Long Short-Term Memory (LSTM) for Water Table Depth Prediction in Agricultural Areas

Anoushka Vyas (20171057)

***Abstract*—Predicting water table depth for long-term in agricultural areas presents great challenges because of complex and heterogeneous hydrogeological characteristics, boundary conditions, and human activities; also, nonlinear interactions occur among these factors. Therefore, a time series model based on Long Short-Term Memory (LSTM), is used as an alternative to computationally expensive physical models. The model is composed of an LSTM layer with another fully connected layer on top of it, with a dropout method applied in the first LSTM layer. In this study, the model was applied and evaluated in the area of Hetao Irrigation District in arid northwestern China using data of 14 years (2000-2013). The results of the model were compared with the other time series models like Feed-Forward Neural Network (FFNN), Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU). The effect of dropout layer is also studied as part of the experiments. The model's architecture is reasonable and can contribute to a strong learning ability on time series data. Thus, one can conclude that the model can serve as an alternative approach in predicting water table depth, especially in areas where hydrogeological data are difficult to obtain.**

***Index Terms*—Machine learning model, Water table depth, Deep learning, LSTM, Hetao Irrigation District**

I. Introduction

Groundwater provides an important source of water for domestic, agricultural, and industrial use. However, groundwater resources are vulnerable to over-exploitation, climate change, and biochemical pollution (Bouwer, 2000; Sophocleous, 2005; Evans and Sadler, 2008; White and Falkland, 2010; Karandish et al., 2015). As a result, many areas over the world face groundwater shortages. An example of such areas is the Hetao Irrigation District, one of the largest irrigation districts in China that is located in the arid area of the Yellow River watershed. The Yellow River serves as the main source of irrigation water in this district. However, the availability of irrigation water from the Yellow River has decreased dramatically, with intensified water resource use in the Yellow River watershed (Yang et al., 2003). Therefore, groundwater has become an important source of supplementary irrigation water in Hetao Irrigation District. Sustainable groundwater planning and management requires accurate forecasting of water table depth (Wang et al., 2014). An accurate and reliable assessment of water table depth can help engineers and decision-makers to: develop optimal water resource allocation strategies, adjust crop patterns in different sub-irrigation areas, and develop optimal irrigation schedules while controlling the effects of salinity related to intensive irrigation (Xu et al., 2010). The objective of this study is to develop an effective and accurate method for predicting water table depth that can be used to help engineers and decision-makers manage groundwater resources and make management decisions.

In the last two decades, predictions of water table depth have usually been made by using physical models such as MODFLOW and HYDRUS (e.g., Pang et al., 2000; Batelaan et al., 2003; Zume and Tarhule, 2008; Faulkner et al., 2009; Xu et al., 2012). However, those models have many practical limitations because they are always data demanding and time consuming during model development and calibration. Using physical models to predict water table depth in Hetao Irrigation District is particularly challenging because the district covers a large area, lacks abundant hydrogeological data, and has strong spatial and seasonal variability in freeze-thaw periods.

In the last decade, many studies have investigated the advantages and disadvantages of various physical models and evaluated their prediction performance with that of emerging data-driven, machine learning methods. The machine learning methods include Multiple Linear Regression (MLR) (Sahoo and Jha, 2013), Support Vector Regression (SVR) (Yu et al., 2006; Yoon et al., 2011; Belayneh et al., 2014; Mirzavand, 2015), and Artificial Neural Networks (ANN) (Raman and Sunilkumar, 1995; Daliakopoulos, 2005; Sarangi et al., 2006; Napolitano et al., 2011; Parchami-Araghi et al., 2013; Seo et al., 2015; Chang et al., 2016). These statistical methods explore the spatial and temporal patterns hidden in historical data without using a physical model because the latter always requires a large number of physical parameters and a deep understanding of the physical processes of the modeling domain. In many cases, machine learning methods can achieve a better predictive performance than physically based models (Coppola et al., 2003; Parkin et al., 2007; Chu and Chang, 2009; Mohanty et al., 2013; Karandish and Šimunek, 2016). For example, Mohanty et al., (2013) found that an ANN model provided a better prediction of water table depth than MODFLOW for short-term predictions. Karandish and Šimunek (2016) found that both Adaptive Neuro-Fuzzy Inference System and SVM models performed well when compared with HYDRUS-2D for water stressed conditions. Therefore, machine learning methods provide promising tools for predicting water table depth.

