2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress

A Deep Learning Algorithm for Groundwater Level Prediction based on Spatial-temporal Attention Mechanism

Chong Chen
College of Information Science and Engineering
China University of Petroleum – Beijing
Beijing, China
chenchong@cup.edu.cn

Xiaoyu Zhu
College of Information Science and Engineering
China University of Petroleum – Beijing
Beijing, China
stonefree925@163.com

Xiaobin Kang
School of Information Engineering
Lanzhou Resource & Environment Voc-Tech College
Lanzhou, China
1328197301@qq.com

Han Zhou
College of Information Science and Engineering
China University of Petroleum – Beijing
Beijing, China
ce_kong1@163.com

***Abstract*—Groundwater modeling is an effective solution for water resources management and exploring complex groundwater-dependent ecosystems. Although groundwater models have been investigated for >5 decades, the computation costs and model accuracy of numerical models are limited by the assumption and conceptualization of the physical mechanisms. Deep learning algorithms have been widely studied and applied in classifications and regressions which provides a feasible solution for groundwater level prediction. This paper proposed a novel deep learning algorithm for groundwater level prediction based on spatial-temporal attention mechanism to improve the accuracy and efficiency of groundwater modeling. Two kind of groundwater level predictions—short-term (one month ahead) and long-term (twelve months ahead) prediction—were conducted. Observed groundwater levels collected in the middle reaches of the Heihe River Basin in northwestern China were used to train and test the algorithm. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to evaluate the performance of the proposed algorithm and several baseline models. The results show that the proposed model is able to effectively improve the prediction accuracy compared to 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 cost-effective and accurate approach for groundwater prediction which may facilitate policy making for water management.**

Keywords—Groundwater, Spatial-temporal, Deep Learning, Attention mechanism

I. Introduction

Groundwater is an important source of drinking water, industrial water and agricultural irrigation water in many countries. In the past several decades, the effects of climatic changes and human activities have changed the distribution and storage of groundwater resources seriously. The direct/indirect influence of climatic changes and human activities will led to hydrological extreme events (e.g., water shortages, floods, draught, etc.), especially in some arid and semi-arid areas [1]. There is a consensus that the climatic changes will increase both frequency and intensity of hydrological extreme events (e.g., rainstorm in Zhengzhou, Henan Province [2]; flood in Taklimakan [3]). Aquifers can be used as water bank, which therefore, can relieve the water scarcity and water abundant caused by climate extremes. Groundwater level is one of the important indicators of groundwater resources which is also highly nonlinear and nonstationary because of heterogeneous impact factors. The first necessity for managing groundwater is the simulation of groundwater system in the aquifers. Research on groundwater simulation may date back to the nineteenth century [4]. At the end of the twentieth century, many numerical models had been developed (e.g., MODFLOW [5], FEFLOW [6], etc.) to provide effective approaches for simulating and analyzing the spatial-temporal variations groundwater. Since then, numerical models have attracted the researcher's attention to conduct groundwater level simulation. With numerical models, the researchers are able to understand the behavior of groundwater system, explore complex groundwater-dependent ecosystems, evaluate groundwater storage, support water resources management decision-making. Furthermore, researchers tend to integrate different models for simulating the multidisciplinary nature of natural systems using agent-based model [7] and modeling environments (e.g., OMS [8], OpenMI [9]) [10]. However, both the application of single numerical models and the integration of multiple numerical models will dramatically enhance the burden of computation. Other issues of numerical models which should be addressed include the relatively low accuracy and the necessary requirements of expert knowledge.

Fortunately, the data-driven methods (e.g., regression analysis, gray theory, machine learning) which have been thoroughly studied [11] and widely used in lots of areas provide cost-effective, high accuracy solutions for groundwater modelling [12]. In hydrology, French et al. developed a three-layer neural network to forecast rainfall intensity fields in space-time. Their results indicated the effectiveness and advantages of neural networks in simulating the space-time evolution of rainfall [13]. Since then, several researches have been conducted using data-driven models. Emery A et al. developed an Artificial Neural Networks (ANNs) model to predict water level elevations at two monitoring wells considering production well extraction and climate conditions [14]. They found that ANNs were more appropriate for making groundwater management strategies. Uddameri constructed Feedforward Neural

