GROUNDWATER LEVEL PREDICTIONS USING HYBRID WAVELET MACHINE LEARNING METHODS

[1]LEYLA GHASEMI, [2]MEYSAM VADIATI

[1]Office of Sustainability, Gamasiab Consulting Engineers Company, Tehran, Iran
[2]Hubert H. Humphrey Fellowship Program, Global Affairs, University of California, Davis, 10 College Park, 95616 Davis, CA, USA
E-mail: [1]le.ghasemi81@gmail.com, [2]mvadiati@ucdavis.edu

Abstract - Groundwater level (GWL) prediction is fundamental for agriculture and water distribution. Machine learning (ML) models,such as artificial neural networks (ANN), have been gradually moreacceptedby scholars to predict GWL becauseof theircapability to simulate the nonlinearities between GWL and its drivers (e.g.,precipitation). The present researchsuggested a traditionalsupervised MLmethod to describe the GWL variationsin the Tehran-Karaj plain in Iran using 14 years monthly interval dataset.The wavelet transform (WT) method was also employedto advance the GWL simulationcapability of ML approaches for three monthsaheadusing various input combinations of temperature, precipitation, evapotranspiration and GWL. The methods were compared based on different error criteria, andthe comparison of the results proved thatthe hybrid-wavelet MLsubstantiallyadvanced the standalone model outcomes. The best GWL results were acquired from the hybrid WT-ANNmodel results forscenario one,and this model produced MAE, RMSE, NSE, and R as 0.20, 0.25, 0.99 and 0.99 for one month ahead of GWL prediction.The results of the two models were similar in predicting groundwater level fluctuations one month ahead of the GWL simulation.

Keywords - Artificial Intelligence, Groundwater Level Forecasting, Wavelet Transform, Standalone Model.

I. INTRODUCTION

The effects of climate change on groundwater have been obtainingimproving international concernrecently, as observed in scientific reviews (Jyrkama and sykes, 2016; Al Atawneh et al., 2021; Hssaisoune et al. 2020),flooding, seawater intrusion and deteriorating groundwater-dependent ecosystems.Forecasting Groundwater level(GWL) in aquifers isessential for the sensible management of groundwater resources.

MLmodels have enduredswiftexpansion in hydrology, and numerousresearchershave employedMLto evaluate the potentialof ML models (Karimi et al. 2019; Roshni et al. 2020; Liu et al. 2020; Su et al. 2020; Azizpour et al. 2022).ML methods were also used for hydrogeological evaluation and GWL prediction in different studies (Almuhaylan et al. 2020; Moghaddam et al. 2019; Chakraborty 2020; Rezaei et al. 2021).

The WTmethodis a promising tool for multitemporal hydrologic time-series evaluation (Zhu et al., 2020;Gordu and Nachabe, 2021).The chief contributions of the present study are: (i) comparing the capability of standardstandalone ML models for prediction of three-time step-ahead GWL in the Tehran-Karajplain; (ii) developingWL artificialmodels (WT–ANN)to advance the ability of MLmethods, and (iii) integrating monthly temperature, evapotranspiration, precipitation and GWL variables to evaluate the performance of the five standard standalone and hybrid ML models in GWL prediction.

II. STUDY AREA

The techniques are employedfor the data of the Tehran-Karaj plain.The Tehran-Karaj Plain, with a mean annual 432.5 mm precipitation, is mainly located inthe Tehran province.(Fig.1).Tehran-Karaj area comprises about5083 km^2, which is about2519 km^2 of the region.

The aquifer has been identified in plain Tehran-Karaj. The climate of the north section is semi-cold, and increasingly the climateconvertsto semi-arid by shifting toward the center and south. High population density and different water suppliesdecrease groundwater levels in the Tehran-Karaj aquifer and make the region vulnerable. There is no balance between groundwater withdrawal and recharge due to uncontrolled groundwater exploration.The trend of falling groundwater levels continues.