The present study focuses on a time series prediction task, so recurrent neural network (RNN) (Rumelhart et al., 1986) is a suitable choice. A RNN model has internal self-looped cells, allowing the RNN to "remember" time series information and making it adept at performing time series tasks. In addition, in this study, LSTM, a special kind of RNN, that works well in processing long term time series data, was chosen due to its sophisticated network structure. Compared with aforementioned advanced FFNN model, the proposed LSTM-based model only applied a very simple data pre-processing method. LSTM is a

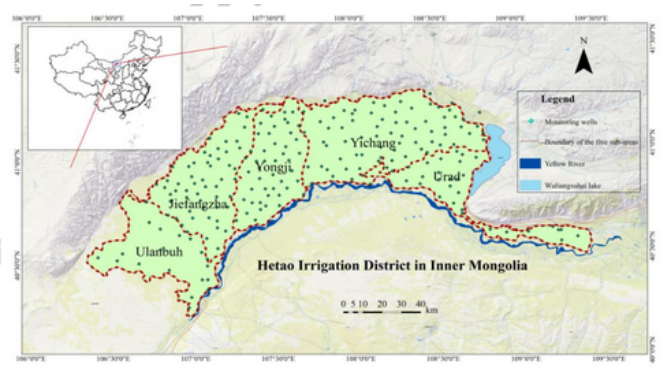

Fig. 1. Location of Hetao district and the obervation wells.

famous deep learning model. It is recurrent, where connections between units form a directed cycle allowing data to flow both forwards and backwards within the network; then the previous information can be preserved for future use.

The goal of this study was to develop a two-layer LSTM-based model for predicting water table depth in Hetao Irrigation District. The model contained one layer of LSTM and a fully connected layer on top of the LSTM layer. The model employed monthly water diversion, evaporation, precipitation, temperature and time as input variables for predicting water table depth in the district. The rest of the paper is organized as follows. In Section 2, the study area and observational data are introduced. Section 3 illustrates RNN, LSTM, GRU, dropout regularization, the proposed model's architecture, and model evaluation criteria. Section 4 presents the experiment results and discussion. Finally, Section 5 concludes the paper.

II. Dataset

A. Area of study

Hetao Irrigation District lies within Bayannaoer City, Inner Mongolia, China (106°20'–109°29' N, 40°14'–41°18' E; elevation 1020–1050 m a.s.l.). Mean annual precipitation is 169.4 mm, about 70% of which falls during June-September; the maximum precipitation occurs in August. The mean, minimum and maximum temperatures were 3.9°C. -14.6°C, and 28.4°C, respectively (China Meteorological Data Service Center, 1954-2013). Hetao Irrigation District covers 11,073 km^2; the central part of the district is about 180 km long and 60 km wide. About 52% (5740 km^2) of the entire terrestrial area is irrigated. A total of 227 monitoring wells are used in the district (Fig. 1) with the water table depth measured every 5 days. Because monthly data are used in the model, the measured water table depth is averaged by month.

B. Dataset description

The experiments used 14 years (2000–2013) of time series data from the Hetao Irrigation District. Six variables were chosen for this study: water diversion, precipitation, evaporation volume, temperature, time, and water table depth. The first 12 years of time series data were used as a training set, and the next two years of data was used as test set.

C. Data pre-processing

All six of the variables were standardized to ensure they remain on the same scale. This pre-processing can guarantee a stable convergence of parameters in the model developed in the present study. The standardization formula is as follows:

$$x_{ij}^{(new)} = \frac{x_{ij} - \overline{x}_i}{\sigma_i} \tag{1}$$

where x_{ij} represents data in i^{th} year, j^{th} month; x_i and σ_i are the average and standard deviation of data in the i^{th} year, respectively.