Network (FNN) models to predict monthly and quarterly time-series water levels at a monitoring well in the confined layer [15]. 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 and found that the AI methods performed better than Auto-Regressive Moving Average [16]. Gong et al. combined Ensemble Empirical Mode Decomposition (EEMD) with three data-driven models (ANN, Support Vector Machine and Adaptive Neuro Fuzzy Inference System) to predict groundwater level fluctuations and improved the prediction accuracy of groundwater level fluctuations [17]. Yadav et al predicted monthly groundwater level fluctuations with ANN and SVM. It should be noted that they pre-processed the input data (groundwater levels, rainfall, temperature, NOI, SOI, NINO3 and monthly population growth rate) by mutual information theory, genetic algorithm and lag analysis. A comparative study analyzed the pros and cons of machine learning methods and numerical models for simulating groundwater dynamics [12] and concluded that regardless of the physical mechanism, the machine learning algorithms reached higher prediction accuracy, modeling efficiency with lower computation costs. As the volume of data and data processing methods evolving, Deep Learning (DL) algorithms have shown great potential in speech recognition, computer vision, natural language processing and intelligent recommendation [11]. Yan et al. introduced a multivariable Long Short-Term Memory neural network (LSTM) to conduct groundwater level prediction and obtained better performance compared with BP (Back Propagation) Neural Network [18]. Mohapatra et al. predicted the seasonal groundwater levels at the country scale using DL algorithms and concluded that Deep Neural Network model is the most efficient in the prediction among ANFIS and SVM [19].

Although progress has been made groundwater level prediction, several issues remain. First, few researches focusing on long-term prediction of groundwater level which is essential for regional water resources management; second, there is a lack of researches focusing on taking full advantage of the spatial and temporal information of data. Du et al. reviewed four machine learning methods concluded that the machine learning methods are suitable to handle spatial data [20]. Given the fact that the boreholes for observing groundwater level are distributed in the whole watershed, certain spatial-temporal correlations among different boreholes exist which would be additional information in improving the accuracy of groundwater level prediction. Therefore, it is important to take the spatial-temporal correlations into consideration when conducting groundwater level prediction.

Attention mechanism was initially motivated by the way people paying different visual attention to different regions of an image. The idea of attention mechanism is to assign different attention weights to manage and quantify the interdependence between the input and output sequence (General Attention) or within the input sequence (Self-Attention). Attention mechanism has been a widely used in developing models for Natural Language Processing (NLP). In addition, the attention mechanism was incorporated into graph mining solutions to deal with the large, complex and noisy graphs. Lee et al. conducted a comprehensive survey of the literature on the emerging field of graph attention models [21]. In other research fields, attention mechanism has been applied to improve the performance of deep learning models. Guo et al. proposed a novel attention based spatial-temporal graph convolutional network model to model the dynamic spatial-temporal correlations of traffic data [22]. Experiments on two real-world datasets showed the better performance compared to the baselines. Li et al. developed a hybrid neural network model combining attention mechanism and a gated recurrent unit to make real-time predictions of open channel flow in coal mines [23]. Liang et al. proposed a multi-level attention-based Recurrent Neural Network (RNN) which considered observations from multiple sensors to predict time series data of a sensor [24]. Therefore, the application of attention mechanism in groundwater level prediction is supposed to improve the performance of models by considering spatial-temporal information and correlations.

In this paper, we propose a DL algorithm based on the spatial-temporal attention mechanism (ST-Att-LSTM) to conduct short-term and long-term groundwater level prediction by combining information from multiple observation boreholes. The main contributions are: (1) the proposal of a DL algorithm based on spatial-temporal attention mechanism to involve multiple data sources; (2) the short-term and long-term groundwater level prediction based on the proposed algorithm with more accurate results compared to state-of-the-art baselines; (3) a potential way of knowledge discovery for domain experts. The rest of the paper is organized as follows: the data sources are described in Section Ⅱ; Section Ⅲ depicted the structure of the proposed algorithm based on spatial-temporal attention mechanism. The experiments and results for evaluating the proposed algorithm is demonstrated in Section Ⅳ; the conclusion and future works are given in Section Ⅴ.

II. Study Area and Data Description

The data used in this study were all observed at the middle reaches of the Heihe River Basin (HRB) (38°38'N~39°53'N, 98°53'E~100°44'E; Fig. 1) which is located in the Hexi Corridor, northwestern China. The study area is a typical arid area with limited rainfall but strong evaporation which makes the groundwater more important. The study area is the primary water consumption area where the irrigated agriculture consumes most of the water supply [25]. The groundwater resource has been overexploited for agricultural, industrial, and domestic use. Groundwater level changes are affected by precipitation, evapotranspiration, surface water, agricultural irrigation. The annual and interannual variations of groundwater level have been summarized by [19]. The groundwater level was observed and collected from 1986 to 2008 by 42 observation boreholes (yellow and blue dots in Fig. 1) which were spatially distributed in the whole study area (~9016 km^2). However, only observations at 6 boreholes were used to construct the DL model for the sake of simplicity. Groundwater level observed at "22" borehole was selected as the prediction target. Groundwater levels observed at several spatially distributed boreholes (i.e., "Daman", "Yanuanzhangwan", "Banqiaodongliu", "Liaoquanwanzi" and "Luocheng") (yellow dots in Fig. 1) were selected to drive the algorithms. The data processing will be introduced in Section 4.

