



# Engineering Applications of Computational Fluid Mechanics

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/tcfm20>

## Time series-based groundwater level forecasting using gated recurrent unit deep neural networks

Haiping Lin, Amin Gharehbaghi, Qian Zhang, Shahab S. Band, Hao Ting Pai, Kwok-Wing Chau & Amir Mosavi

To cite this article: Haiping Lin, Amin Gharehbaghi, Qian Zhang, Shahab S. Band, Hao Ting Pai, Kwok-Wing Chau & Amir Mosavi (2022) Time series-based groundwater level forecasting using gated recurrent unit deep neural networks, Engineering Applications of Computational Fluid Mechanics, 16:1, 1655-1672, DOI: [10.1080/19942060.2022.2104928](https://doi.org/10.1080/19942060.2022.2104928)

To link to this article: <https://doi.org/10.1080/19942060.2022.2104928>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 16 Aug 2022.



Submit your article to this journal [↗](#)



Article views: 824





View related articles [↗](#)



View Crossmark data [↗](#)

# Time series-based groundwater level forecasting using gated recurrent unit deep neural networks

Haiping Lin<sup>a</sup>, Amin Gharehbaghi <sup>b</sup>, Qian Zhang<sup>c</sup>, Shahab S. Band<sup>d</sup>, Hao Ting Pai<sup>e</sup>, Kwok-Wing Chau <sup>f</sup> and Amir Mosavi<sup>g,h</sup>

<sup>a</sup>College of Information Engineering, Hangzhou Vocational & Technical College, Hangzhou, People's Republic of China; <sup>b</sup>Faculty of Engineering, Department of Civil Engineering, Hasan Kalyoncu University, Şahinbey, Turkey; <sup>c</sup>School of Data Science and Artificial Intelligence, Wenzhou University of Technology, Wenzhou, People's Republic of China; <sup>d</sup>Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Yunlin, Taiwan; <sup>e</sup>International Graduate Institute of Artificial Intelligence, National Yunlin University of Science and Technology, Yunlin, Taiwan; <sup>f</sup>Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hong Kong, People's Republic of China; <sup>g</sup>John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary; <sup>h</sup>Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia

## ABSTRACT

In this research, the mean monthly groundwater level with a range of 3.78 m in Qoşaçay plain, Iran, is forecast. Regarding three different layers of gated recurrent unit (GRU) structures and a hybrid of variational mode decomposition with gated recurrent unit (VMD-GRU), deep learning-based neural network models are developed. As the base model for performance comparison, the general single-long short-term memory-layer network model is developed. In all models, the module of sequence-to-one is used because of the lack of meteorological variables recorded in the study area. For modeling, 216 monthly datasets of the mean monthly water table depth of 33 different monitoring piezometers in the period April 2002–March 2020 are utilized. To boost the performance of the models and reduce the overfitting problem, an algorithm tuning process using different types of hyperparameter accompanied by a trial-and-error procedure is applied. Based on performance evaluation metrics, the total learnable parameters value and especially the model grading process, the new double-GRU model coupled with multiplication layer ( $\times$ ) (GRU2 $\times$  model) is chosen as the best model. Under the optimal hyperparameters, the GRU2 $\times$  model results in an  $R^2$  of 0.86, a root mean square error (RMSE) of 0.18 m, a corrected Akaike's information criterion (AICc) of  $-280.75$ , a running time for model training of 87 s and a total grade (TG) of 6.21 in the validation stage; and the hybrid VMD-GRU model yields an RMSE of 0.16 m, an  $R^2$  of 0.92, an AICc of  $-310.52$ , a running time of 185 s and a TG of 3.34.

## ARTICLE HISTORY

Received 22 November 2021  
Accepted 19 July 2022

## KEYWORDS

Groundwater level; deep neural network; machine learning; gated recurrent unit; artificial intelligence

## 1. Introduction

In groundwater hydrology, the hydraulic head is regarded as one of the most significant criteria, which has been chiefly evaluated as groundwater level (GWL) with natural and unnatural bases. It determines the potentiometric condition of water recharge, discharge and storage of aquifers (Conlon et al., 2005). Continuous monitoring of GWL over time provides a valuable information source for hydrologists to understand the evolution and geometry of aquifers, particularly spatial hydraulic heterogeneity and anisotropy, by analyzing diverse spatiotemporal attributes, including the hydrometeorological cycle and groundwater resources development under global long-standing climate unpredictability and human-caused influences, and applying rational policies on logical water

resources management (Asadi et al., 2020; Bekesi et al., 2009; Valdes et al., 2014).

Assessing long-term groundwater resources in a potential agronomic site requires specific knowledge, high expenditure on the installation of observation piezometers, and the gathering of diverse data, such as hydrogeological and meteorological parameters, over time. It is unreliable to estimate GWL through conventional linear and deterministic statistical models since GWL is neither a continuous stationary multifactor hydrological time series nor a recurrent physical procedure with a lagged response (Post & von Asmuth, 2013; Schuurmans et al., 2007). Consequently, it is vital to develop robust approaches that take into consideration the dynamic behavior of observations such as

**CONTACT** Qian Zhang  20200420@wzu.edu.cn; Shahab S. Band  shamshirbands@yuntech.edu.tw; Amir Mosavi  amir.mosavi@kvk.uni-obuda.hu

heteroscedasticity, non-stationarity and autoregression (Fathian et al., 2016). It can thus be inferred that estimating monthly scale GWL for irrigation purposes is often thought of as a burdensome task compared to surface water computations (Favero et al., 2007).

In the past few decades, different methods have been developed to assess GWL fluctuations, which are generally divided into four subgroups: physical, statistical, artificial intelligence (AI) and hybrid methods. Because of the extreme randomness of GWL fluctuations, it is difficult to determine which model has the best estimation performance. Each model has its strengths and weaknesses, and many models are site dependent. Hence, it is hard to forecast all GWL fluctuation series with a definite model. Since the late 1960s, numerous studies have been carried out on the numerical emulation of groundwater quality and quantity (e.g. solute and contaminant transport) in porous media by different methods (Anderson et al., 1992; Dalkılıç & Gharehbaghi, 2021; Gharehbaghi, 2021a, 2021b; Henriksen et al., 2008). This research contemplates flexible and reliable techniques for solving groundwater flow problems in complicated field conditions (Hemker & Bakker, 2006). Nevertheless, emulating the groundwater flow by numerical methods for various real-world problems is too difficult owing to the uncertainty and nonlinearity corresponding to influential parameters such as conductivity and hydraulic gradient, depth of saturated zone, long computational time for emulation, rate of discharge and recharge, and hydraulic head (Gómez-Hernández & Gorelick, 1989; Kovács et al., 2015).

To avoid the restrictions and disadvantages of physically based methods, several empirical techniques have been developed as a significant and renowned branch of machine learning models (MLMs). They are viewed as dependable methods with acceptable precision and a lower computational time for forecasting complex phenomena in different sciences by making connections between input/output without requiring a basic knowledge of physical relationships (Ahmadi et al., 2020; Choubin et al., 2019; Ghalandari et al., 2019; Golestani Kermani et al., 2019; Joloudari et al., 2020; Mahmoudi et al., 2021; Qasem et al., 2019; Shabani et al., 2020; Taherei Ghazvinei et al., 2018; Torabi et al., 2019).

The application of empirical techniques in geosciences is increasing rapidly, mainly in the prediction of extreme events such as GWL fluctuations (Mosavi et al., 2018). Many kinds of research have been performed using traditional MLMs for the prediction of GWL in different places. Shiri et al. (2013) utilized several classical soft computing methods to predict GWL fluctuations with meteorological impact implications in Hongcheon Well station, South Korea. The results of the prediction

showed that for the 1-day-ahead prediction, gene expression programming (GEP) and adaptive neuro-fuzzy inference system (ANFIS) models presented small testing values ( $t$ -statistic) of 0.86 and 0.126, respectively, and high significance levels of 0.932 and 0.9, respectively. Sharafati et al. (2020) developed a novel AI model called gradient boosting regression (GBR) for forecasting GWL over Rafsanjan aquifer in Iran. Using correlation analysis, depending upon the lead periods, the  $R^2$  value changed from 0.66 to 0.94. MLM-based deep neural network (DNN) models, which are an advanced type of recurrent neural network (RNN), are receiving more attention in various sciences (Band et al., 2020; Nabipour et al., 2020). Deep learning (DL) algorithms are black-box methods in which the input/output data are directly connected through a huge-dimension matrix of biases and weights in the hidden layers. In the training process, through the intrinsic ability of this method, the network feedback to the output of calibration data is matched using an optimization algorithm (Mosavi et al., 2020; Rajaei et al., 2019). The emulation results rely on the quantity and quality of calibration data and also the construction of the network. The increased use of DL models, with gated recurrent unit (GRU) and long short-term memory (LSTM) neural networks as improvised architecture, provides vast opportunities for improvements in the fields of water resources and hydrology (Shen, 2018). Although a time scale was embedded into the RNN, theoretical and practical experiments confirmed that conventional gradient-based networks cannot dependably operate with information from more than the last 10 time steps (Gers et al., 1999). To permit the networks to recall inputs for a long period, an explicit memory, called LSTM, was formulated to enhance their ability (Hochreiter & Schmidhuber, 1997). Consequently, the exploding and vanishing gradients that arise in RNNs could be circumvented by substituting the traditional neuron in the hidden layer with an LSTM cell, which has a periodically self-connected linear component. The GRU, according to Cho et al. (2014), has an inherent capability for learning long period dependencies using raw time-series data. One benefit of such a model is that it can efficiently capture nonlinear relationships in historical GWL datasets (Mosavi et al., 2018).

So far, LSTM-based models have been effectively used for modeling in the field of hydrological science (Bowes et al., 2019; Jeong et al., 2020; Vu et al., 2021; Yin et al., 2020; Zhang et al., 2018). Nevertheless, inadequate research has been conducted on different applications of GRU-based models for complex hydrological processes such as GWL prediction. Ghasemlounia et al. (2021) predicted GWL fluctuations of four observation piezometers in an agrarian region through the structure of the novel

bi-directional long short-term memory (BiLSTM)-based DNN model, and found the BiLSTM $2\times$  to be the best model. This model resulted in a root mean square error (RMSE) of 0.17 m and an  $R^2$  of 0.89 in piezometer 4, with a range of 4.49 m. Gao et al. (2020) forecast short-time runoff with an artificial neural network (ANN), GRU and LSTM neural networks without using time-step optimization in sample generation. The results confirmed that DNN models outperformed the typical ANN and their estimation precision was raised by increasing the time step. Finally, GRU was chosen as the best method as it needed less running time for model training. Chen et al. (2021) developed a surrogate method for groundwater emulation using GRU to enhance the efficacy of parameter autocalibration and global sensitivity analysis. The results showed that the GRU surrogate combined with the particle swarm optimization (PSO) algorithm could be used to adopt high-dimensionality parameter calibration tasks with outstanding ability. Pan et al. (2020) forecast the water level in the Yangtze River using different models. They reported that for predicting the 8 o'clock water levels for the next 5 days, the ability of the hybrid convolutional neural network-gated recurrent unit (CNN-GRU) model with three water stations was greater than the other applied models. This model resulted in an RMSE of 0.13, a mean relative error of 3.13% and a Nash-Sutcliffe efficiency (NSE) coefficient of 0.97 as the average value of three seasons. Jeong and Park (2019) applied GRU, LSTM, autoregressive exogenous (ARX) and nonlinear autoregressive exogenous (NARX) models to forecast GWL. They found that LSTM and NARX neural network models worked better than ARX and GRU for monitoring GWL.