III. Model Description

Before introducing LSTM, understanding RNN is important because LSTM is a special kind of RNN. RNNs were first developed in the 1980s; these networks have connections between neurons and form a directed cycle. This type of structure creates an internal self-looped cell, which allows it to display dynamic temporal behavior. RNNs have chain-like structures of repeating modules (Fig. 2(a)). These structures can help RNNs to "remember" previous information, which allows the RNNs to process arbitrary (long time) sequences. Therefore, RNNs have been successful in learning sequences.

Forward propagation begins with a specific initial state $h_0 = 0$. For each time step from $t = 1$ to $t = \tau$, the following update equations were applied:

$$h_t = tanh(b_h + Wh_{t-1} + Ux_t) \tag{2}$$

$$o_t = b_o + Vh_t \tag{3}$$

where x_t is the input vector at time t and h_{t-1} is the hidden cell state at time $t-1$. The parameters are the bias vector b_h and b_o, as well as the weight matrices U, W and V for input-to-hidden, hidden-to-hidden and hidden-to-output connections, respectively.

A. Long Short-Term Memory Network (LSTM)

The gradients of the RNNs can be computed via Back-Propagation Through Time (Rumelhart et al., 1986; Werbos, 1990). However, Back-Propagation Through Time is not sufficiently efficient to learn a pattern from long term dependency because of a gradient vanishing problem (Hochreiter, 1998). This problem can be solved by the structure of LSTMs (Hochreiter and Schmidhuber, 1997; Graves, 2012;Jozefowicz et al., 2015). Like RNNs, LSTMs also have chain like modules, but the repeating modules have more complicated structures. Each repeating module of LSTMs contain a memory block. This memory block is specifically designed to store information over long time periods.

The memory block contains four parts: a CEC (the Constant Error Carousel) cell in addition to three special multiplicative units called gates. The CEC cell runs straight down the entire chain without any activation function and thus the gradient does not vanish when Back-Propagation Through Time is applied to train a LSTM. Therefore, LSTMs have been shown to learn long-term dependencies more easily than

RNNs because information can easily flow along the cells unchanged. Furthermore, the input, forget (Gers et al., 2000) and output gates, in each memory block can control the flow of information inside memory block. The input, forget, and output gates control the extent to which new input flows into a CEC cell, information is stored in a cell, and output flows of the cell into the rest of the networks, respectively. A schematic of memory block is shown in Fig. 2(b). It includes block input, three gates (input, forget, output), and a CEC cell.

Similar to RNNs, LSTMs computes a mapping from an input sequence x to an output sequence y by calculating the network unit activations using the following equations iteratively from $t = 1$ to $t = \tau$ with initial values $C_0 = 0$ and $h_0 = 0$:

$$i_t = \sigma(b_i + W_i x_t + U_i h_{t-1}) \tag{4}$$

$$f_t = \sigma(b_f + W_f x_t + U_f h_{t-1}) \tag{5}$$

$$o_t = \sigma(b_o + W_o x_t + U_o h_{t-1}) \tag{6}$$

$$\tilde{C}_t = tanh(b_c + W_c x_t + U_c h_{t-1}) \tag{7}$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \tag{8}$$

$$h_t = o_t \otimes tanh(C_t) \tag{9}$$

where W_i, W_f, and W_o denote the matrix of weights from the input, forget, and output gates to the input, respectively. Similarly, U_i, U_f, and U_o denote the matrix of weights from the input, forget, and output gates to the hidden, respectively. b_i, b_f, b_o denote the input, forget, and output gate bias vectors, respectively. σ is an element-wise non-linear activation function: logistic sigmoid. i_t, f_t, o_t and C_t are the input, forget, output gates and the cell state vectors at time t, respectively, all of which are the same size as the cell output vector h_t. The element-wise multiplication of two vectors is denoted with $\otimes$.