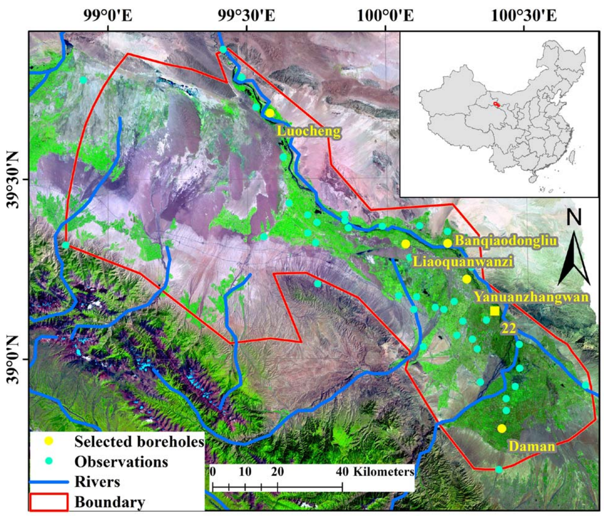

Fig. 1. Location of the middle reaches of the Heihe River Basin

III. Model Structure

In this paper, the structure of the proposed DL algorithm for groundwater level prediction is based on LSTM, sequence-to-sequence model (seq2seq) and attention mechanism. The purpose of the proposed algorithm is to involve spatial-temporal information while predicting the groundwater level. Spatial information is mainly considered in the influences of different observation boreholes on the target borehole. Temporal information is considered in the influences of the previous time steps on the predicted time step. Different from the conventional usage of attention mechanism which is only applied in the inputs of decoder to consider the temporal attention, the attention mechanism is also introduced in the inputs of encoder to consider the spatial attention between different observation boreholes.

A. LSTM

LSTM is a kind of Recurrent Neural Network (RNN) with recurrent units which is different from the Feedforward Neural Network (FNN). FNN with back propagation establishes weight connections between the inputs, layers and outputs. The outputs of FNN at time t only depends on the inputs at the same time step. Therefore, RNN was then proposed to overcome this limitation of temporal dependencies [26]. In RNN, the hidden layer's information is able to combine its own information at the previous time step through the interconnection between hidden layers. However, the problem of long-term dependencies prevents RNN from large scale applications. Therefore, Hochreiter and Schmidhuber [27] introduced Long Short-Term Memory neural networks by introducing a storage unit to enable the network to forget or retain the (state) information. Fig. 2 shows the structure of LSTM in which x_t refers to the input vector of time step t, h_t is the final output of the cell. From this structure, we can derive that h_t contains information from h_{t-1} and the input x_t. The information is then transmitted to the next time step iteratively through the interconnection which ensures the memorization of information.

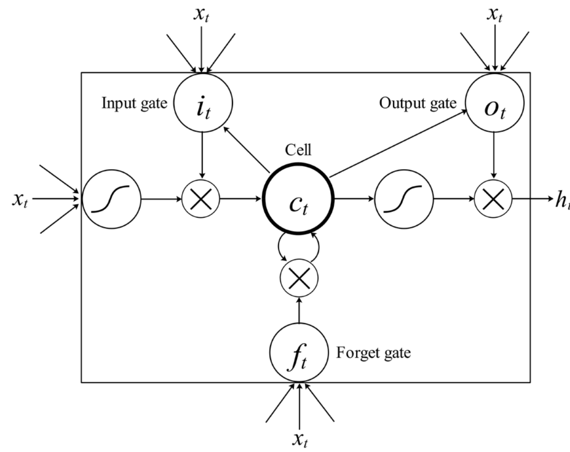

Fig. 2. Structure of LSTM

The difference between LSTM and RNN lies in the "processor" unit which is used to judge whether the information is useful for the current outputs or not. Three gates (input gate, forget gate and output gate) are contained in the processor units (also known as Cells). The information enters LSTM which complies with the algorithm's certification will be kept (remembered), otherwise, the information will be discarded (forgotten) [28]. Equations (1) ~ (6) describe the forward procedures that the LSTM cell maps the input sequence x to the hidden vector sequence h. In the equations, f_t, i_t, o_t, and C_t represent the output vectors of the forget gate, input gate, output gate and the memory cell, respectively; W_f, W_i, W_o, W_C, b_f, b_i, b_C and b_o are trainable parameters; σ and *tanh* are activation functions.