In recent years, many scientists have investigated hybrid models. These models can utilize the benefits of diverse models by incorporating their functions. Typical hybrid models are data preprocessing and parameter choice and optimization. The decomposition technique is one of the generally employed data preprocessing models. Commonly used decomposition methods include singular spectrum analysis (SSA), variational mode decomposition (VMD), empirical mode decomposition (EMD) and wavelet decomposition. VMD (Dragomiretskiy and Zosso, 2013) is a groundbreaking non-recessive technique that was developed to disentangle nonlinear eurhythmic signals from intricate data signals with extreme computational efficacy. In VMD, a variational model tries to find a sequence of elements with an assorted rate of recurrence bands and time resolutions (Fang et al., 2019; H. Li et al., 2019). In contrast to EMD, VMD can successfully improve the procedure of the integration problem, where one element encompasses two or further subsignals with distinct variances or one subsignal is separated

into two or more elements with comparable traits (Fosso and Molinas, 2018; Naik et al., 2018). Several investigators have developed hybrid decomposition-based models for forecasting the time-series GWL (Bahmani et al. 2020; Gehrels et al. 1994; Nourani & Mousavi, 2016; Rezaie-balf et al. 2017). Zare and Koch (2018) estimated GWL fluctuations by ANFIS and hybrid wavelet-ANFIS/fuzzy C-means (FCM) clustering approaches in Miandarband plain, Iran. They concluded that the developed methods could be used with satisfactory precision; the hybrid wavelet-ANFIS approach with a Symlet mother wavelet outperformed the other models, with RMSE and  $R^2$  values of 0.17 and 0.984, respectively, in the validation stage. Bahmani and Ouarda (2021) developed different hybrid decomposition models for predicting GWL. The results confirmed the hybrid M5 model tree with wavelet transform (WT-M5) model as the best technique, with an  $R^2$  value of 0.91 and RMSE of 0.26 in the validation stage. Azari et al. (2021) used a coupling preprocessing method with linear stochastic models (i.e. a linear model with one autoregressive and one seasonal moving average parameter) to develop a novel model for GWL prediction. The statistical evaluation of this model confirmed its acceptable performance for GWL estimation, with a scatter index (SI) of 0.0004, root mean squared relative error (RMSRE) of 0.0004, corrected Akaike's information criterion (AICc) of 151 and mean absolute percentage error (MAPE) of 0.0003. Wu et al. (2021) predicted GWL using different stand-alone and hybrid models. They found that the hybrid wavelet transform-multivariate long short-term memory (WT-MLSTM) technique, which integrates exterior correlated parameters into forecasts, worked better than the traditional methods. Azizpour et al. (2021) estimated GWL using the self-adaptive extreme learning machine (SAELM) and the hybrid SAELM with WT models. They reported that the hybrid (WT-SAELM) model was the best model, with variance accounted for (VAF), NSE and correlation coefficient ( $R$ ) values of 97.450, 0.973 and 0.988, respectively. Tao et al. (2022) reviewed advanced MLMs developed to predict GWL in together with accomplishments in this field from 2008 to 2020 (138 papers). They evaluated the types of methods, data extent, time scale, input/output variables, performance metrics applied and the excellent models recognized.

In this study, since the nature of modeling consists of temporal dependencies, GRU neural networks as emulators are investigated as a suitable possibility to predict GWL in Qoşaçay plain, Iran. Thus, several GRU-based models with the sequence-to-one regression module are developed because of the lack of meteorological variables recorded in the study area. The developed models are a general single-gated recurrent

unit-layer network, a simple double-gated recurrent unit-layers network (GRU2), a novel proposed double-gated recurrent unit model coupled with multiplication layer ( $\times$ ) (GRU2 $\times$ )-layers network and a hybrid variational mode decomposition-gated recurrent unit (VMD-GRU) model. GRU2 $\times$  is a state-of-the-art coupling version of the GRU2-layers network model with a multiplication layer ( $\times$ ), and the hybrid VMD-GRU model is a new combination of VMD with a general single-GRU-layer network model. Of these models, the hybrid VMD-GRU and GRU2 $\times$  models are considered as the newly combined version models among the traditional DL models, which improve the accuracy of prediction. The novelty of the present research lies in developing an advanced new architecture of the GRU2 $\times$  neural network and hybrid VMD-GRU models with the sequence-to-one regression module for predicting elaborate natural events such as the mean monthly time-series GWL. To the best of our knowledge, despite many studies having been performed for estimating GWL using the conventional architecture of DL-based models, no studies have been reported in the literature aimed at the application of the exceptional and newly proposed GRU2 $\times$  and hybrid VMD-GRU models.

The scope and main purposes of the present research are:

- (1) To develop different layer structures of GRU-based and hybrid VMD-GRU models for the prediction of GWL using a long-term water table depth with 33 different monitoring piezometers with time-series traits without operating meteorological variables in Qoşaçay plain.
- (2) To determine the ideal numbers of different hyper-parameters in each model for better configuration of the developed models, to reduce the adverse influences of the overfitting problem, using a trial-and-error process.
- (3) To assess and match the outcomes of the emulation to determine the perfect model through statistical metrics.

## 2. Site description and data collection

Qoşaçay plain is located to the south of Lake Urmia in Qoşaçay town, West Azerbaijan Province, Northwest Iran, with an area of approximately 1100 km<sup>2</sup>. It has coordinates of 46°06′10″ E latitude and 36°58′10″ N longitude, lies 1314 m above sea level, comprises cultivable land, and has semi-arid and cold weather. Figure 1 depicts the geographical location of the study district.

Since the real evapotranspiration ratio ( $ET_a$ ) is high in summer, groundwater stores are used as a supplementary or alternative source. The irrigation period is

from April to September and the average annual potential evapotranspiration ( $ET_{Pot}$ ) is about 900–1500 mm (Jalilvand et al., 2019). Peak rainfall generally occurs from January to May. The mean annual temperature and precipitation were reported to be 12°C and 284 mm, respectively (EARWO, 2020). Since this area is very vulnerable to drought, groundwater plays an important role. The stream direction of the groundwater is from the south-east to the north-west of the plain (Norouzi & Moghaddam, 2020).

To predict the average monthly GWL in the study region, the current investigation used 216 average monthly measured water table depths from 33 monitoring piezometers with different statistical characteristics in the period of April 2002–March 2020. Figure 2 shows the sites of the observation piezometers in Qoşaçay plain.

Figure 3 presents the long period time series of average monthly GWL between April 2002 and March 2020 in the study area.

Descriptive statistics of the average monthly GWL in Qoşaçay plain are presented in Table 1, where CV and STDV are the coefficient of variation and standard deviation, respectively.

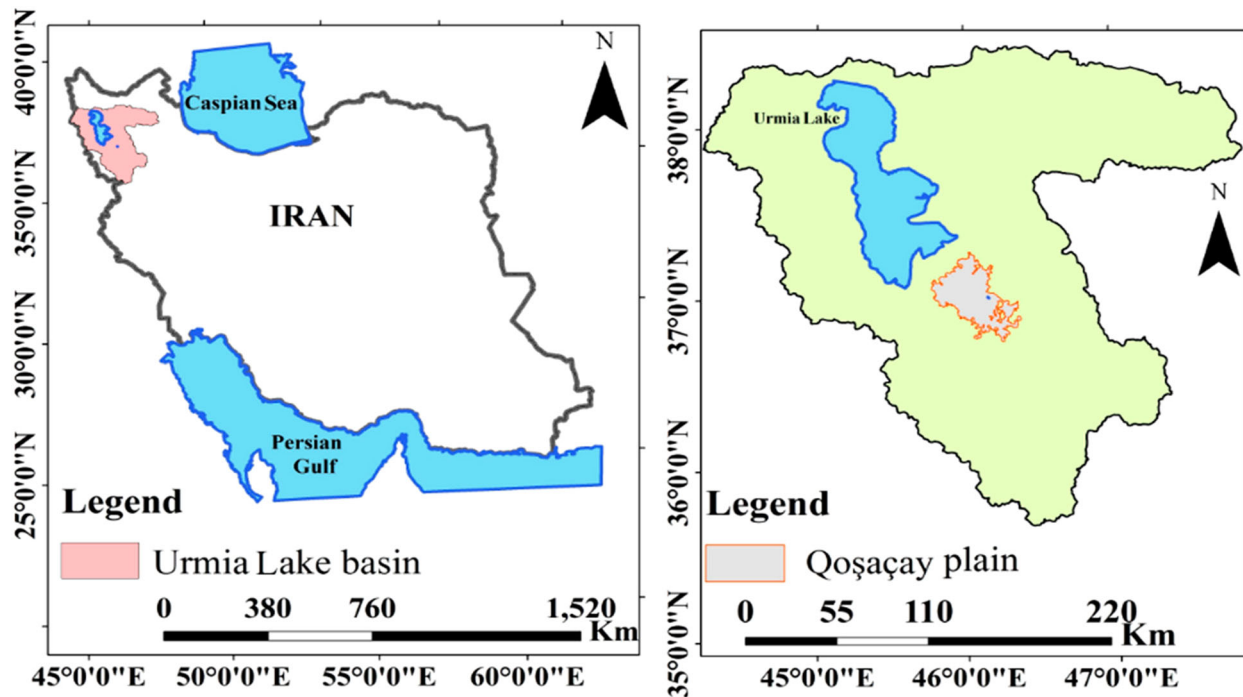
Based on Table 1, it can be seen that the variation trend in GWL is very unsteady, with an extremely complicated basic nature and nonlinearity, owing to the high range, skewness and kurtosis. Therefore, robust accurate techniques are required for modeling.

## 3. Methodology

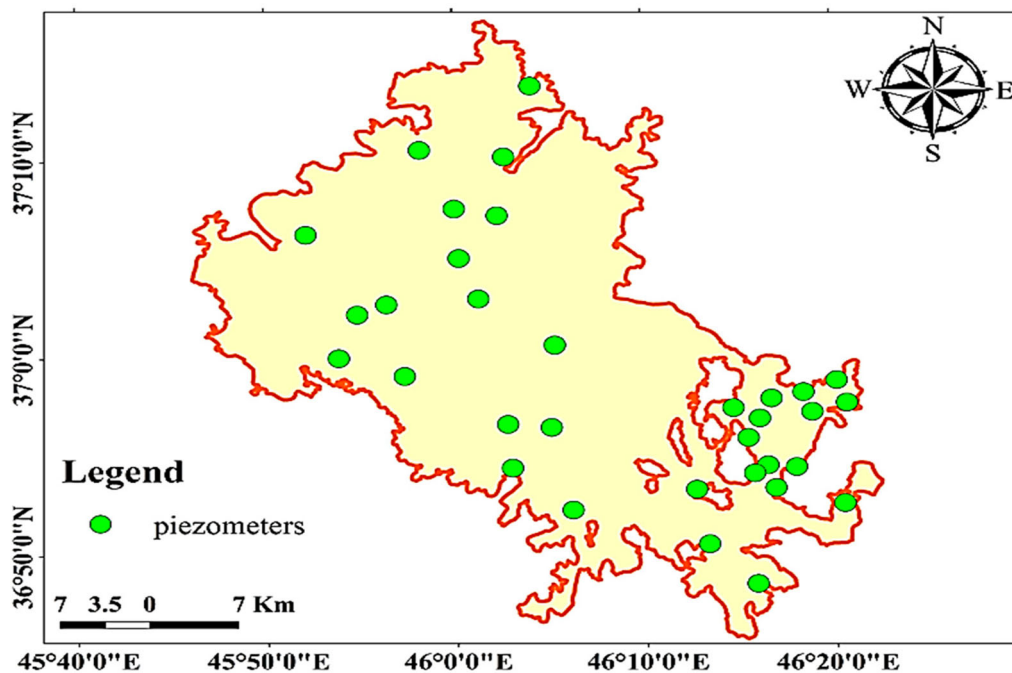
In the current study, to forecast the time-series GWL, different GRU-based neural network models are developed. In this regard, first, all datasets are normalized to zero average and unit variance, as suggested by Lawrence et al. (1997). Then, the average monthly recorded GWL datasets are divided into two subsets: the first 70% of the total datasets (150 months), between April 2002 and September 2014, is used for calibration of models; and the remaining 30% of the total datasets (66 months), between October 2014 and March 2020, is used for validation.