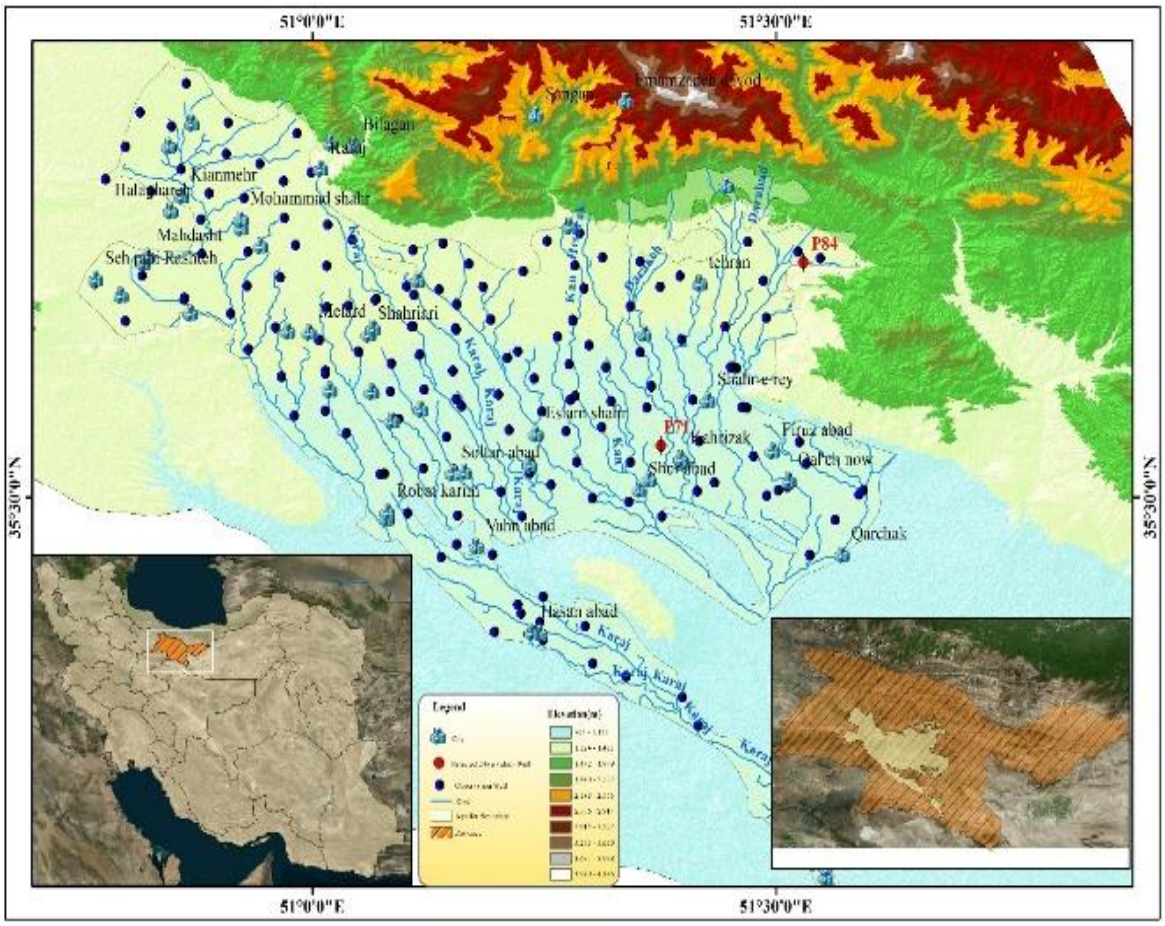


Fig. 1. Geographical map of the Tehran-Karaj area.