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

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

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

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

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

$$h_t = o_t * \tanh(C_t) \tag{6}$$

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

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

B. seq2seq model

RNN/LSTM requires fixed inputs and outputs dimension which is a serious limitation. Therefore, seq2seq model was proposed to solve the problem. The most distinguished contribution of seq2seq model is the inclusion of the encoder and decoder [29]. The basic idea of seq2seq model is that the input sequence are compressed to a specified length by the encoder. The most direct way of obtaining the hidden vector is to use the hidden state of the encoder. Besides, it is also possible to transform all hidden states of the input sequence to obtain hidden variables. The decoder is used to generate the specified results based on the hidden vector C at the current time step. A common used method is to regard the hidden vector C obtained by the encoder as the initial state to the decoder. The outputs will be used as the inputs of the current time step. The hidden vector C is used and only used as the initial state to participate in the operation.

C. Attention mechanism

The attention mechanism has recently been successfully applied in a wide range of tasks. LSTM uses the last hidden state or the mean of all hidden states as an output, while the attention mechanism allows for a more specific dependency between the states of the model at different time steps. According to the definition of [30], given a hidden state h_t, the attention-based model calculates a "context" vector c_t as a weighted average of the state sequence h. c_t is calculated as:

$$c_t = \sum_{j=1}^{T} \alpha_{tj} h_j \tag{9}$$

where T represents the total number of time steps; α_{tj} denote the weights computed for each state h_j.

c_t are then used to calculate state sequence s as:

$$s_t = f\left(s_{t-1}, y_{t-1}, c_t\right) \tag{10}$$

which indicates that s_t depends on s_{t-1}, c_t and the model outputs at $t-1$.

The weights α_{tj} are then calculated as:

$$e_{tj} = a(s_{t-1}, h_j) \tag{11}$$

$$\alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^{T} \exp(e_{tk})} \tag{12}$$

D. Spatial-temporal attention model for groundwater level prediction

The spatial-temporal dynamics of the groundwater level located in the spatial distribution of observation boreholes and the data dependencies through time. The sensors in observation boreholes record the groundwater level through time. In the perspective of geography, the groundwater level observed at different boreholes may have implicit relations which can be demonstrated as spatially connected. The groundwater level between different time steps may also have implicit connections. To this end, we propose a structure based on attention mechanism to capture this spatial-temporal dynamic relationship (refer to ST-Att-LSTM hereafter). The goal of the spatial attention is to involve the influence of the data from other observation boreholes on the target borehole at the same time step; the purpose of the temporal attention is to involve the influence of the data at previous times steps on the current time step.

In spatial attention, certain attention weights are assigned to observation boreholes according to the impact factors. The weight α_t^k in (1) measures the importance of the k-th observation borehole at time t. The inputs after weighting are:

$$\tilde{x}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \cdots \alpha_t^n x_t^n) \tag{13}$$

The weights α_t^k can be calculated from the hidden state and cell state of the encoder as:

$$e_t^k = v_e^t \, tanh(W_e[h_{t-1}; s_{t-1}] + U_e x_k) \tag{14}$$

$$\alpha_t^k = \frac{exp(e_t^k)}{\sum_{j=1}^{N} exp(e_t^j)} \tag{15}$$

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

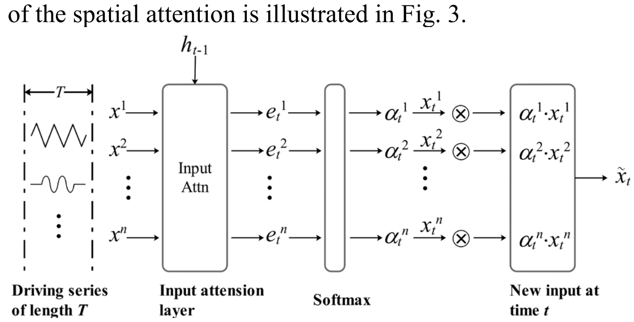

Fig. 3. Structure of spatial attention mechanism

The spatial attention mechanism updated the inputs by multiplying the attention weights with the inputs. Then, the encoder based on the LSTM encodes the updated inputs. Different decoding strategies are designed for the short-term and long-term prediction. In short-term prediction, the outputs of LSTM pass through the attention layer. The final outputs (y) are calculated as the weighting of the LSTM outputs. In (16) ~ (18), *LSTM* represents the calculation of a hidden layer unit using LSTM; *ATT* represents the calculation of the attention mechanism; c_t represents the context vector; FC represents a fully connected layer.

$$h_t = LSTM(x_t, h_{t-1}) \tag{16}$$

$$c_t = ATT(h_t, e_t) \tag{17}$$

$$y = FC(c_t) \tag{18}$$

In long-term prediction, the seq2seq model structure is implemented (LSTM as the decoder). Each time step of the decoding process obtains different encoding vectors through the temporal attention mechanism. In Equations (19) ~ (21), *decoder* represents the calculation of a hidden layer unit; c_t represents the context vector.

$$s_t = decoder(y_{t-1}, s_{t-1}) \tag{19}$$

$$\widehat{s_t} = tanh(W_c[s_t; c_t]) \tag{20}$$

$$y_t = FC(\widehat{s_t}) \tag{21}$$

The structure of spatial-temporal attention mechanism is shown in Fig. 4.