B. Overall Model

The input data is first put into the LSTM layer. The input gate of the LSTM layer will recompose input data and decide which input data is important; this process is similar to principal component analysis (PCA). However, the LSTM layer can preserve previous information, which can help to improve the ability of the model to learn time series data. A fully connected layer is set atop the LSTM layer in order to improve the model's learning ability. Moreover, dropout is set on the LSTM layer to prevent overfitting. The proposed model is illustrated in Fig. 3. The loss function is defined below:

$$L = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N} \tag{10}$$

where y_i is measured value at time i; and $\hat{y}_i$ is predicted value at time i. The flowchart of the proposed model framework is displayed in Fig. 4.

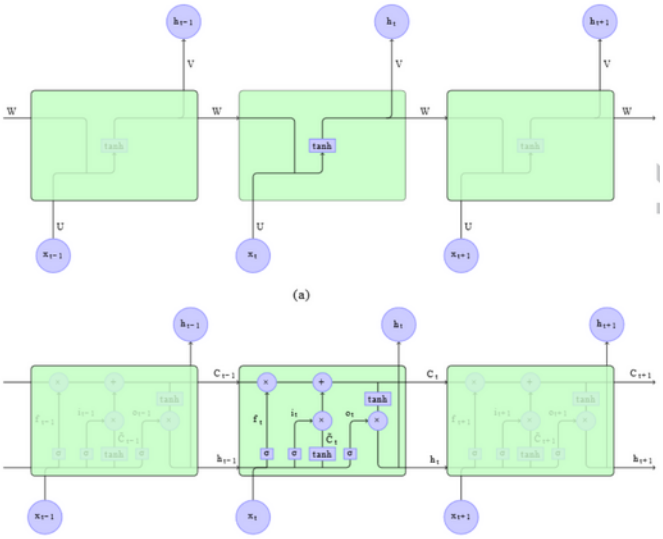

Fig. 2. (a) Chain like structure of the recurrent neural network. The self-connected hidden units allow information to be passed from one step to the next. (b) A graphical representation of LSTM's memory block.

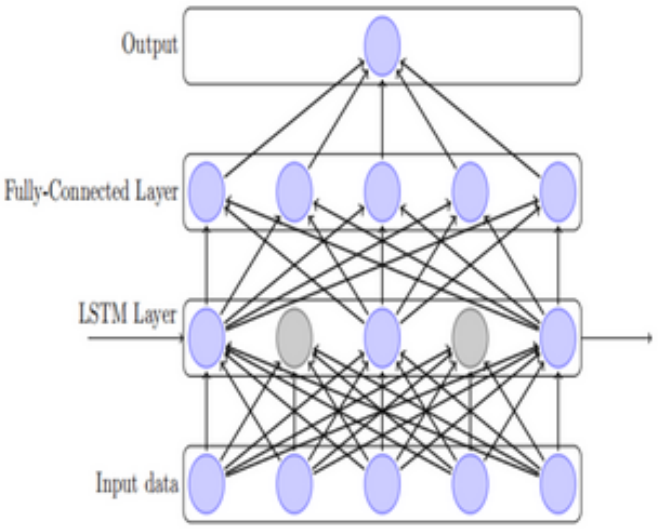

Fig. 3. Structure of the proposed model. Dropout has been applied at the Long Short-Term Memory (LSTM) layer. Grey units have been dropped.