III. MATERIALS AND METHOD

3.1. Artificial Neural Network (ANN)

The basic concept of Artificial Neural Network (ANN) modeling, one of the well-accepted ML models, is constructed on a simplified imitation of calculationimplemented by human neurons. It bears a resemblance to the brain in two main viewpoints: (a) knowledge is attained by the learning process, and (b) interneuron connection strengths are employed to accumulate the acquired knowledge (Haykin, 2004).Single-layer perceptron (SLP) or multi-layer perceptron (MLP) are two standard ANN algorithms. Many have recommended the reasonable results of the SLP methodto estimatecomplicated nonlinear hydrological events. However, nowadays, the most common variety of the ANN model is the MLP network (Monjezi et al., 2014). This study used a standard ANN model, i.e., MLP, developed initially by McClelland et al. (1987).TheMATLABsoftware was applied to advanceall ML models.

3.2. Wavelet Transformation (WT)

The wavelet transform has been broadlyutilized in time series de-noising due to its remarkable time-frequency interpretationability(Graf et al. 2019).The conjunctiveWT–ML methodspursuetwo principalprocedures:applying the WT to preprocess the time series dataset and miningfeatures from the primary signal to feedtheML model as an input dataset. Typically, a trial-and-errormethod is utilized to verify the standard mother waveleton the basis of the appearancecomparison of the mother wavelet and centralinput signal. Daubechies-1, Daube-chies-4, and Haarmother wavelets are in generalcompetentin creating time localization characteristics for time series(Nourani et al. 2014).

3.3. Coupled Wavelet and ML(WT–ML) Models

The potential of WT–ML methods to classify data components in a time series is asignificantbenefit compared to standalone techniques (Adamowski and Sun 2010).The WTdecomposes the GWL and climatictime signals high- and low-frequency sub-signals. To attain a precise and accurate GWL simulation, signal analysis is essential to essential abstractdata such as stochastic and trendterms (Moeeni et al. 2017).Hence, the WT isemployed in this research to preprocess the decomposed time series and the ML methods to predict the nonlinear GWL.All utilizedmethods have been implemented in the MATLAB software.Each signal (GWL and meteorological) was decomposed intotwo central sub-signals (approximation and details)to predict the GWL. Hence,the Wavelet-ML methods have data set inputs containingapproximation (A component) and

details (D component) sub-signals.The trial-and-errormethodwasapplied to distinguish the decomposition levelcorresponding to the two. Therefore, each input time seriesdeemedtopredict GWL isdivided into three sub-series, A2, D1 and D2.The db4 mother wavelet was chosen in the presentresearchfor the WT procedure.

3.4. Evaluation Statistics

When the testing and training set isapplied to adjust themodel configuration, the outcome of differentmodels can be evaluated according to the degree of data fitting.The metrics of the root mean square errors (RMSE), mean absolute error (MAE), Nash-Sutcliffe criterion (NASH) and Coefficient of correlation (R)are used as performance indicators for analysis, as shown below:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|H_{pi} - H_{oi}\right| \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(H_{pi} - H_{oi}\right)^2}{N}} \quad (2)$$

$$R = \frac{\left(\sum_{i=1}^{N}\left(H_{oi} - \overline{H_o}\right)\left(H_{pi} - \overline{H_p}\right)\right)^2}{\sum_{i=1}^{N}\left(H_{oi} - \overline{H_o}\right)^2 \sum_{i=1}^{N}\left(H_{pi} - \overline{H_p}\right)^2} \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(H_{oi} - H_{pi}\right)^2}{\sum_{i=1}^{N}\left(H_{oi} - \overline{H_o}\right)^2} \quad (4)$$

H_{oi} is the observed GWL, H_{pi} is the predicted GWL, $\overline{H_O}$ is the average observed GWL, $\overline{H_P}$ is the mean GWL simulation,and N is the number of observations.

IV. RESULTS AND DISCUSSION

4.1. Model Development

Assessing the importance of each input time series on the target time series prior to simulating is essential. Potential input signalsaffecting the GWLconsist of precipitation (P), temperature (T), and evapotranspiration (ET).The present study's monthly time scale waschosen to forecast groundwater levels.