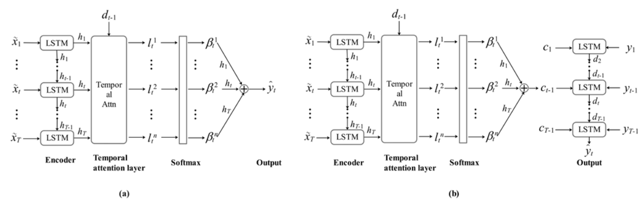

Fig. 4. Structure of spatial-temporal attention mechanism ((a) the short-term prediction model; (b) the long-term prediction model)

IV. EXPERIMENTS AND RESULTS

A. Raw data processing

The introduction of the middle reaches of HRB as well as the data are briefly described in Section Ⅱ. Fig.1 illustrates the location of the target observation borehole ("22") and auxiliary boreholes ("Daman", "Yanuanzhangwan", "Banqiaodongliu", "Liaoquanwanzi" and "Luocheng") used in our experiments. The datasets of groundwater levels from January 1986 to December 2008 were used and processed to monthly interval which was further divided into training set (90%) and validation set (10%). The purpose of the experiment is to predict the groundwater level at the target observation borehole "22" considering the information of the auxiliary boreholes. The groundwater level at previous time (refer to time window hereafter) are also used to obtain the groundwater level at the current/future time steps. The time window for the input is T=12 which indicates the input sequence is $\{x_t, x_{t+1}, x_{t+2}, \cdots x_{t+11}\}$. The output is x_{t+12} (with the time window being 1) for short-term prediction and $\{x_{t+11+1}, x_{t+11+2}, \cdots x_{t+11+\tau}\}$ (with the time window being τ) for the long-term prediction. τ is the length of the output sequence.

Because of the existence of different dimensions, the data was scaled to [0,1] as:

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{22}$$

where x_i represents the normalized data; x denotes the raw data; x_{min} and x_{max} are the minimum (i.e., 0) and maximum value (i.e., 1) of x.

B. Parameters and performance evaluation

Table I shows the software and hardware environments for conducting the experiment. Backpropagation with Adaptive moment estimation (Adam) [31] is used to train the algorithm. The training process is to minimize the loss function with respect to the parameters. The hyper-parameters of the algorithm and related baseline models will be depicted in the relevant sections.

TABLE I. 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 and calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{23}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{24}$$

where n stands for the number of samples; y_i refers to the observed groundwater level, $\hat{y}_i$ refers to the calculated groundwater level.

C. Short-term prediction

In the experiment of short-term prediction, the time window is set to 1 (τ=1 in Section Ⅳ A) to predict the groundwater level of the next month. Support Vector Regression (SVR), FNN, LSTM, and LSTM with spatial attention mechanism (S-Att-LSTM) are used as baseline models in order to evaluate the ST-Att-LSTM. The hyper-parameters of all the models are shown in Table II. The corresponding results are shown in Fig. 5 which indicate reasonable match for all the models. The performance of the LSTM-based models are better than SVR and FNN because of the inclusion of time dependencies which is a nature of LSTM. However, S-Att-LSTM performs better which may benefit from useful information extracted from different observation boreholes spatially distributed in the study area. The best performance is obtained by ST-Att-LSTM which may attributed to the spatial-temporal attention mechanism. The spatial-temporal attention mechanism is capable of considering both the spatial and temporal relations with attention weights. RMSE and MAE are used to conduct quantitative comparisons between different models (Table III).

ST-Att-LSTM performs the best among all the models which is in accordance with Fig. 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 the consideration of information through time. The S-Att-LSTM algorithm with spatial attention mechanism performs better than the conventional LSTM. The ST-Att-LSTM algorithm with spatial-temporal attention mechanisms performs best among three LSTM-based algorithms. Comparisons between LSTM-based algorithms indicates that the attention mechanism will further improve the performance of the algorithms.