C. Evaluation Metrics

Root mean square error ($RMSE$), and coefficient of determination (R^2) are used to evaluate the model's accuracy between the measured and predicted values. R^2 measures the degree of how well the outcomes are replicated by the model, ranging between $[-\infty, 1]$ where for optimal model prediction an R^2 score close to 1 is preferred. The R^2 equation is as follows,

$$R^2 = \frac{\sum_{i=1}^{N}(y_i - \overline{y})^2 - \sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{11}$$

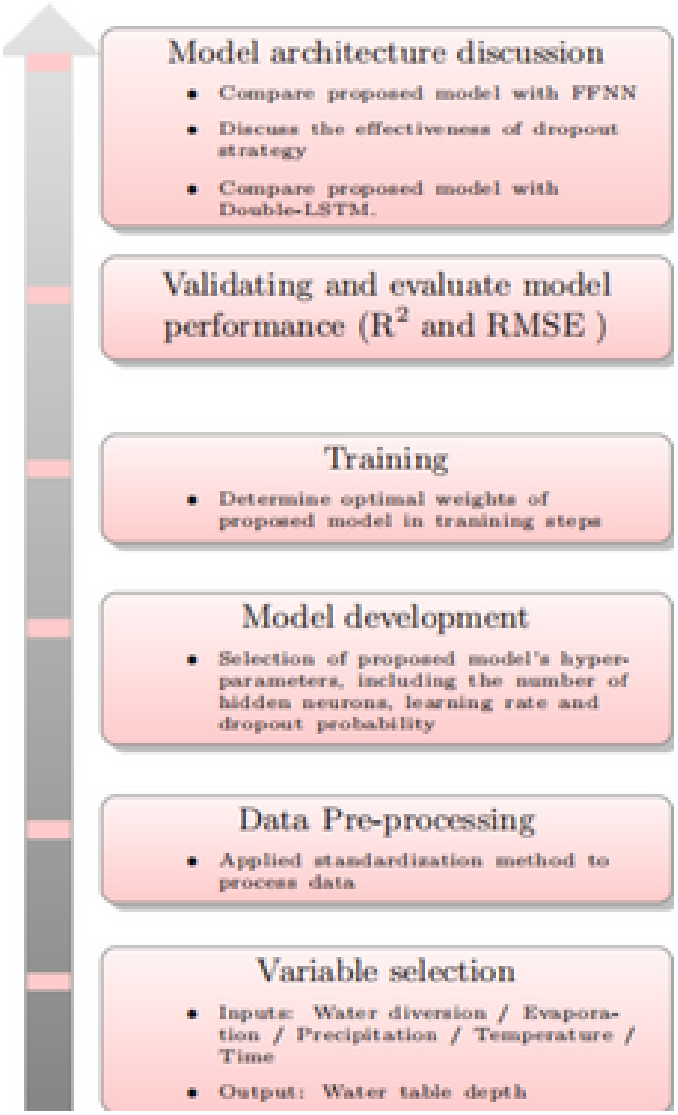

Fig. 4. Flowchart of this study.

where y_i is measured value at time i, $\overline{y}$ is mean of y_i, $(i = 1, \cdots, N)$ and $\hat{y_i}$ is predicted value at time i.

$RMSE$ scores range between $[0, \infty]$, and the model prediction is ideal if $RMSE$ is 0. $RMSE$ is defined as,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y_i})^2}{N}} \tag{12}$$

IV. Results

In order to observe the changes in water table depth in the entire Hetao Irrigation District, the proposed model also has been evaluated in the entire area, whose temperature was the average of five sub-areas and precipitation, water diversion and evaporation volume were sum of five sub-areas.

In order to illustrate how hyper-parameters affect the results, different hyper-parameters were set for the proposed model and parameters are shown in TABLE I. The optimal hyper-parameters of the proposed model used for water table depth prediction are shown in TABLE I (first row).