### 3.1. Gated recurrent unit (GRU) cell structure

The RNN, which is a specific type of neural network model, is very appropriate for time-series and serial data modeling. In RNNs, an unfurled loop cell induces to inflow the previous data. Nevertheless, because of the explosion gradient dilemma, the accuracy of the typical RNN model diminishes during the learning process of a long period sequence during backpropagation (Rangapuram et al., 2018).



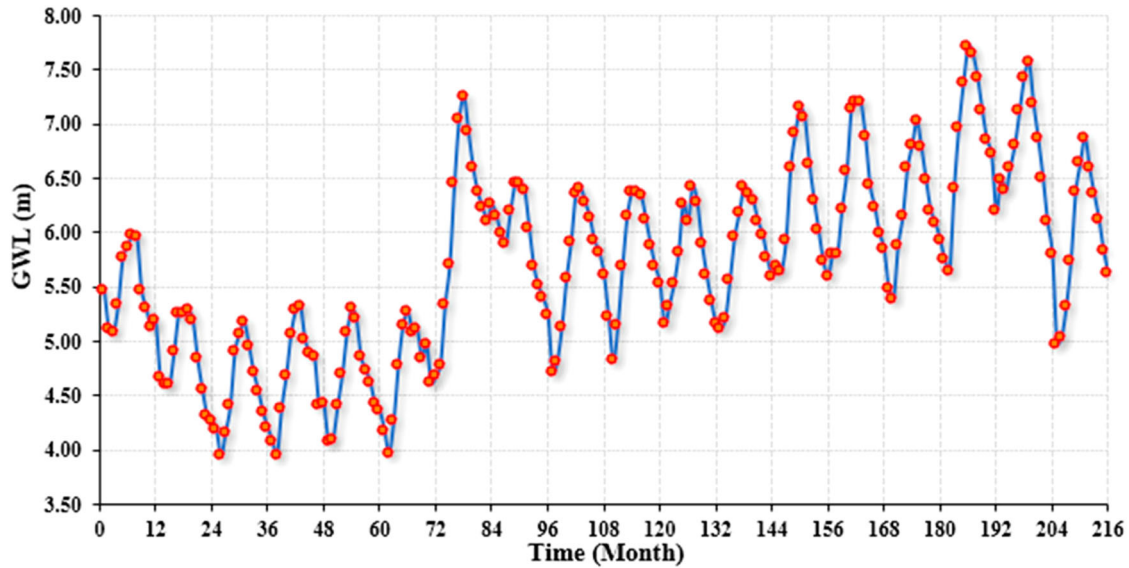
**Figure 1.** Geographical location map of the study district.



**Figure 2.** Locations of the 33 piezometers in Qoşaçay plain, West Azerbaijan Province, Iran.

GRU, as a special kind of RNN, has been developed to preclude the fundamental difficulties in the common RNN using a specific infrastructure (Sak et al., 2014). It unites three gates of LSTM i.e. input, forget and output, into two control gates:  $Z_t$  (update gate) and  $r_t$  (reset gate). The interior structure of a GRU memory cell is depicted as a sequence diagram in Figure 4.

As described in Figure 4, there is an input layer comprised of several neurons, where the number of neurons is controlled through the dimensions of feature space. Correspondingly, the number of neurons in the output layer corresponds to the output space (Cho et al., 2014). The structure of the GRU can be formulated using Equations (1) and (2) (Cho et al.,

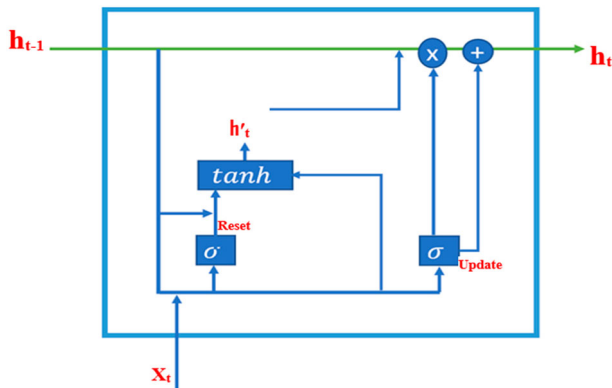


**Figure 3.** Time-series plots of observed average monthly groundwater level (GWL) between April 2002 and March 2020 (216 months).

**Table 1.** Statistical factors of the average monthly groundwater level (GWL) between April 2002 and March 2020 in the study area.

| Variable | Data period  | Min. | Max. | Range | Mean | STDV | Skewness | CV   |
|----------|--------------|------|------|-------|------|------|----------|------|
| GWL (m)  | Whole period | 3.94 | 7.72 | 3.87  | 5.71 | 0.86 | 0.023    | 0.15 |

Note: STDV = standard deviation; CV = coefficient of variation.



**Figure 4.** Internal structure of a gated recurrent unit (GRU) memory cell.

2014):

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t \quad (1)$$

$$h'_t = \tanh(r_t * (U_h h_{t-1}) + W_h x_t + b_h) \quad (2)$$

where  $h_t$ ,  $h_{t-1}$  and  $h'_t$  are the hidden layer of step  $t$ , the hidden layer of step  $t - 1$  and the present new state of step  $t$ , which are considered as a summation of the information at  $h_{t-1}$  and the input information at step  $t$ ,  $x_t$ . The network weight matrices and bias vectors are signified as  $U$ ,  $W$  and  $b$ , which are attained in the training phase. The hidden layer contains memory cells that take into consideration the principal functions of the GRU network.

GRU uses control gates to adjust the retaining, forgetting and updating of sequence data.  $z_t$  determines the volume of ignored previous information and newly added data, while  $r_t$  controls the volume of the above information which is passed into the new state.  $z_t$  and  $r_t$  can be formulated by the following equations (Cho et al., 2014):

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (3)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (4)$$

where  $x_t$  denotes the input at step  $t$ , and  $\sigma$  and  $\tanh$  are the logistic sigmoid and hyperbolic tangent functions, respectively.

In contrast to LSTM, GRU has no remote memory cells; instead, it operates a single hidden state  $h_t$  to send information over time steps. Both GRU and LSTM perform in nearly the same way; nonetheless, as GRU has a lower number of learnable parameters and a more solid structure, it can converge readily, when the volume of data is not too big (R. Fu et al., 2016; Greff et al., 2016).

### 3.2. Variational mode decomposition (VMD)

As an advanced data decomposition technique, the VMD algorithm takes advantage of variational modes to generate a certain number of distinct elements for determining the variational problem. Based on the set modes number, the VMD algorithm separates a signal into certain

bandwidths with diverse center frequencies and upgrades progressively by operating the alternating direction multiplier optimization method (ADMM) (Bai et al., 2021; Wang et al., 2015).

In VMD, it is presumed that every model has a bandwidth with a given center rate of recurrence. In more detail, it is supposed that a basic signal  $f(t)$  is broken into  $K$  dissimilar separate modes  $u_k(t)$  to minimize the total computed bandwidth of every mode, where every mode is accumulated around a center frequency  $w_k$  (He et al., 2019; Nechikkat et al., 2015). A constrained variational problem, depicted by Equation (5), is applied to answer the decomposition problem for a time series  $f(t)$ :

$$\left\{ \begin{array}{l} \min_{u_k(t), w_k(t)} \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k(t)} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{k=1}^K u_k(t) = f(t) \end{array} \right\} \quad (5)$$

where  $K$  is the number of modes and  $t$  is time.  $u_k(t)$  and  $w_k(t)$  are the decomposed modes and relevant center frequencies, respectively.  $\delta(t)$  is the Dirac distribution and  $j$  is an imaginary unit, so that  $j^2 = -1$ .  $\otimes$  is the convolution operator and  $f(t)$  is the  $t$ th data of the basic signal (Wang et al., 2015).

To find the optimal resolution to the constrained variational problem and progress the performance efficiency, a quadratic penalty parameter ( $a$ ) and Lagrange multiplier ( $\lambda$ ) are applied to obtain an unconstrained version of Equation (5) (J. Li et al., 2019; Niu et al., 2018). The augmented Lagrangian  $L$  is expressed by the following formula:

$$\begin{aligned} L(u_k(t), w_k(t), \lambda) &= a \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-jw_k(t)} \right\|_2^2 \\ &+ \left\| \sum_{k=1}^K u_k(t) - f(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \quad (6)$$

The ADMM algorithm, which is a well-known dividing technique for discrete optimization problems, is applied to search for the saddle point of the augmented Lagrange function and solve Equation (6) in the VMD algorithm (Dragomiretskiy and Zosso, 2013; W. Fu et al., 2019). In ADMM, decision variables ( $\hat{u}_k^{n+1}$ ,  $w_k^{n+1}$  and  $\hat{\lambda}^{n+1}$ ) are divided into numerous blocks on the basis of their characteristics, and the main problem is separated into several lesser subproblems of a single block, which will be optimized reiteratively by affixing all other blocks in every internal sequence (Chang et al., 2016; W. Fu et al., 2020).

Equations (7)–(9) are used to upgrade the mode  $u_k(\omega)$  in the frequency sphere, center frequencies  $\omega_k$  and concurrently  $\lambda$ , respectively. In the time sphere, the mode  $u_k(t)$  is the real section of the inverse Fourier transform of  $u_k(\omega)$ , described by Equation (7).

$$\hat{u}_k^{n+1} = \frac{\hat{f}(w) - \sum_{i < k}^K \hat{u}_i^{n+1}(w) - \sum_{i > k}^K \hat{u}_i^n(w) + 0.5\hat{\lambda}(w)}{1 + 2a(w - w_k^n)^2} \quad (7)$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k^{n+1}(w)|^2 dw}{\int_0^\infty |\hat{u}_k^{n+1}(w)|^2 dw} \quad (8)$$

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left[ \hat{f}(w) - \sum_{k=1}^K \hat{u}_k^{n+1}(w) \right] \quad (9)$$

where  $n$  is the number of repetitions and  $\tau$  is the time step of the dual ascent (iterative parameter) signifying the noise tolerance of VMD.  $\hat{f}(w)$ ,  $\hat{u}_i(w)$ ,  $\hat{\lambda}(w)$  and  $\hat{u}_k^{n+1}(w)$  are Fourier transforms of  $f(t)$ ,  $u_i(t)$ ,  $\lambda(t)$  and  $\hat{u}_k^{n+1}(t)$ , respectively (Wang et al., 2015).