TABLE II. 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

TABLE III. 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**

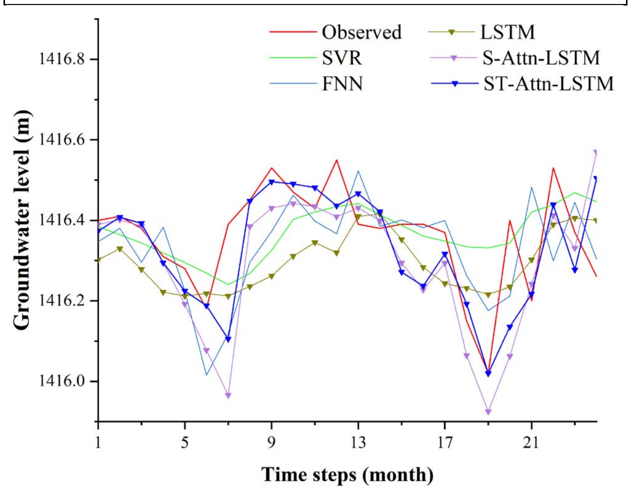

Fig. 5. Short-term prediction of groundwater level

Furthermore, the attention weights of ST-Att-LSTM are extracted to examined the effectiveness of spatial-temporal attention mechanism on combining the relations from different observation boreholes and different time steps (shown Fig. 6). The attention weights among different time steps indicates strong correlations between the current time step and the previous-two time steps. One can imagine that certain annual patterns exist in the groundwater level dynamics because of the routine groundwater pumping, agricultural irrigation, precipitation, upstream runoff and evapotranspiration. Therefore, the groundwater level at the target time step correlates with the groundwater level at the adjacent time steps the most. The spatial attention weights depicted in Fig. 6 (b) indicates that the groundwater level of the prediction target (observation borehole "22"), besides having the strongest correlation with its own historical data, is also highly relevant to the groundwater level of borehole "Banqiaodongliu". Compare Fig. 6 with Fig. 1, it is interest that the attention weights of borehole "Liaoquanwanzi" which is the nearest to the borehole "22" is the second lowest. The attention weights of the farthest observation borehole "Luocheng" is second highest among all the boreholes. The interesting results may offer some hints to the hydrologists when analyzing the hydraulic properties of the study area. Hydraulic connections between "22" and "Banqiaodongliu", fault between "22" and "Liaoquanwanzi" should be considered. The attention weights indicate that the spatial-temporal attention mechanism is effective and the spatial-temporal dynamic relationships between different observation boreholes and time steps are successfully extracted which will improve the prediction accuracy.

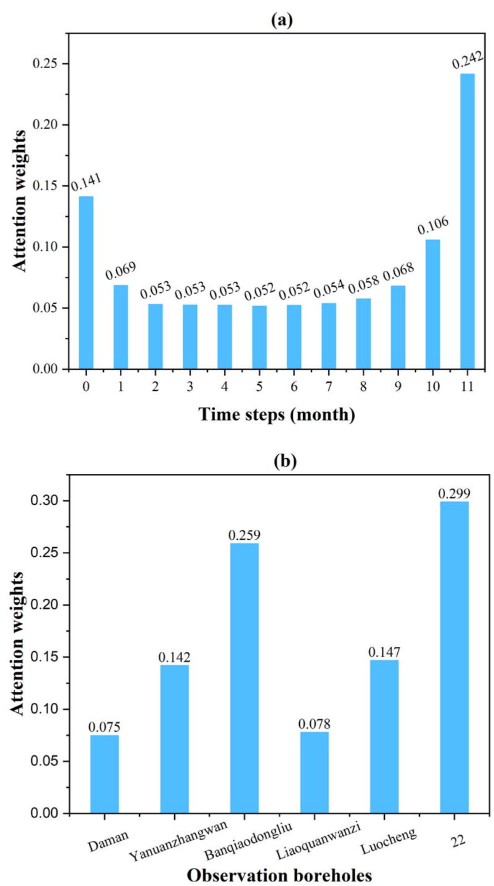

Fig. 6. Spatial-temporal attention weight: (a) attention weights among different time steps; (b) attention weights among different observation boreholes

D. Long-term prediction

In the long-term prediction experiment, the time window is set to 12 (τ=12 in Section Ⅳ A) to predict the groundwater level of the next year. Different baseline models (ARIMA, LSTM, seq2seq, seq2seq with time attention (seq2seq-Att)) are selected deliberately to evaluate ST-Att-LSTM. The same methods of pre-processing and training neural networks as in the short-term prediction experiment are used. The hyper-parameters of all the models are optimized and shown in Table IV. The results of all the models are shown in Fig. 7. One can conclude that the performance of conventional methods (ARIMA, LSTM) for long-term prediction are worse than those of 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 should is noted that a local minimum value around the 9th time step is mismatched by all the models. This may be acceptable considering the purpose of the long-term prediction being to capture the trend of the data (which has already been captured well) rather than some special value in case of overfitting.