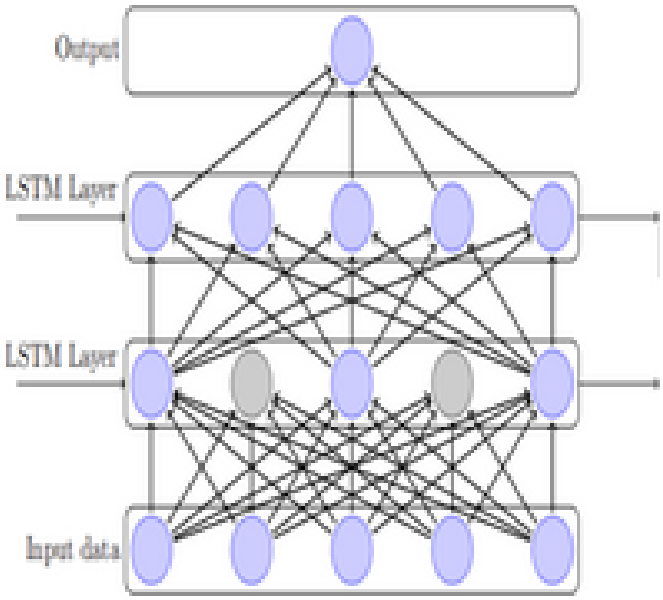

Fig. 5. Structure of the Double-LSTM model. Dropout has been applied at the first LSTM layer. Grey units have been dropped.

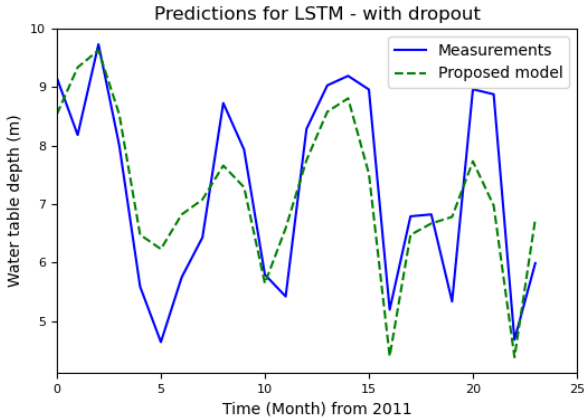

Fig. 6. Prediction of water table depth using LSTM with dropout.

Other models like RNN, FFNN and GRU were trained using the same hyperparameters and dropout. The Double-LSTM model is also used to do model ablation study. It is very similar to the main model; only change the fully connected layer to LSTM layer in Double-LSTM model as shown in Fig. 5. Effect of dropout is also observed in LSTM. From TABLE II and Fig. 6-11, it is evident that LSTM with dropout is performing the best. LSTM performs better than GRU because it able to keep track of history of long sequences.

V. Conclusion

In this study, a framework composed by one LSTM layer and one fully connected layer is used for predicting water table

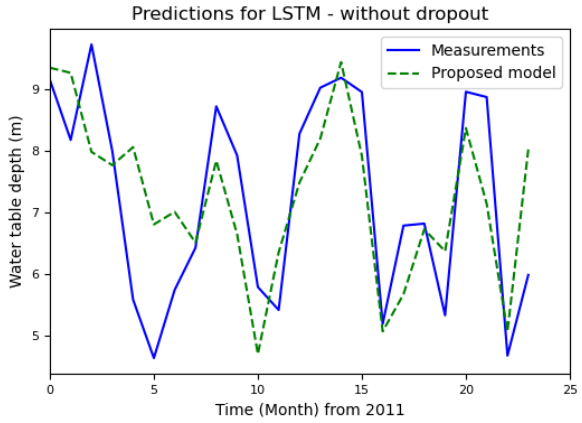

Fig. 7. Prediction of water table depth using LSTM with dropout.

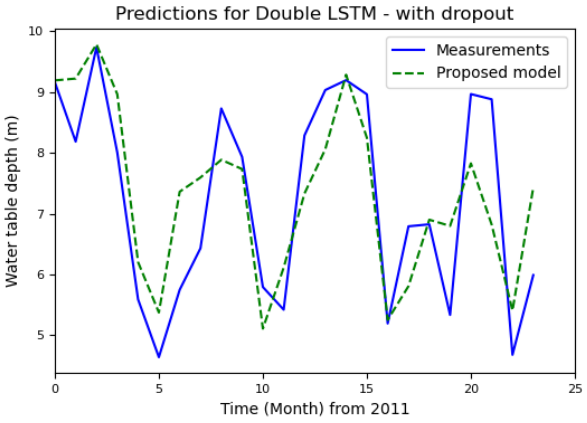

Fig. 8. Prediction of water table depth using Double LSTM.