This study selected two observation wells that characterizemost of the aquifersin the study area and the aquifer system's hydrogeological characteristics.The Tehran Regional Water Company issued this data set.For the models employed in this study, 168-time steps (month) were chosen as the model period. Then, for simulation purposes, the input datasetis divided into two parts. The first part comprises 70% of the time series and isemployed as input data set to adjust the ML models to simulateGWL one month ahead.Then, the followingsection, which comprises 30% of the data set, is usedto validate the models.

4.2. Scenarios Determination

This research used and evaluated WT and standalone methodstopredict GWL in Tehran-Karaj Plain by considering the input dataset. Variousscenarioswere developed with different lag times from the one-month "GWLt-1" to three months prior "GWLt-3" to simulate the GWL for various lead times (one to three months ahead) (Table1).

In the present study,perfectscenarios were specificbased ondata relationshipsbetween the inputs and groundwater level.The groundwater level (H), monthly temperature average (T) in C, total monthly precipitation (P) in mm, total monthly evapotranspiration (ET) in mm, the average of monthly stream discharge (Q) in cubic meters per day, and a maximum of three-time delays for GWL, were considered as model inputs.Precipitation, evapotranspiration and temperature are selected because they directly influence groundwater levels (Rezaie-balf etal., 2017).The groundwater level,temperature,precipitation, evapotranspiration andriver discharge signals are decomposed to sub-signals using DWT for various decomposition levels. Next, the sub-signals are employed as inputs to the ML models to generate hybrid models for one to three months ahead of GWL prediction (Table2).

Entry	input
1	H_{t-1} ,H_{t-2} ,H_{t-3}
2	H_{t-1}, T_t, ET_t, P_t, Q_t
3	H_{t-1}, H_{t-2}, T_t, ET_t, P_t, Q_t
4	H_{t-1} ,H_{t-2} ,$H_{t-3,}$ T_t ,ET_t ,P_t ,Q_t
5	H_{t-1} ,H_{t-2} ,$H_{t-3,}$ T_t ,ET_t ,P_t ,Q_t ,T_{t-1} ,ET_{t-1}

Table 1. Different scenarios for standalone models.

4.3. Standalone Model Performance

This subdivisiondescribes the accuracy and precision of the five most common standalone models, ANN,in five scenarios, for one, two and three months ahead GWL simulationevaluatedbased ondifferent goodness of fit statistics.Applying the ANN methoddemonstrates that the optimalframework for the study area is the FNN structure and the Levenberq-Marquardt algorithm—the number of neurons and hidden layers in each hidden layer estimated by the trial-and-error technique.

Entry	input
1	)Decomposed in levels 1, 2(H_{t-1} ,H_{t-2} ,H_{t-3}
2	H_{t-1}, T_t, E_t, P_t, Q_t (H_{t-1} Decomposed in levels 1, 2)
3	H_{t-1}, H_{t-2}, P_t, Q_t (H_{t-1}, H_{t-2}Decomposed in levels 1, 2)
4	H_{t-1}, H_{t-2}, $H_{t-3,}$ P_t, Q_t (H_{t-1}, H_{t-2}, H_{t-3}Decomposed in levels 1, 2)
5	H_{t-1}, H_{t-2}, $H_{t-3,}$ T_t, E_t (H_{t-1}, H_{t-2}, H_{t-3}Decomposed in levels 1, 2)

Table 2. Different scenarios for wavelet conjunction models.

4.4. Hybrid Model Performance

The input dataset was decomposed into sub-signals utilizingthe db4 wavelet (mother), and two decomposition levels werecontemplatedto design the conjunctive wavelet models. So, thesub-signals were employedas input datasets to the ANN inGWL prediction. WT triggersa data transition from a time-value to a frequency system in the time domain.The performance of standalone ML models with wavelet ML conjunction models was evaluated by calculating R, RMSE, MAE and NSE statistical parameters. The performanceof standalone ML models is lower than that of the wavelet ML conjunction models because the wavelet ML conjunction models simultaneously take advantage of both standalone ML models and WT. In particular, WT can consider various viewpoints of a time series, such as trends, discontinuities and breakpoints. This time-scale localization feature was employed to create the input signaldecomposition up to the two levels, allowing for a separation of the approximation and providing details of the nonstationary noisy signal.