Once the stopping conditions have been fulfilled, the whole reiterative optimization procedure in Equation (8) is stopped; otherwise, the calculation will be repeated with Equations (7) and (8) until Equation (10) is satisfied.

$$\frac{\sum_{k=1}^K \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon \quad (10)$$

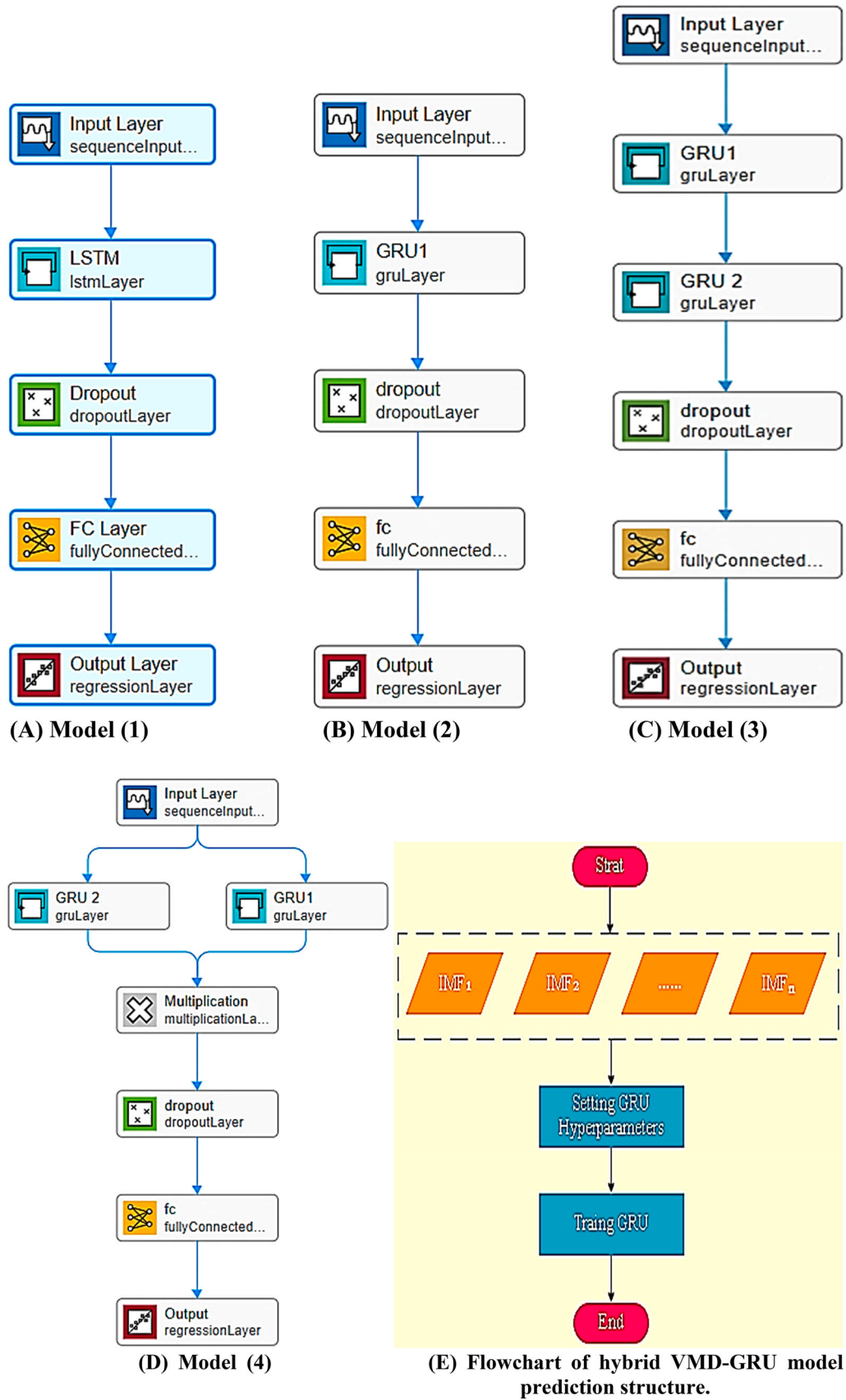
In Equation (10),  $\varepsilon$  is the convergence tolerance that affects the reformation error of VMD.

### 3.3. Model development

As recommended by Maier et al. (2010), the initial architectural construction of layer network models should be designed according to the intentions of the investigation. Therefore, the general single-LSTM layer network model (model 1) was first developed in MATLAB release 2021a (MATLAB 2021). This is taken as our base model and is used to match the performance of other developed models. Then, to tune the network topology pattern (NTP) hyperparameter in the general single-GRU layer network model (model 2), two deeper structures (models 3 and 4) are developed. Finally, model 2 is combined with the VMD technique to generate the hybrid VMD-GRU model. Figure 5 illustrates the structure of the designed models.

As depicted in Figure 5, model 3 is a simple double-GRU model (GRU2), whereas the newly proposed model (model 4) is a double-GRU coupled with multiplication





**Figure 5.** Architecture of the designed layer network models. (A) Structure of the general single-long short-term memory (LSTM)-layer network model (model 1); (B) structure of the general single-gated recurrent unit (GRU)-layer network model (model 2); (C) structure of the simple double-GRU-layer network model (model 3); (D) structure of the newly proposed double-GRU layer network coupled with multiplication layer ( $\times$ ) (GRU2 $\times$ ) (model 4); (E) structure of the hybrid variational mode decomposition-gated recurrent unit (VMD-GRU) model.

layer ( $\times$ ) with the sequence output mode. The number of layers in the construction of models 1, 2, 3 and 4 is 5, 5, 6 and 7, respectively. In the hybrid VMD-GRU model, the VMD algorithm is applied to decompose the original time series of GWL into a series of elements to reduce the complexity and non-stationary of GWL (decomposition phase). Next, the general single-GRU-layer network model (model 2) is trained to estimate the complex relationships among the input/output variables and create the ultimate outcome by uniting the information from all extracted subsequences (ensemble phase).

In this research, to achieve a suitable configuration to reduce the impact of overfitting and improve the capability of the developed models, the algorithm tuning process, using different types of hyperparameters, is operated. Because there are no effective instructions to determine the appropriate hyperparameters in a certain model and dataset, the current procedure is considered a time-consuming and arduous undertaking. In this respect, various scenarios are characterized to determine the appropriate hyperparameters, in combination with a trial-and-error procedure by the user.

In the building of the developed models, the input layer intakes time-series datasets with a size of 1. Since GWL has a time dimension, time steps on the monthly scale should be used as an independent variable. To tune the state activation function (SAF), diverse unions of *tanh* and *softsign*, including *softsign-tanh*, *softsign-softsign*, *tanh-softsign* and *tanh-tanh*, in each GRU layer with the sigmoid gate activation function are used separately. Moreover, to tune the number of hidden units (NHU), diverse values in each model are examined.

The multiplication layer multiplies inputs from multiple neural network layers element-wise. The number of inputs in this layer is determined characteristically by the operated program; nonetheless, it has a sole output (MATLAB, 2021).

DNNs are suitable for assessing large datasets; however, a very large dataset will cause overfitting problems. Because a dropout layer provides a functioning regularization system, it has been used to reduce overfitting (Hinton et al., 2012). Theoretically, it automatically ignores or drops some neurons with a certain probability rate of  $P$  in a layer allocated at each training reiteration (Srivastava et al., 2014). In this way, it breaks up positions where network layers coadjust to errors from previous layers and, consequently, makes the model more robust. In this emulation, values of 0.3, 0.5 and 0.7 are used to tune the learning dropout rate ( $P$ -rate).

Although LSTM and GRU neural networks have a high capability to learn long-term time-series datasets, their fitting ability can be unsatisfactory (Zhang et al., 2018), and thus a fully connected layer must be operated.

The fully connected layer multiplies the inputs via a matrix of weights and then adds a vector of bias to progress the fitting ability of the developed models (MATLAB, 2021). In this emulation, the input size of this layer in all developed models in the training phase is set to 'auto' to control its size automatically via the operated program, and its output size is set to 1. The final layer of all developed networks is set as a regression output layer to estimate the half-mean-squared-error loss for regression via the loss function, as follows:

$$\text{Loss} = \sum_{i=1}^N (y_m - y_p)^2 \quad (11)$$

where  $y_p$  and  $y_m$  are the predicted and measured values of GWL at time  $i$ , respectively.

To update the weights and bias of each network in the developed models, the calibration options are set as in the Adam optimization algorithm, with a maximum iteration number of 2000, a learn rate drop period of 125, an initial learning rate of 0.001 and a learn rate drop factor of 0.2. In addition, to avoid the vanishing gradients problem, the initial batch size and gradient threshold are set to 128 and 1, respectively. These parameters are fundamental components of the developed structures, which have a noticeable influence on their operation. On the whole, in the hybrid VMD-GRU model,  $K$ ,  $\alpha$ ,  $\tau$  and  $\varepsilon$  factors affect the decomposition performance of VMD, but it is difficult to specify their optimal values. Hence, in this study,  $K$  is deliberated as a key hyperparameter and is tuned with different values, while the other factors are set to the default values.

#### 4. Performance evaluation criteria

In this study, the RMSE and coefficient of determination ( $R^2$ ) were applied to assess the effectiveness and accuracy of the designed model for predicting the time-series GWL. These statistical metrics are formulated by the following equations:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (12)$$

$$R^2 = \left( \frac{\sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x\sigma_y} \right)^2 \quad (13)$$

where  $N$  is the number of datasets,  $x_i$  and  $y_i$  are the measured and estimated GWL at the time  $i$ ,  $\sigma_x$  and  $\sigma_y$  are the standard deviation of the measured and estimated GWL, respectively, and  $\mu_x$  and  $\mu_y$  are the average of the measured and estimated GWL, respectively. The optimal values for these metrics are 0 and 1, respectively.

## 5. Results and discussion

### 5.1. Validation of general single-LSTM and single-GRU models

After numerous experiments, the optimal values of the  $P$ -rate and NHU for the general single-LSTM model (model 1) are determined as 0.5 and 100, respectively, and for the general single-GRU model (model 2) as 0.5 and 30, respectively. Moreover, the appropriate kind of SAF for both models is identified as *tanh*. In the training phase of the optimal scenario of both models, after 2000 reiterations, the final loss, RMSE and learning rate values are approximately  $1\text{E-}4$ ,  $1\text{E-}3$  and  $1.638\text{E-}13$ , respectively. After calibration, the model is tested through the testing datasets. The results of both models with respect to SAF,  $P$ -rate and NHU in the training and testing phases for emulating GWL in Qoşaçay plain are presented in Tables 2 and 3, respectively, where the bold values signify the attributes of the optimal scenarios.

Based on Table 2, for both models, with the same NHU, by increasing the  $P$ -rate, the value of the running time for model training decreases; and with the same  $P$ -rate, by increasing the value of NHU, the running time increases. Moreover, with the same NHU and  $P$ -rate, scenarios with an SAF of *tanh* are faster than those with *softsign*. Besides, in accordance with Table 3, it can be discerned that in the single-GRU model (model 2), too high a value of NHU (50) causes to an increase in the RMSE for the overfitting problem, whereas a smaller value of NHU (15) reduces the learning ability of the network for the underfitting problem.