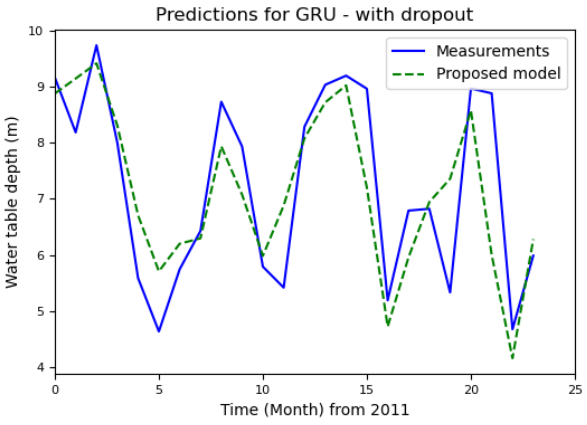

Fig. 9. Prediction of water table depth using GRU.

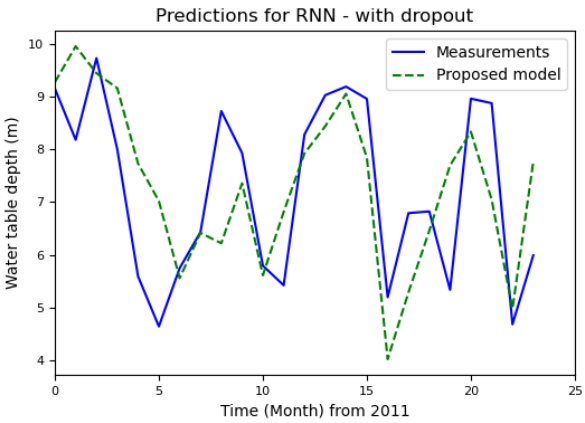

Fig. 10. Prediction of water table depth using Double RNN.

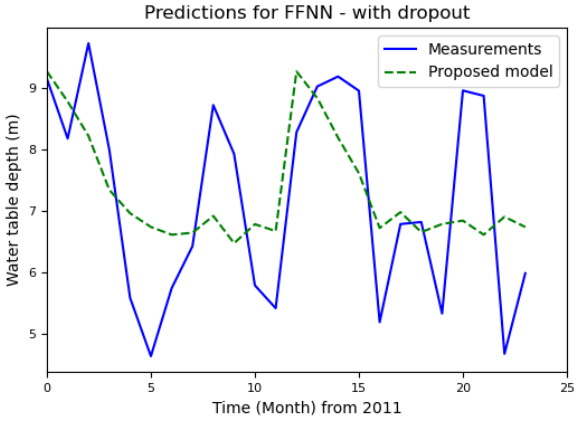

Fig. 11. Prediction of water table depth using FFNN.

Neurons	Learning rate	Dropout	Iterations
40	**10^{-4}**	**0.5**	**20000**
40	10^{-4}	0.5	40000
40	10^{-4}	0.5	5000
40	10^{-4}	0	20000
40	10^{-4}	0.8	20000
40	10^{-3}	0.5	20000
40	10^{-5}	0.5	20000
70	10^{-4}	0.5	20000
10	10^{-4}	0.5	20000

TABLE I
DIFFERENT HYPERPARAMETER SETTINGS

Models	MSE	RMSE	R^2
LSTM	**0.884**	**0.94**	**0.841**
LSTM w/o dropout	1.4	1.18	0.48
Double LSTM	0.925	0.962	0.657
GRU	1.02	1	0.623
RNN	1.73	1.31	0.359
FFNN	1.74	1.31	0.358

TABLE II
EVALUATION METRIC VALUES

depth, in order to help engineers and decision-makers to plan and manage groundwater resources in agricultural areas.

