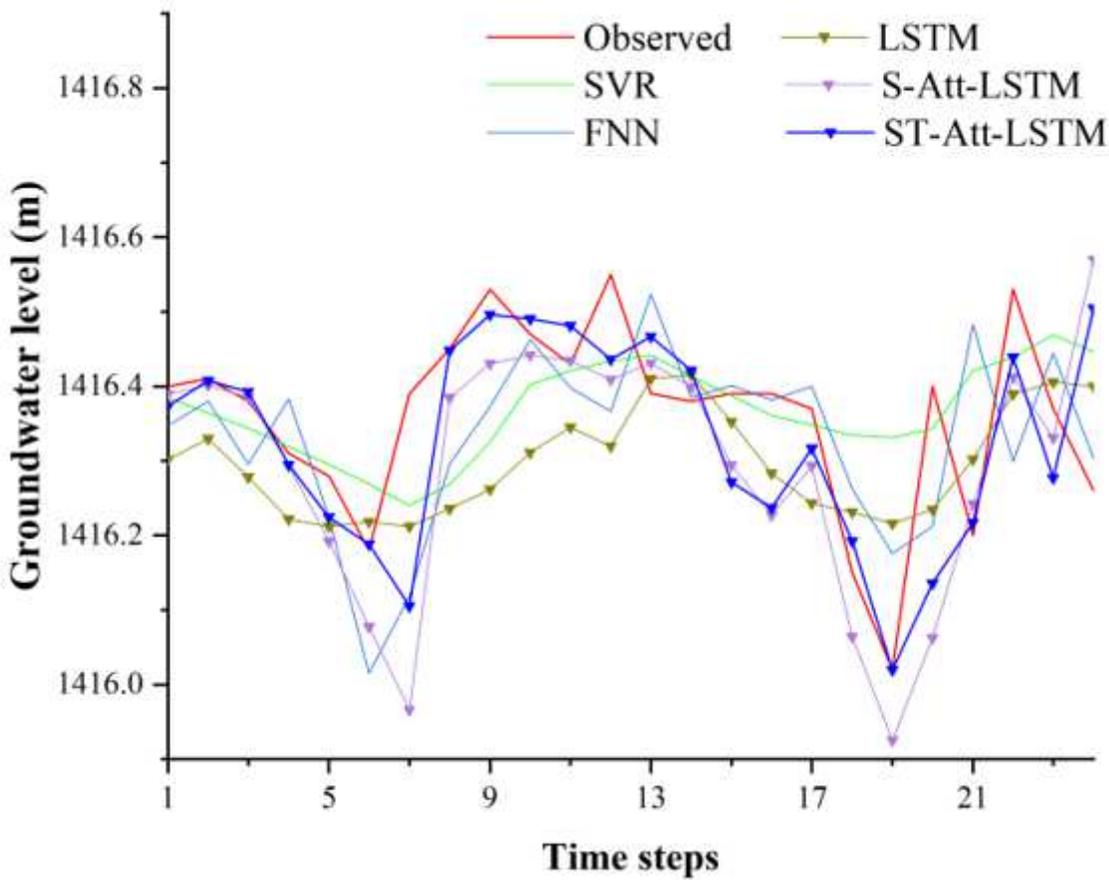


Figure 5

Short-term prediction of groundwater level

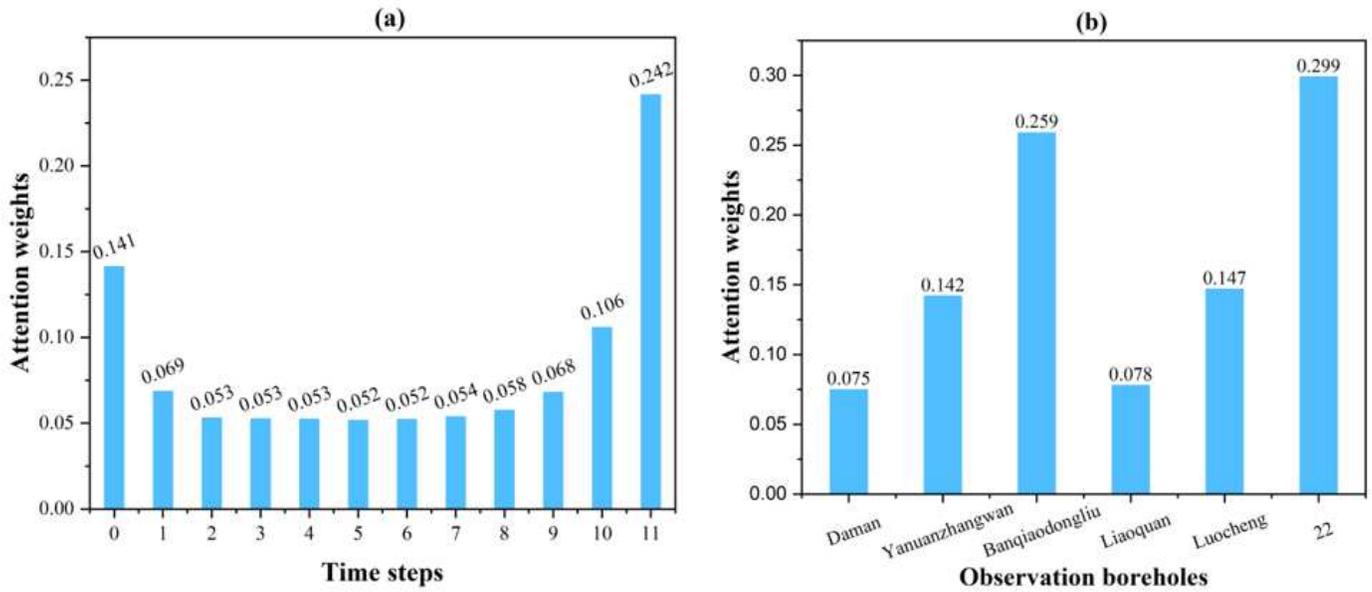


Figure 6

Spatiotemporal attention weight: (a) attention