All in all, both models are more precise in the calibration stage than in the validation stage. Although model 1 is very slightly more accurate than model 2, as it needs more NHU for convergence, it takes too long in comparison with model 2.

### 5.2. Validation of double-GRU models

Various combinations of activation function were used in the SAFs of GRU layers in the hidden layer of the simple double-GRU (GRU2) model (model 3) and the newly proposed GRU2 $\times$  model (model 4). In this respect, after abundant experiments, the *softsign-tanh* union qualified as the best choice in both models. Likewise, the optimal values for  $P$ -rate and NHU in both models are determined as 0.5 and 25 (each layer), respectively.

Applying different activation functions in the hidden layers of both models results in more complex and nonlinear functions being learnt. In this regard, the *softsign-tanh* combination encourages the models to be less susceptible to the overfitting problem in the training phase. The *softsign* activation function can increase the

rapidity of model calibration, while *tanh* can suitably capture complex relationships in long-period time series (Yin et al., 2020). Nevertheless, in this case, the type of data has a substantial impact the model quality.

In the training stage of the optimal scenario after 2000 reiterations, the values of RMSE and loss obtained for model 3 are approximately  $1\text{E-}3$  and  $1\text{E-}4$ , respectively, and for model 4 are approximately  $1\text{E-}4$  and  $1\text{E-}5$ , respectively. The value of the learning rate for both models is  $1.638\text{E-}13$ . The results of models 3 and 4 with respect to  $P$ -rate and NHU (each layer) in the training and testing phases are presented in Table 4, where the expression *softsign-tanh* indicates that the kind of SAF in layers 1 and 2 of the GRU is *softsign* and *tanh*, respectively. The bold values indicate the characteristics of the optimal scenario.

Consistent with Table 4, it can be noticed that, in both models, too high a value of NHU (50) leads to an increase in RMSE because of the overfitting problem, but a smaller value of NHU (15) causes the underfitting problem. Furthermore, for the same NHU, by increasing the  $P$ -rate value, the running time decreases, and at the same  $P$ -rate, as the value of NHU increases, the running time increases. In summary, both models are more precise in the calibration stage than in the validation stage, but the GRU2 $\times$  model is noticeably faster than the GRU2 model.

### 5.3. Results of the hybrid VMD-GRU model

In the hybrid VMD-GRU model, after several tests, the optimal values of  $P$ -rate, NHU and  $K$  are 0.5, 128 and 3, respectively, and the appropriate sort of SAF is determined to be *tanh*. In the validation stage, the hybrid VMD-GRU model under optimal hyperparameters results in an  $R^2$  of 0.92 and an RMSE of 0.16 m. Figure 6 presents the values of RMSE against NHU under the optimal value of hyperparameters for the  $K$  values used in the validation stage.

### 5.4. Performance comparison

In the current research, the authors emphasize the substructure and capacity of the developed models (models 1–4) to assess their performance. In this regard, MATLAB 2021a (MATLAB 2021) offers a noteworthy parameter, namely, total learnable parameters (TLP), as a vital measure to judge the operational capacity of the models. A model with optimal NHU and NTP results in an appropriate TLP, and, accordingly, encourages a reduction in the impacts of the overfitting condition. In this direction, Table 5 presents the TLP values of the developed models (models 1–4) under the optimal scenario.

According to Table 5, model 3, owing to the maximal value of TLP, is seen as the most exact and best of the

**Table 2.** Results of the general single-long short-term memory (LSTM) model (model 1) and single-gated recurrent unit (GRU) model (model 2) in the calibration stage.

| Single-LSTM model (model 1) |                |                  |                 | Single-GRU model (model 2) |                |                  |                 |
|-----------------------------|----------------|------------------|-----------------|----------------------------|----------------|------------------|-----------------|
| NHU                         | <i>P</i> -rate | Running time (s) |                 | NHU                        | <i>P</i> -rate | Running time (s) |                 |
|                             |                | <i>tanh</i>      | <i>softsign</i> |                            |                | <i>tanh</i>      | <i>softsign</i> |
| 75                          | 0.3            | 110              | 114             | 15                         | 0.3            | 47               | 51              |
| 75                          | 0.5            | 103              | 108             | 15                         | 0.5            | 45               | 48              |
| 75                          | 0.7            | 95               | 98              | 15                         | 0.7            | 43               | 45              |
| 85                          | 0.3            | 115              | 121             | 20                         | 0.3            | 51               | 55              |
| 85                          | 0.5            | 109              | 114             | 20                         | 0.5            | 49               | 53              |
| 85                          | 0.7            | 102              | 108             | 20                         | 0.7            | 48               | 50              |
| 100                         | 0.3            | 145              | 151             | 25                         | 0.3            | 57               | 61              |
| <b>100</b>                  | <b>0.5</b>     | <b>139</b>       | <b>146</b>      | 25                         | 0.5            | 55               | 58              |
| 100                         | 0.7            | 130              | 135             | 25                         | 0.7            | 53               | 55              |
| 110                         | 0.3            | 211              | 219             | 30                         | 0.3            | 77               | 79              |
| 110                         | 0.5            | 204              | 212             | <b>30</b>                  | <b>0.5</b>     | <b>71</b>        | <b>74</b>       |
| 110                         | 0.7            | 197              | 206             | 30                         | 0.7            | 66               | 61              |
| 125                         | 0.3            | 231              | 236             | 50                         | 0.3            | 95               | 99              |
| 125                         | 0.5            | 223              | 228             | 50                         | 0.5            | 90               | 94              |
| 125                         | 0.7            | 215              | 221             | 50                         | 0.7            | 82               | 88              |

Note: Characteristics of the optimal scenarios are highlighted in bold.  
 NHU = number of hidden units; *P*-rate = learning dropout rate.

**Table 3.** Results of the general single-long short-term memory (LSTM) model (model 1) and single-gated recurrent unit (GRU) model (model 2) in the testing stage.

| Single-LSTM model (model 1) |                |             |                 |             |                 | Single-GRU model (model 2) |                |             |                 |             |                 |
|-----------------------------|----------------|-------------|-----------------|-------------|-----------------|----------------------------|----------------|-------------|-----------------|-------------|-----------------|
| NHU                         | <i>P</i> -rate | RMSE (m)    |                 | $R^2$       |                 | NHU                        | <i>P</i> -rate | RMSE (m)    |                 | $R^2$       |                 |
|                             |                | <i>tanh</i> | <i>softsign</i> | <i>tanh</i> | <i>softsign</i> |                            |                | <i>tanh</i> | <i>softsign</i> | <i>tanh</i> | <i>softsign</i> |
| 75                          | 0.3            | 0.32        | 0.35            | 0.39        | 0.37            | 15                         | 0.3            | 0.35        | 0.39            | 0.36        | 0.33            |
| 75                          | 0.5            | 0.31        | 0.33            | 0.43        | 0.40            | 15                         | 0.5            | 0.34        | 0.37            | 0.39        | 0.37            |
| 75                          | 0.7            | 0.33        | 0.35            | 0.34        | 0.33            | 15                         | 0.7            | 0.36        | 0.38            | 0.31        | 0.29            |
| 85                          | 0.3            | 0.28        | 0.32            | 0.51        | 0.47            | 20                         | 0.3            | 0.31        | 0.35            | 0.46        | 0.43            |
| 85                          | 0.5            | 0.25        | 0.28            | 0.56        | 0.52            | 20                         | 0.5            | 0.28        | 0.31            | 0.51        | 0.48            |
| 85                          | 0.7            | 0.27        | 0.31            | 0.53        | 0.48            | 20                         | 0.7            | 0.30        | 0.33            | 0.49        | 0.44            |
| 100                         | 0.3            | 0.23        | 0.24            | 0.72        | 0.70            | 25                         | 0.3            | 0.28        | 0.32            | 0.62        | 0.58            |
| <b>100</b>                  | <b>0.5</b>     | <b>0.21</b> | <b>0.23</b>     | <b>0.76</b> | <b>0.74</b>     | 25                         | 0.5            | 0.26        | 0.29            | 0.65        | 0.62            |
| 100                         | 0.7            | 0.22        | 0.25            | 0.74        | 0.68            | 25                         | 0.7            | 0.29        | 0.31            | 0.6         | 0.59            |
| 110                         | 0.3            | 0.25        | 0.30            | 0.57        | 0.53            | 30                         | 0.3            | 0.25        | 0.27            | 0.71        | 0.69            |
| 110                         | 0.5            | 0.23        | 0.27            | 0.59        | 0.57            | <b>30</b>                  | <b>0.5</b>     | <b>0.22</b> | <b>0.24</b>     | <b>0.75</b> | <b>0.73</b>     |
| 110                         | 0.7            | 0.26        | 0.29            | 0.53        | 0.50            | 30                         | 0.7            | 0.24        | 0.26            | 0.72        | 0.70            |
| 125                         | 0.3            | 0.35        | 0.38            | 0.41        | 0.39            | 50                         | 0.3            | 0.38        | 0.41            | 0.38        | 0.36            |
| 125                         | 0.5            | 0.31        | 0.35            | 0.37        | 0.41            | 50                         | 0.5            | 0.34        | 0.38            | 0.43        | 0.40            |
| 125                         | 0.7            | 0.34        | 0.37            | 0.40        | 0.37            | 50                         | 0.7            | 0.37        | 0.4             | 0.40        | 0.39            |

Note: Characteristics of the optimal scenarios are highlighted in bold.  
 NHU = number of hidden units; *P*-rate = learning dropout rate; RMSE = root mean square error.

developed GRU-based models. However, based on the values of RMSE and  $R^2$ , it can be differentiated that for the same hyperparameters the novel proposed model 4 works better than model 3. This incongruity and incompatibility can mainly be explained by the excessing the TLP value in model 3 compared to models 4 and 2. In other words, as a result of overcapacity, model 3 has memorized the calibration dataset and has overfitted and, consequently, has lost out in the optimization procedure. All things considered, model 3 has poor performance and cannot be considered as an appropriate model to forecast the average monthly time-series GWL in the study region.

From the TLP values in Table 5, it can be inferred that for the same hyperparameters, adding a multiplication layer yields a fit NTP accompanied by a suitable TLP in comparison with model 3. Model 4 decreases the value of RMSE by 25% and 18%, respectively, compared with models 3 and 2. Overall, it cannot be unconditionally concluded that only by increasing the number of layers and NHU will the precision and ability of GRU-based models definitively improve. In particular, as a significant outcome, to accomplish an effective GRU-based model, a fitting structural NTP with suitable TLP and an appropriate NHU should be designed and operated.