The architecture of the proposed model is reasonable. The LSTM layer helps to maintain previous information and contributes to learning the time series data. The dropout method helps to prevent overfitting during the training process. A fully connected layer atop the LSTM layer helps to improve the learning and fitting ability of the model.

REFERENCES

Bouwer, H., 2000. Integrated water management: emerging issues and challenges. Agric. Water Manage., 45, 217-228.

Sophocleous, M., 2005. Groundwater recharge and sustainability in the High Plains aquifer in Kansas, USA. Hydrogeol. J., 13, 351-365.

Evans, R. G., Sadler, E. J., 2008. Methods and technologies to improve efficiency of water use. Water Resour. Res., 44, W00E04, doi:10.1029/2007WR006200.

White, I., Falkland, T., 2010. Management of freshwater lenses on small Pacific Islands. Hydrogeol. J., 18, 227-246.

Karandish, F., Salari, S., Darzi-Naftchali, A., 2015. Application of virtual water trade to evaluate cropping pattern in arid regions. Water Resour. Manage., 29, 4061-4074.

Yang, L., Shen, R., Cao, X., 2003. Scheme of groundwater use in Hetao Irrigation District in Inner Mongolia. Trans. of the CSAE. 19, 56-59.

Wang, W., Yu, Z., Zhang, W., Shao, Q., Zhang, Y., Luo, Y., Jiao, X., Xu, J., 2014. Responses of rice yield, irrigation water requirement and water use efficiency to climate change in China: Historical simulation and future projections. Agric. Water Manage., 146, 249-261.

Xu, X., Huang, G., Qu, Z., Pereira, L. S., 2010. Assessing the groundwater dynamics and impacts of water saving in the Hetao Irrigation District, Yello River basin. Agric. Water Manage., 98, 301-313.

Pang, L. P., Close, M. E., Watt, J. P. C., Vincent, K. W., 2000. Simulation of picloram, atrazine, and simazine leaching through two New Zealand soils and into groundwater using HYDRUS-2D. J. Contam. Hydrol., 44, 19-46.

Batelaan, O., Smedt, F. D., Triest, L., 2003. Regional groundwater discharge: phreatophyte mapping, groundwater modelling and impact analysis of land-use change. J. Hydrol., 275, 86-108.

Zume, J., Tarhule, A., 2008. Simulating the impacts of groundwater pumping on stream-aquifer dynamics in semiarid northwestern Oklahoma, USA. Hydrogeol. J., 16, 797-810.

Xu, X., Huang, G., Zhan, H., Qu, Z., Huang, Q., 2012. Integration of SWAP and MODFLOW-2000 for modeling groundwater dynamics in shallow water table areas. J. Hydrol., 412-413, 170-181.

Sahoo, S., Jha, M. K., 2013. Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. Hydrogeol. J., 21, 1865-1887.

u, P. S., Chen, S. T., Chang, I. F., 2006. Support vector regression for real-time flood stage forecasting. J. Hydrol., 328, 704-716.

Yoon, H., Jun, S. C., Hyun, Y., Bae, G. O., Lee, K. K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J. Hydrol., 396, 128-138.

Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. J. Hydrol., 508, 418-429.

Mirzavand, M., Ghazavi, R., 2015. A Stochastic Modelling Technique for Groundwater Level Forecasting in an Arid Environment Using Time Series Methods. Water Resour. Manage., 29, 1315-1328.

Raman, H. Sunilkumar, N., 1995. Multivariate modelling of water resources time series using artificial neural networks. Hydrolog. SCI. J., 40, 145-163.

Daliakopoulos, I. N., Coulibaly, P., Tsanis, I. K., 2005. Groundwater level forecasting using artificial neural networks. J. Hydrol.,309, 229-240.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput., 9, 1735-1780. Jozefowicz, R., Zaremba, W., Sutskever, I., 2015. An empirical exploration of recurrent network architectures. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France, pp. 2342-2350.

Graves, A., 2012. Supervised Sequence Labelling with Recurrent Neural Networks. Springer-Verlag Berlin, Heidelberg.