MAE and RMSE are calculated to conduct qualitative comparisons between different models (Table V). The performance of ST-Att-LSTM model is the best among all the algorithms which is also in accord with Fig. 7. This experiments proved that the inclusion of spatial-temporal information by spatial-temporal attention mechanism is a successful attempt in the long-term prediction of groundwater level.

TABLE IV. 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

TABLE V. 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**

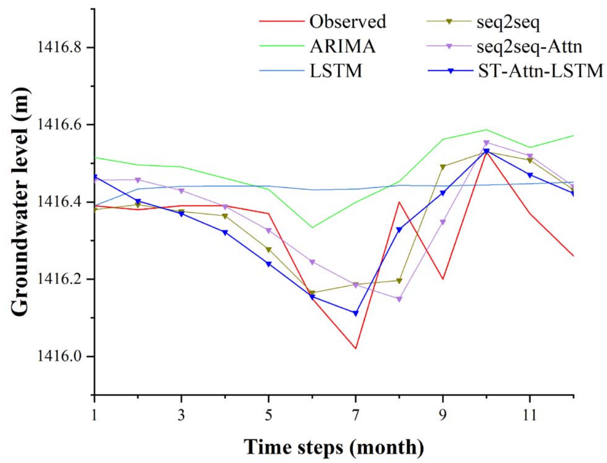

Fig. 7. Long-term prediction of groundwater level

V. CONCLUSION

In this paper, a deep learning algorithm for groundwater level prediction was proposed based on the spatial-temporal attention mechanism to explore the potential correlations in spatial-temporal distribution of observations. The groundwater level at six spatially distributed observation boreholes (i.e., “Daman”, “Yanuanzhangwan”, “Banqiaodongliu”, “Liaoquanwanzi”, “Luocheng” and “22”) in the middle reaches of the HRB were 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 of the proposed algorithm and several baseline algorithms. The results showed that the inclusion of spatial-temporal information by attention mechanisms was able to further improve the performance of deep learning algorithms. The attention weights among different spatial-distributed observation boreholes were examined which may provide information for potential hydraulic connections or faults hints for hydrologist. Therefore, the proposed algorithm provides (1) a potential way to fully utilize the spatial-temporal data in groundwater modeling; (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.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under grant No. 62006247, Science Foundation of China University of Petroleum-Beijing under grant No. 2462020YXZZ025 and 2462020XKJS03.

REFERENCES

[1] R. Q. Grafton, M. B. Ward, H. To, and T. Kompas, "Determinants of residential water consumption: Evidence and analysis from a 10-country household survey," Water Resour. Res., vol. 47, no. 8, 2011, doi: 10.1029/2010wr009685.

[2] Z. Wang, A. Wang, Y. Qiu and Q. Tan, Heavy rain in Henan (in Chinese), Health times, 2021.

[3] "Floods in Taklimakan Desert," (in Chinese) Retrieved August 2021, from https://www.163.com/dy/article/GG6060400519D4UH.html.

[4] E. D. Swain and E. J. Wexler, "A coupled surface-water and ground-water flow model (MODBRANCH) for simulation of stream-aquifer interaction," in "Techniques of Water-Resources Investigations," Report 06-A6, 1996. [Online]. Available: http://pubs.er.usgs.gov/publication/twri06A6

[5] J. D. Hughes, C. D. Langevin, and E. R. Banta, "Documentation for the MODFLOW 6 framework," in "Techniques and Methods," Reston, VA, Report 6-A57, 2017. [Online]. Available: http://pubs.er.usgs.gov/publication/tm6A57

[6] H.-J. Diersch, FEFLOW: Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media. Springer-Verlag Berlin Heidelberg, 2014, p. 996.

[7] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," (in English), Proceedings of the National Academy of Sciences of the United States of America, Article; Proceedings Paper vol. 99, pp. 7280-7287, May 2002, doi: 10.1073/pnas.082080899.

[8] O. David et al., "Object Modeling System v3.0: Developer and User Handbook," 2010.

[9] J. B. Gregersen, P. J. A. Gijsbers, and S. J. P. Westen, "OpenMI: Open modelling interface,", J Hydroinform, vol. 9, no. 3, pp. 175-191, Jul 2007, doi: 10.2166/hydro.2007.023.

[10] C. Guevara-Ochoa, A. Medina-Sierra, and L. Vives, "Spatio-temporal effect of climate change on water balance and interactions between groundwater and surface water in plains," Sci. Total Environ., vol. 722, p. 137886, June 2020, doi: https://doi.org/10.1016/j.scitotenv.2020.137886.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," , Nature., Review vol. 521, no. 7553, pp. 436-444, May 2015, doi: 10.1038/nature14539.