**Table 4.** Results of double-gated recurrent unit (GRU) models 3 and 4 (GRU2 and GRU2 $\times$ ) in the training and testing phases.

| NHU       | P-rate     | SAF <sub>1</sub> -SAF <sub>2</sub> | Training stage   |           | Testing stage |             |                |             |
|-----------|------------|------------------------------------|------------------|-----------|---------------|-------------|----------------|-------------|
|           |            |                                    | Running time (s) |           | RMSE (m)      |             | R <sup>2</sup> |             |
|           |            |                                    | Model 3          | Model 4   | Model 3       | Model 4     | Model 3        | Model 4     |
| 15        | 0.3        | <i>softsign-tanh</i>               | 85               | 79        | 0.38          | 0.35        | 0.44           | 0.50        |
| 15        | 0.5        | <i>softsign-tanh</i>               | 79               | 74        | 0.36          | 0.31        | 0.52           | 0.59        |
| 15        | 0.7        | <i>softsign-tanh</i>               | 74               | 69        | 0.37          | 0.33        | 0.48           | 0.55        |
| 20        | 0.3        | <i>softsign-tanh</i>               | 92               | 85        | 0.33          | 0.29        | 0.54           | 0.60        |
| 20        | 0.5        | <i>softsign-tanh</i>               | 85               | 80        | 0.30          | 0.27        | 0.59           | 0.66        |
| 20        | 0.7        | <i>softsign-tanh</i>               | 78               | 71        | 0.31          | 0.28        | 0.50           | 0.62        |
| 25        | 0.3        | <i>softsign-tanh</i>               | 118              | 93        | 0.29          | 0.21        | 0.62           | 0.76        |
| <b>25</b> | <b>0.5</b> | <b><i>softsign-tanh</i></b>        | <b>102</b>       | <b>87</b> | <b>0.24</b>   | <b>0.18</b> | <b>0.68</b>    | <b>0.86</b> |
| 25        | 0.7        | <i>softsign-tanh</i>               | 91               | 81        | 0.28          | 0.22        | 0.65           | 0.74        |
| 30        | 0.3        | <i>softsign-tanh</i>               | 144              | 98        | 0.35          | 0.26        | 0.48           | 0.68        |
| 30        | 0.5        | <i>softsign-tanh</i>               | 131              | 91        | 0.32          | 0.23        | 0.55           | 0.75        |
| 30        | 0.7        | <i>softsign-tanh</i>               | 109              | 84        | 0.34          | 0.25        | 0.50           | 0.72        |
| 50        | 0.3        | <i>softsign-tanh</i>               | 178              | 134       | 0.44          | 0.39        | 0.31           | 0.45        |
| 50        | 0.5        | <i>softsign-tanh</i>               | 161              | 128       | 0.38          | 0.35        | 0.41           | 0.51        |
| 50        | 0.7        | <i>softsign-tanh</i>               | 148              | 120       | 0.42          | 0.38        | 0.37           | 0.49        |

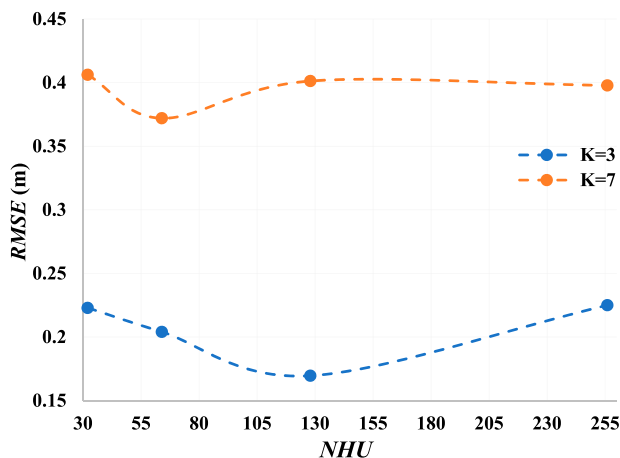
Note: Characteristics of the optimal scenario are highlighted in bold.

NHU = number of hidden units; P-rate = learning dropout rate; SAF = state activation function; RMSE = root mean square error.

**Table 5.** Characteristics and statistical metrics in the optimal scenario of the developed models in the validation stage.

| Model                         | NHU | RMSE (m) | R <sup>2</sup> | Running time (s) | TLP   |
|-------------------------------|-----|----------|----------------|------------------|-------|
| General single-LSTM (model 1) | 100 | 0.21     | 0.76           | 139              | 40901 |
| General single-GRU (model 2)  | 30  | 0.22     | 0.75           | 71               | 2911  |
| GRU2 (model 3)                | 25  | 0.24     | 0.68           | 102              | 5876  |
| GRU2 $\times$ (model 4)       | 25  | 0.18     | 0.86           | 87               | 4076  |

Note: LSTM = long short-term memory; GRU = gated recurrent unit; NHU = number of hidden units; RMSE = root mean square error; TLP = total learnable parameters.

**Figure 6.** Value of root mean square error (RMSE) versus number of hidden units (NHU) in different applied K.

In connection with the number of iterations and running time for model training in the optimal scenarios, model 2, because of the lowest TLP value, can learn sub-optimal weight sets more quickly (2000 reiterations in 71 s). Nonetheless, model 3, for the maximum TLP value, can learn further and even globally optimum weight sets, and, consequently, it takes too long to train (2000 reiterations in 102 s).

## 6. Selection of the best model

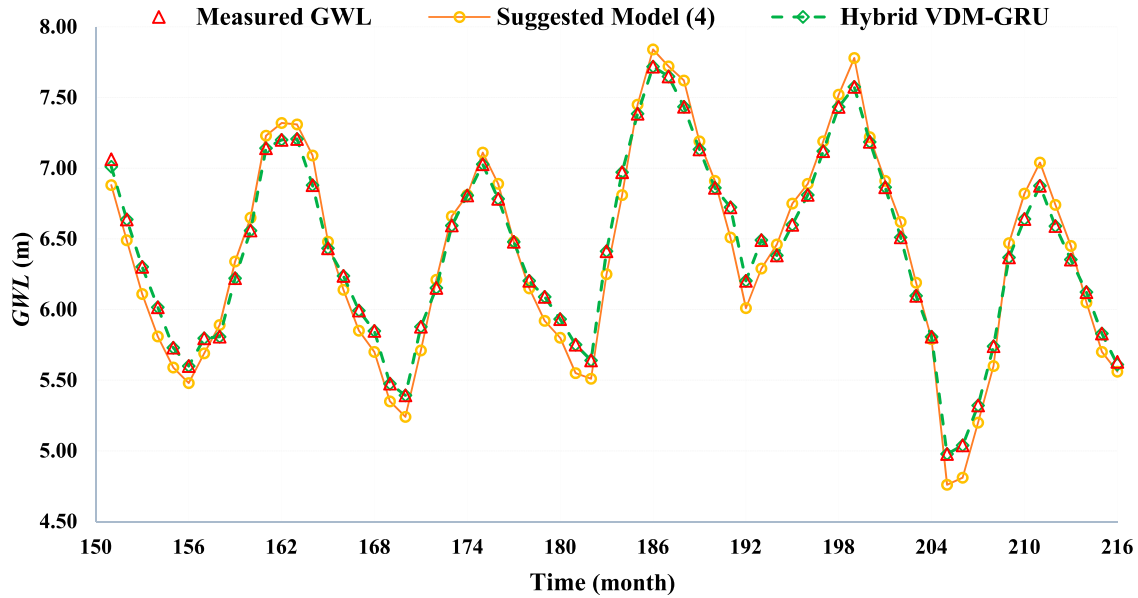
In this research, to select the best model, taking into account the effect of running time for the model training scale, the AICc and model grading process are performed.

Because the applied models have different layers, their tuned parameters are not similar. Accordingly, it is necessary to equate the performance of the developed models not only in terms of accuracy but also in terms of simplicity. In this regard, the AICc is usually used to scrutinize and compare the developed models by considering the precision and intricacy of the model concurrently (Zeynoddin et al., 2019). The AICc index is defined as follows (Ebtehaj et al., 2019):

$$\text{AICc} = \frac{(n \ln(\sigma_{\varepsilon}^2)(n - k - 1)) + 2kn}{n - k - 1} \quad (14)$$

where  $n$  and  $k$  are the numbers of months and parameters, respectively, and  $\sigma_{\varepsilon}$  is the standard deviation of the residuals.

The model grading procedure presented by Vahedoost et al. (2016) is used for a complementary statistical judgment. In this method, the success grade (SG) and failure grade (FG) of performance significant parameters



**Figure 7.** Comparison of the predicted and measured groundwater level (GWL) by model 4 and the hybrid variational mode decomposition–gated recurrent unit (VMD-GRU) model under optimal hyperparameters in the testing phase (66 months between October 2014 and March 2020).

are expressed as follows, along with the NSE:

$$NSE = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \mu_x)^2} \tag{15}$$

$$SG(AICc) = \frac{AICc_i}{AICc_{min}} \times 10 \tag{16}$$

$$SG(NSE) = \frac{NSE_i}{NSE_{max}} \times 10 \tag{17}$$

$$FG(Time) = \frac{Time_i}{Time_{max}} \times 10 \tag{18}$$

$$FG(RMSE) = \frac{RMSE_i}{RMSE_{max}} \times 10 \tag{19}$$

The total grade (TG) of each model is obtained by adding the SG and FG of each model separately; the TG ranges between  $-20$  and  $+20$ , and is defined as

$$\begin{aligned} \text{Total grade} &= SG(NSE) + SG(AICc) \\ &\quad - FG(Time) - FG(RMSE) \end{aligned} \tag{20}$$

The values of AICc and TG obtained by Equations (14)–(20) are given in Table 6. Figure 7 shows a graphical comparison of the observed and forecast time-series GWL by model 4 and the hybrid VMD-GRU model in the testing phase.

Examination of Figure 7 validates that the hybrid VMD-GRU model can remember the previous measured time-series GWL and capture the variational trend, especially in the bottom and peak points.

**Table 6.** Comparison of the optimal developed models in the testing stage and related grades.

| Model          | $R^2$ | AICc    | NSE  | Time | RMSE | TG    |
|----------------|-------|---------|------|------|------|-------|
| 1              | 0.76  | -208.32 | 0.86 | 139  | 0.21 | -0.5  |
| 2              | 0.75  | -197.78 | 0.85 | 71   | 0.22 | 2.31  |
| 3              | 0.68  | -158.37 | 0.81 | 102  | 0.24 | -1.89 |
| 4              | 0.86  | -280.75 | 0.89 | 87   | 0.18 | 6.21  |
| Hybrid VMD-GRU | 0.92  | -310.52 | 0.92 | 185  | 0.16 | 3.34  |

Note: VMD-GRU = variational mode decomposition–gated recurrent unit; AICc = corrected Akaike’s information criterion; NSE = Nash–Sutcliffe efficiency; RMSE = root mean square error; TG = total grade.

According to Table 6, where the hybrid VMD-GRU model has the lowest AICc and RMSE, combined with the highest  $R^2$  and NSE values, by discarding the effect of running time criteria in the selection process of the best method, this hybrid model may be considered the most precise technique for predicting the mean monthly time-series GWL. However, considering the effect of the running time, because the highest TG value is found for the GRU2× model (model 4), it is considered to be the best method. All in all, based on Table 6 and Figure 7, since the hybrid VMD-GRU model is slightly more accurate than model 4, the authors prefer model 4 as the optimum model for forecasting GWL. Despite the superiority of model 4, it has some limitations, such as the need for continuous measurement of different water table depths and the high number of monitoring piezometers required over a long period to forecast exactly the average monthly GWL in the study region. Furthermore, as GWL is not an unremitting stationary multifactor time series

with a lagged response, forecasting the GWL for farming purposes is often expected to be time-consuming.