weights among different time steps; (b) attention weights among different observation boreholes

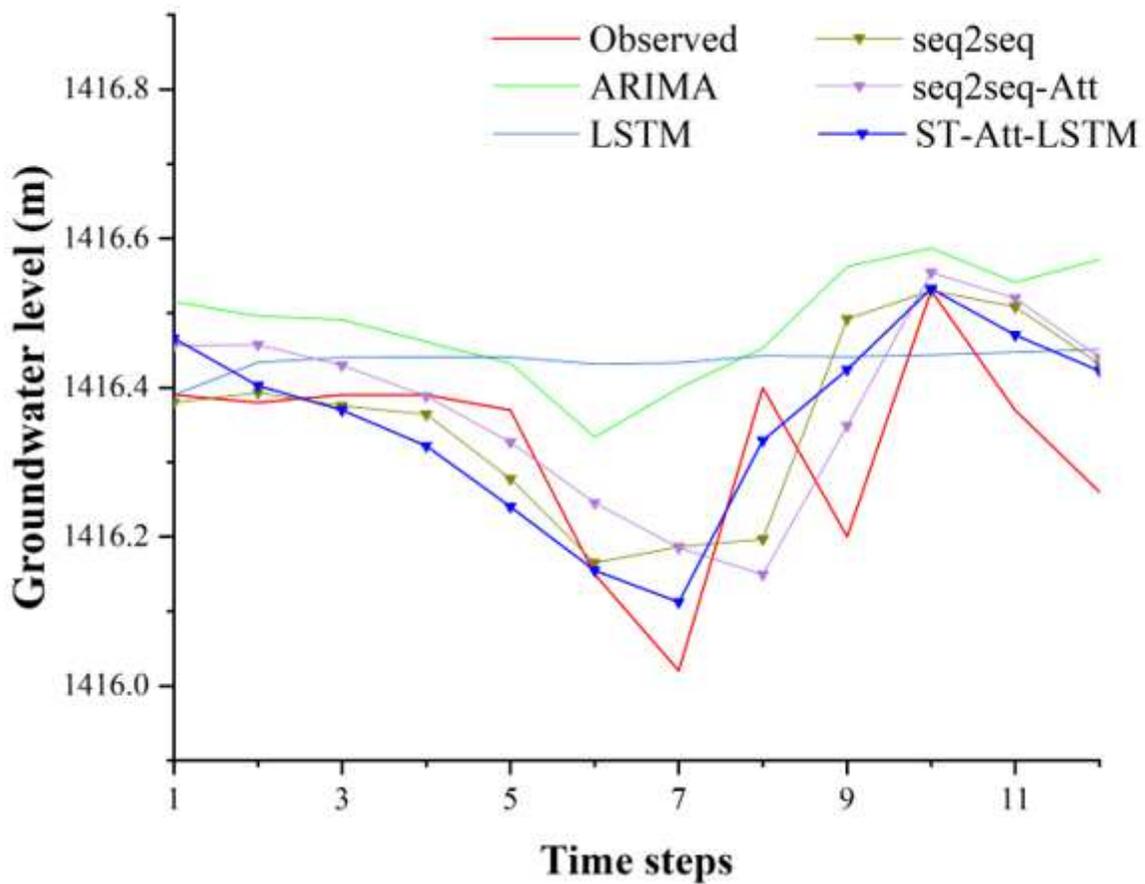


Figure 7

Long-term prediction of groundwater level