[12] C. Chen, W. He, H. Zhou, Y. Xue, and M. Zhu, "A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China," Scientific Reports, vol. 10, no. 1, p. 3904, March 2020, doi: 10.1038/s41598-020-60698-9.

[13] M. N. French, W. F. Krajewski, and R. R. Cuykendall, "Rainfall forecasting in space and time using a neural network," Journal of Hydrology, vol. 137, no. 1, pp. 1-31, August 1992, doi: https://doi.org/10.1016/0022-1694(92)90046-X.

[14] E. A. Coppola, Jr., A. J. Rana, M. M. Poulton, F. Szidarovszky, and V. W. Uhl, "A neural network model for predicting aquifer water level elevations," Groundwater, vol. 43, no. 2, pp. 231-41, Mar-April 2005, doi: 10.1111/j.1745-6584.2005.0003.x.

[15] V. Uddameri, "Using statistical and artificial neural network models to forecast potentiometric levels at a deep well in South Texas," Environmental Geology, vol. 51, no. 6, pp. 885-895, January 2007, doi: 10.1007/s00254-006-0452-5.

[16] J. Shiri, O. Kisi, H. Yoon, K.-K. Lee, and A. Hossein Nazemi, "Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques," Computers & Geosciences, vol. 56, pp. 32-44, July 2013, doi: https://doi.org/10.1016/j.cageo.2013.01.007.

[17] Y. C. Gong, Z. J. Wang, G. Y. Xu, and Z. X. Zhang, "A Comparative Study of Groundwater Level Forecasting Using Data-Driven Models Based on Ensemble Empirical Mode Decomposition," (in English), Water, Article vol. 10, no. 6, p. 20, Jun 2018, Art no. 730, doi: 10.3390/w10060730.

[18] B. Yan, J. Sun, X. Wang, N. Han and B. Liu, “Multivariable LSTM Neural Network Model for Groundwater Levels Prediction,” (in Chinese) Journal of Jilin University (Earth Science Edition), vol. 50, no. 1, pp. 208-216, January 2020, doi: 10.13278/j.cnki.jjuese.20190055.

[19] J. B. Mohapatra, P. Y. Jha, M. K. Jha, and S. Biswal, "Efficacy of machine learning techniques in predicting groundwater fluctuations in agro-ecological zones of India," Sci. Total Environ., Article vol. 785, p. 14, September 2021, no. 147319, doi: 10.1016/j.scitotenv.2021.147319.

[20] P. Du et al., "Advances of Four Machine Learning Methods for Spatial Data Handling: a Review," Journal of Geovisualization and Spatial Analysis, vol. 4, no. 1, p. 13, May 2020, doi: 10.1007/s41651-020-00048-5.

[21] J. Lee, R. A. Rossi, S. Kim, N. Ahmed, and E. Koh, "Attention Models in Graphs," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 13, pp. 1 - 25, 2019.

[22] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 922-929, July 2019, doi: 10.1609/aaai.v33i01.3301922.

[23] Z. Li, T. Gao, C. Guo, and H.-A. Li, "A Gated Recurrent Unit Network Model for Predicting Open Channel Flow in Coal Mines Based on Attention Mechanisms," Ieee Access, vol. 8, pp. 119819-119828, 2020, doi: 10.1109/access.2020.3004624.

[24] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, GeoMAN: Multi-level Attention Networks for Geo-sensory Time Series Prediction. 2018, pp. 3428-3434.

[25] F. Wu, J. Zhan, and İ. Güneralp, "Present and future of urban water balance in the rapidly urbanizing Heihe River Basin, Northwest China," Ecol. Model., vol. 318, pp. 254-264, December 2015, doi: https://doi.org/10.1016/j.ecolmodel.2014.11.032.

[26] S. E. Hihi and Y. Bengio, "Hierarchical Recurrent Neural Networks for Long-Term Dependencies," in neural information processing systems, 1995, pp. 493-499.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[28] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," Transportation Research Part C: Emerging Technologies, vol. 85, pp. 591-608, December 2017, doi: https://doi.org/10.1016/j.trc.2017.10.016.

[29] A. Zaytar and C. E. Amrani, "Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks," International Journal of Computer Applications, vol. 143, pp. 7-11, 2016.

[30] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," CoRR, vol. abs/1409.0473, 2015.

[31] X. Zhao and Z. Song, "Adam Optimized CNN Super-Resolution Reconstruction," Journal of Frontiers of Computer Science and Technology, vol. 13, no. 5, pp. 858-865, doi: 10.3778/j.issn.1673-9418.1809020.