## 7. Conclusions

In this study, the time-series GWL was predicted in Qoşaçay plain, an arable region in Qoşaçay city, West Azerbaijan Province, Northwest Iran. For this purpose, 216 monthly measured water table depths of 33 monitoring piezometers in the long period between April 2002 and March 2020 were employed. For modeling, three different-layer GRU-based neural network models, along with the hybrid VMD-GRU model, were developed. To compare the performance of these models, the general single-LSTM-layer network model was developed as the base model. For better configuration and reduced overfitting effects, algorithm tuning was performed using various types of activation function combinations,  $P$ -rate and NHU as hyperparameters, along with a trial-and-error procedure. The main conclusions of this study are as follows:

- (1) The general single-LSTM-layer network model (model 1) was very slightly more precise than the general GRU-layer network model (model 2), but, as it required more NHU, it took too long to reach convergence in comparison with model 2. Also, in both models 1 and 2,  $\tanh$  was selected as the most suitable and fastest type of SAF.
- (2) In the calibration phase, all developed models were more precise than in the validation stage. As well, for the same NHU, by increasing  $P$ -rate values, the running time for model training decreased. However, for the same  $P$ -rate, by increasing the value of NHU, the running time increased.
- (3) After several experiments, the *softsign-tanh* combination in SAF was determined to be an optimal assemblage in both the GRU2 model (model 3) and GRU2 $\times$  model (model 4). In addition, model 4 was noticeably faster than model 3.
- (4) The hybrid VMD-GRU model under the optimal hyperparameter values ( $P$ -rate = 0.5, NHU = 128,  $K = 3$ ) and  $\tanh$  as the most suitable type of SAF resulted in an  $R^2$  of 0.92 and an RMSE of 0.16 m. This hybrid model was a little more precise than the GRU2 $\times$  model, but it took too long to reach convergence compared to all of the developed models.
- (5) Despite the maximum value of TLP being obtained by model 3, because of the highest value of TG in the GRU2 $\times$  model, the latter is preferred as the best model. As an important consequence, it can be deduced that to achieve an operative model, the

optimal NTP, NHU and TLP must be used in the emulation process.

The newly proposed GRU-based structure (model 4) is different from typical neural network models and has not been previously utilized in the field of hydrology. It can be viewed as an accurate DL-based structure and a robust intelligent model since its performance was verified by statistical metrics. In contrast to other models, it can significantly reduce the computational costs and is therefore cheaper to run. Moreover, owing to the innovative framework and construction of the model (model 4), which was developed in environments lacking records on meteorological parameters, it can be straightforwardly used in various climatic regions around the world. Also, for the same hyperparameters, its innovative and suitable layer network structure resulted in a suitable TLP value among the designed GRU-based models. Although the current study evaluated the ability of different structures of the GRU-based model and hybrid VMD-GRU model in forecasting GWL, future investigations could be performed using other types of models by hybridizing DNN methods with optimization algorithms such as the genetic algorithm or PSO. The relevant outcomes could be equated with the results of this research so that the best AI approaches can be determined. The authors also strongly recommend predicting GWL in an environment in which meteorological and hydrological data records are available, to investigate the effect of these parameters on the ability and precision of the designed models. Finally, the authors advocate using the TLP factor as a useful measure to judge the predictive ability, overfitting conditions and performance of DL-based models, since it can specify the functioning dimensions.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Amin Gharehbaghi  <http://orcid.org/0000-0002-2898-3681>  
Kwok-Wing Chau  <http://orcid.org/0000-0001-6457-161X>

## References

- Ahmadi, M. H., Baghban, A., Sadeghzadeh, M., Zamen, M., Mosavi, A., Shamshirband, S., ... Mohammadi-Khanaposh tani, M. (2020). Evaluation of electrical efficiency of photovoltaic thermal solar collector. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 545–565. <https://doi.org/10.1080/19942060.2020.1734094>
- Anderson, M. P., Woessner, W. W., & Hunt, R. J. (1992). Applied groundwater modeling: Simulation of flow and advective transport. *Journal of Hydrology*, 140(1-4), 393–394. [https://doi.org/10.1016/0022-1694\(92\)90251-P](https://doi.org/10.1016/0022-1694(92)90251-P)

- Asadi, E., Isazadeh, M., Samadianfard, S., Ramli, M. F., Mosavi, A., Nabipour, N., ... Chau, K. W. (2020). Groundwater quality assessment for sustainable drinking and irrigation. *Sustainability*, 12(1), 177–190. <https://doi.org/10.3390/su12010177>.
- Azari, A., Zeynoddin, M., Ebtehaj, I., Sattar, A., Gharabaghi, B., & Bonakdari, H. (2021). Integrated preprocessing techniques with linear stochastic approaches in groundwater level forecasting. *Acta Geophysica*, 69(4), 1395–1411. <https://doi.org/10.1007/s11600-021-00617-2>
- Azizpour, A., Izadbakhsh, M. A., Shabanlou, S., Yosefvand, F., & Rajabi, A. (2021). Estimation of water level fluctuations in groundwater through a hybrid learning machine. *Groundwater for Sustainable Development*, 15(5), 100687–100699. <https://doi.org/10.1016/j.gsd.2021.100687>.
- Bahmani, R., & Ouarda, T. B. (2021). Groundwater level modeling with hybrid artificial intelligence techniques. *Journal of Hydrology*, 595, 125659. <https://doi.org/10.1016/j.jhydrol.2020.125659>
- Bahmani, R., Solgi, A., & Ouarda, T. (2020). Groundwater level simulation using gene expression programming and M5 model tree combined with wavelet transform. *Hydrological Sciences Journal*, (7), 66–78. <https://doi.org/10.1080/0262667.2020.1749762>.
- Bai, Y., Liu, M. D., Ding, L., & Ma, Y. J. (2021). Double-layer staged training echo-state networks for wind speed prediction using variational mode decomposition. *Applied Energy*, 301(6), 117461. <https://doi.org/10.1016/j.apenergy.2021.117461>.
- Band, S. S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakraborty, R., Shokri, M., & Mosavi, A. (2020). Novel ensemble approach of deep learning neural network (DLNN) model and particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility. *Sensors*, 20(19), 5609–5633. <https://doi.org/10.3390/s20195609>.
- Bekesi, G., McGuire, M., & Moiler, D. (2009). Groundwater allocation using a groundwater level response management method – Gngangara groundwater system, Western Australia. *Water Resources Management*, 23(9), 1665–1683. <https://doi.org/10.1007/s11269-008-9346-5>
- Bowes, B. D., Sadler, J. M., Morsy, M. M., Behl, M., & Goodall, J. L. (2019). Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks. *Water*, 11(5), 1098–1105. <https://doi.org/10.3390/w11051098>.
- Chang, X., Liu, S., & Li, X. (2016). Modified alternating direction method of multipliers for convex quadratic semidefinite programming. *Neurocomputing*, 214, 575–586. <https://doi.org/10.1016/j.neucom.2016.06.043>
- Chen, Y., Liu, G., Huang, X., Chen, K., Hou, J., & Zhou, J. (2021). Development of a surrogate method of groundwater modeling using gated recurrent unit to improve the efficiency of parameter auto-calibration and global sensitivity analysis. *Journal of Hydrology*, 598, 125726. <https://doi.org/10.1016/j.jhydrol.2020.125726>
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. <https://doi.org/10.3115/v1/D14-1179>.
- Choubin, B., Mosavi, A., Alamdarloo, E. H., Hosseini, F. S., Shamshirband, S., Dashtekian, K., & Ghamisi, P. (2019). Earth fissure hazard prediction using machine learning models. *Environmental Research*, 179(6), 108770. <https://doi.org/10.1016/j.envres.2019.108770>.
- Conlon, T. D., Wozniak, K. C., Woodcock, D., Herrera, N. B., Fisher, B. J., Morgan, D. S., ... Hinkle, S. R. (2005). Groundwater hydrology of the Willamette Basin, Oregon. US Geological Survey.
- Dalkilic, H. Y., & Gharehbaghi, A. (2021). Numerical modeling of groundwater flow based on explicit and fully implicit schemes of finite volume method. *Journal of Engineering Research*, 9(4B), 535–566. <https://doi.org/10.36909/jer.9253>.
- Dragomiretskiy, K., & Zosso, D. (2013). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544. <https://doi.org/10.1109/TSP.2013.2288675>
- EARWO (East Azerbaijan Regional Water Organization). (2020). Preparation of water balance and water cycle in the Malekan region. 56p.
- Ebtehaj, I., Bonakdari, H., & Gharabaghi, B. (2019). A reliable linear method for modeling lake level fluctuations. *Journal of Hydrology*, 570, 236–250. <https://doi.org/10.1016/j.jhydrol.2019.01.010>
- Fang, W., Huang, S., Ren, K., Huang, Q., Huang, G., Cheng, G., & Li, K. (2019). Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. *Journal of Hydrology*, 568, 534–550. <https://doi.org/10.1016/j.jhydrol.2018.11.020>
- Fathian, F., Modarres, R., & Dehghan, Z. (2016). Urmia Lake water-level change detection and modeling. *Modeling Earth Systems and Environment*, 2(4), 1–16. <https://doi.org/10.1007/s40808-016-0253-0>
- Favero, L., Mattiuzzo, E., & Franco, D. (2007). Practical results of a water budget estimation for a constructed wetland. *Wetlands*, 27(2), 230–239.
- Fosso, O. B., & Molinas, M. (2018, October). EMD mode mixing separation of signals with close spectral proximity in smart grids. In 2018 IEEE PES innovative smart grid technologies conference Europe (ISGT-Europe) (pp. 1–6). IEEE. doi:10.1109/ISGTEurope.2018.8571816.
- Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC) (pp. 324–328). IEEE. doi:10.1109/YAC.2016.7804912.
- Fu, W., Wang, K., Tan, J., & Shao, K. (2020). Vibration tendency prediction approach for hydropower generator fused with multi-scale dominant ingredient chaotic analysis, adaptive mutation grey wolf optimizer and KELM. *Complexity*, 2020(5), 1–20.
- Fu, W., Wang, K., Zhang, C., & Tan, J. (2019). A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine. *Transactions of the Institute of Measurement and Control*, 41(15), 4436–4449. <https://doi.org/10.1177/0142331219860279>.
- Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology*, 589(8), 125188. <https://doi.org/10.1016/j.jhydrol.2020.125188>.
- Gehrels, J. C., Van Geer, F. C., & De Vries, J. J. (1994). Decomposition of groundwater level fluctuations using



- transfer modelling in an area with shallow to deep unsaturated zones. *Journal of Hydrology*, 157(1-4), 105–138. [https://doi.org/10.1016/0022-1694\(94\)90101-5](https://doi.org/10.1016/0022-1694(94)90101-5)
- Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Continual prediction using LSTM with forget gates. In *Neural nets WIRN Vietri-99* (Vol. 9, pp. 133–138). Springer. <https://doi.org/10.1007/978-1-4471-0877-1-10>.
- Ghalandari, M., Ziamolki, A., Mosavi, A., Shamshirband, S., Chau, K. W., & Bornassi, S. (2019). Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments. *Engineering Applications of Computational Fluid Mechanics*, 13(1), 892–904. <https://doi.org/10.1080/19942060.2019.1649196>
- Gharehbaghi, A. (2021a). Fully implicit form of differential quadrature method for multi-species solute transport in porous media. *Teknik Dergi*, 33(4), 655–6665. <https://doi.org/10.18400/tekderg.975457>.
- Gharehbaghi, A. (2021b). A numerical model with finite difference schemes for multi-species solute transport in porous media. *Journal of Engineering Research*. <https://doi.org/10.36909/jer.13705>.
- Ghasemlouia, R., Gharehbaghi, A., Ahmadi, F., & Saadatnejadgharahassanlou, H. (2021). Developing a novel framework for forecasting groundwater level fluctuations using bi-directional long short-term memory (BiLSTM) deep neural network. *Computers and Electronics in Agriculture*, 191, 106568. <https://doi.org/10.1016/j.compag.2021.106568>
- Golestani Kermani, S., Sayari, S., Kisi, O., & Zounemat-Kermani, M. (2019). Comparing data driven models versus numerical models in simulation of waterfront advance in furrow irrigation. *Irrigation Science*, 37(5), 547–560. <https://doi.org/10.1007/s00271-019-00635-5>
- Gómez-Hernández, J. J., & Gorelick, S. M. (1989). Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge. *Water Resources Research*, 25(3), 405–419. <https://doi.org/10.1029/WR025i003p00405>
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. doi: 10.1109/TNNLS.2016.2582924
- He, F., Zhou, J., Feng, Z. K., Liu, G., & Yang, Y. (2019). A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Applied Energy*, 237, 103–116. <https://doi.org/10.1016/j.apenergy.2019.01.055>
- Hemker, K., & Bakker, M. (2006). Analytical solutions for whirling groundwater flow in two-dimensional heterogeneous anisotropic aquifers. *Water Resources Research*, 42(12), 55–65. <https://doi.org/10.1029/2006WR004901>.
- Henriksen, H. J., Trolldborg, L., Højberg, A. L., & Refsgaard, J. C. (2008). Assessment of exploitable groundwater resources of Denmark by use of ensemble resource indicators and a numerical groundwater–surface water model. *Journal of Hydrology*, 348(1-2), 224–240. <https://doi.org/10.1016/j.jhydrol.2007.09.056>
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jalilvand, E., Tajrishy, M., Hashemi, S. A. G. Z., & Brocca, L. (2019). Quantification of irrigation water using remote sensing of soil moisture in a semi-arid region. *Remote Sensing of Environment*, 231(6), 111226. <https://doi.org/10.1016/j.rse.2019.111226>
- Jeong, J., & Park, E. (2019). Comparative applications of data-driven models representing water table fluctuations. *Journal of Hydrology*, 572, 261–273. <https://doi.org/10.1016/j.jhydrol.2019.02.051>
- Jeong, J., Park, E., Chen, H., Kim, K. Y., Han, W. S., & Suk, H. (2020). Estimation of groundwater level based on the robust training of recurrent neural networks using corrupted data. *Journal of Hydrology*, 582, 124512–124522. <https://doi.org/10.1016/j.jhydrol.2019.124512>.
- Joloudari, J. H., Hassannataj Joloudari, E., Saadatfar, H., Ghasemigol, M., Razavi, S. M., Mosavi, A., ... Nadai, L. (2020). Coronary artery disease diagnosis; ranking the significant features using a random trees model. *International Journal of Environmental Research and Public Health*, 17(3), 731–750. <https://doi.org/10.3390/ijerph17030731>.
- Kovács, J., Márkus, L., Szalai, J., & Kovács, I. S. (2015). Detection and evaluation of changes induced by the diversion of River Danube in the territorial appearance of latent effects governing shallow-groundwater fluctuations. *Journal of Hydrology*, 520, 314–325. <https://doi.org/10.1016/j.jhydrol.2014.11.052>
- Lawrence, S., Back, A. D., Tsoi, A. C., & Giles, C. L. (1997). On the distribution of performance from multiple neural-network trials. *IEEE Transactions on Neural Networks*, 8(6), 1507–1517. <https://doi.org/10.1109/72.641472>
- Li, H., Bao, T., Gu, C., & Chen, B. (2019). Vibration feature extraction based on the improved variational mode decomposition and singular spectrum analysis combination algorithm. *Advances in Structural Engineering*, 22(7), 1519–1530. <https://doi.org/10.1177/1369433218818921>.
- Li, J., Yao, X., Wang, H., & Zhang, J. (2019). Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 126(6), 568–589. <https://doi.org/10.1016/j.ymsp.2019.02.056>.
- Mahmoudi, M. R., Heydari, M. H., Qasem, S. N., Mosavi, A., & Band, S. S. (2021). Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. *Alexandria Engineering Journal*, 60(1), 457–464. <https://doi.org/10.1016/j.aej.2020.09.013>
- Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, 25(8), 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- MATLAB. (2021). MATLAB 2021a user's guide (Deep Learning Toolbox). The MathWorks Inc. Computer software. [www.mathworks.com/](http://www.mathworks.com/)

- Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., & Band, S. S. (2020). Comprehensive review of deep reinforcement learning methods and applications in economics. *Mathematics*, 8(10), 1640–1650. <https://doi.org/10.3390/math8101640>.
- Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>
- Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *IEEE Access*, 8, 150199–150212. <https://doi.org/10.1109/ACCESS.2020.3015966>
- Naik, J., Dash, S., Dash, P. K., & Bisoi, R. (2018). Short term wind power forecasting using hybrid variational mode decomposition and multi-kernel regularized pseudo inverse neural network. *Renewable Energy*, 118, 180–212. <https://doi.org/10.1016/j.renene.2017.10.111>
- Nechikkat, N., Sowmya, V., & Soman, K. (2015). A comparative analysis of variational mode and empirical mode features on hyperspectral image classification. *International Journal of Applied Engineering Research [IJAER]*, 10(73), 95–98.
- Niu, M., Hu, Y., Sun, S., & Liu, Y. (2018). A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. *Applied Mathematical Modelling*, 57, 163–178. <https://doi.org/10.1016/j.apm.2018.01.014>
- Norouzi, H., & Moghaddam, A. A. (2020). Groundwater quality assessment using random forest method based on groundwater quality indices (case study: Miandoab plain aquifer, NW of Iran). *Arabian Journal of Geosciences*, 13(18), 1–13. <https://doi.org/10.1007/s12517-020-05904-8>
- Nourani, V., & Mousavi, S. (2016). Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method. *Journal of Hydrology*, 536, 10–25. <https://doi.org/10.1016/j.jhydrol.2016.02.030>
- Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., & Chen, C. H. (2020). Water level prediction model based on GRU and CNN. *IEEE Access*, 8, 60090–60100. <https://doi.org/10.1109/ACCESS.2020.2982433>
- Post, V. E., & von Asmuth, J. R. (2013). Hydraulic head measurements – New technologies, classic pitfalls. *Hydrogeology Journal*, 21(4), 737–750. <https://doi.org/10.1007/s10040-013-0969-0>
- Qasem, S. N., Samadianfard, S., Sadri Nahand, H., Mosavi, A., Shamshirband, S., & Chau, K. W. (2019). Estimating daily dew point temperature using machine learning algorithms. *Water*, 11(3), 582. <https://doi.org/10.3390/w11030582>.
- Rajaei, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, 572, 336–351. <https://doi.org/10.1016/j.jhydrol.2018.12.037>
- Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T. (2018). Deep state space models for time series forecasting. *Advances in Neural Information Processing Systems*, 31(8), 344–450.
- Rezaie-balf, M., Naganna, S. R., Ghaemi, A., & Deka, P. C. (2017). Wavelet coupled MARS and M5 model tree approaches for groundwater level forecasting. *Journal of Hydrology*, 553, 356–373. <https://doi.org/10.1016/j.jhydrol.2017.08.006>
- Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128.
- Schuermans, J. M., Bierkens, M. F. P., Pebesma, E. J., & Uijlenhoet, R. (2007). Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar. *Journal of Hydrometeorology*, 8(6), 1204–1224. <https://doi.org/10.1175/2007JHM792.1>
- Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., & Várkonyi-Kóczy, A. R. (2020). Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*, 11(1), 66–77. <https://doi.org/10.3390/atmos11010066>.
- Sharafati, A., Asadollah, S. B. H. S., & Neshat, A. (2020). A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. *Journal of Hydrology*, 591, 125468–125480. <https://doi.org/10.1016/j.jhydrol.2020.125468>.
- Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 54(11), 8558–8593. <https://doi.org/10.1029/2018WR022643>
- Shiri, J., Kisi, O., Yoon, H., Lee, K. K., & Nazemi, A. H. (2013). Predicting groundwater level fluctuations with meteorological effect implications – A comparative study among soft computing techniques. *Computers & Geosciences*, 56, 32–44. <https://doi.org/10.1016/j.cageo.2013.01.007>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K. B. W., Alizamir, M., Shamshirband, S., & Chau, K. W. (2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Engineering Applications of Computational Fluid Mechanics*, 12(1), 738–749. <https://doi.org/10.1080/19942060.2018.1526119>
- Tao, H., Hameed, M. M., Marhoon, H. A., Zounemat-Kermani, M., Salim, H., Sungwon, K., ... Yaseen, Z. M. (2022). Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing*, 7(22), 133–150. <https://doi.org/10.1016/j.neucom.2022.03.014>
- Torabi, M., Hashemi, S., Saybani, M. R., Shamshirband, S., & Mosavi, A. (2019). A hybrid clustering and classification technique for forecasting short-term energy consumption. *Environmental Progress & Sustainable Energy*, 38(1), 66–76. <https://doi.org/10.1002/ep.12934>
- Vaheddoost, B., Aksoy, H., & Abghari, H. (2016). Prediction of water level using monthly lagged data in Lake Urmia, Iran. *Water Resources Management*, 30(13), 4951–4967. <https://doi.org/10.1007/s11269-016-1463-y>
- Valdes, D., Dupont, J. P., Laignel, B., Slimani, S., & Delbart, C. (2014). Infiltration processes in karstic chalk investigated through a spatial analysis of the geochemical properties of the groundwater: The effect of the superficial layer of clay-with-flints. *Journal of Hydrology*, 519, 23–33. <https://doi.org/10.1016/j.jhydrol.2014.07.002>
- Vu, M. T., Jardani, A., Massei, N., & Fournier, M. (2021). Reconstruction of missing groundwater level data by using long

- short-term memory (LSTM) deep neural network. *Journal of Hydrology*, 597, 125776. <https://doi.org/10.1016/j.jhydrol.2020.125776>.
- Wang, Y., Markert, R., Xiang, J., & Zheng, W. (2015). Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system. *Mechanical Systems and Signal Processing*, 60-61, 243–251. <https://doi.org/10.1016/j.ymssp.2015.02.020>
- Wu, C., Zhang, X., Wang, W., Lu, C., Zhang, Y., Qin, W., ... Shu, L. (2021). Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model. *Science of The Total Environment*, 783(6), 146948. <https://doi.org/10.1016/j.scitotenv.2021.146948>.
- Yin, J., Deng, Z., Ines, A. V., Wu, J., & Rasu, E. (2020). Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). *Agricultural Water Management*, 242(7), 106386–1063295. <https://doi.org/10.1016/j.agwat.2020.106386>.
- Zare, M., & Koch, M. (2018). Groundwater level fluctuations simulation and prediction by ANFIS-and hybrid wavelet-ANFIS/fuzzy C-means (FCM) clustering models: Application to the Miandarband plain. *Journal of Hydro-Environment Research*, 18, 63–76. <https://doi.org/10.1016/j.jher.2017.11.004>
- Zeynoddin, M., Bonakdari, H., Ebtehaj, I., Esmaeilbeiki, F., Gharabaghi, B., & Haghi, D. Z. (2019). A reliable linear stochastic daily soil temperature forecast model. *Soil and Tillage Research*, 189, 73–87. <https://doi.org/10.1016/j.still.2018.12.023>
- Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, 561, 918–929. <https://doi.org/10.1016/j.jhydrol.2018.04.065>