Selecting the most practicalcombination of input data setsis an importantstage in ML modeling. Containing irrelevant signalswith various time lagsas input in the developed scenariosdeclinesthe prediction qualityof the models. The general correlation analysis shows that climatic components followed by the GWL are the essential predictor of GWL(Fig. 2 and 3). Hence, the overall conclusionsexhibit that scenarios 4 and 5 can be suitable based on the applied dataset (Tables 3 and 4). Scenario number 1 shows less precision and accuracy than the other developed scenarios.The conclusionsshow that differentsignal analysesmayimprovethe generalmodels'results (Yaseen et al. 2019).

V. CONCLUSION

In this work, the simulation of GWLs was done using ML and W-ML models. It might be interesting to assess the potential of these and other methods for long-term GWL simulation. This is significantas efficiently managingseveral environmental issues needs a long-runperspective. In the presentstudy, WT was employed to increase the models' accuracy. The capability of coupled WT and MLmethods was compared to ML models. Giving to the results of the WT combination, the model results weresatisfactoryin estimating GWL changesin the aquifer, such that the variance between the simulated andobserved values was insignificant, resultingin the incorporation of the WT in standalone ML models improved the results of ML models, particularly in the simulationstep. The WT combination typicallyattained better because standalone methodscannotdeal with the seasonality and nonlinearity of data like hybrid-WT models. It could be concluded that all methods in developed scenarios had a reasonable result onemonth ahead of GWL prediction. WT-FL and WT-ANFIS provided GWL simulation one month ahead compared to standalone ML models. The relative superiority of prediction results could be rooted in the more used variables in scenario five than in the other scenarios. Based on the model's outcome, the modelscanaccuratelypredict the GWL one month ahead. However, the accuracy of the models decreased when two and three months were chosen to predict.

Input	structure	Step	Ht+1				Ht+2				Ht+3			
			R	RMSE	MAE	NS	R	RMSE	MAE	NS	R	RMSE	MAE	NS
Comb. 1	3-2-1	Training	0/94	0/76	0/6	0/88	0/88	1/04	0/85	0/77	0/81	1/25	1/03	0/66
		test	0/94	0/66	0/5	0/86	0/9	0/9	0/74	0/7	0/88	0/85	0/72	0/71
Comb. 2	5-3-1	Training	0/95	0/67	0/54	0/9	0/9	0/94	0/75	0/81	0/86	1/09	0/9	0/74
		test	0/94	0/62	0/49	0/87	0/89	0/86	0/71	0/72	0/85	0/84	0/64	0/72
Comb. 3	6-3-1	Training	0/95	0/68	0/55	0/9	0/91	0/9	0/73	0/82	0/86	1/11	0/9	0/73
		test	0/96	0/54	0/44	0/9	0/95	0/72	0/57	0/81	0/88	0/83	0/7	0/72
Comb. 4	7-4-1	Training	0/95	0/66	0/53	0/91	0/91	0/89	0/71	0/83	0/85	1/17	0/96	0/71
		test	0/95	0/57	0/45	0/89	0/93	0/69	0/62	0/82	0/87	0/87	0/74	0/7
Comb. 5	9-5-1	Training	0/96	0/63	0/5	0/91	0/92	0/85	0/66	0/84	0/89	1/01	0/77	0/78
		test	0/95	0/53	0/41	0/91	0/92	0/73	0/58	0/8	0/87	0/79	0/64	0/75

Table 3. The scenarios' outcomesemploy the ANN methodsto predict GWL_{t+1}, GWL_{t+2}, and GWL_{t+3}.

Input	Wavelet type	Step	Ht+1				Ht+2				Ht+3			
			R	RMSE	MAE	NS	R	RMSE	MAE	NS	R	RMSE	MAE	NS
Comb. 1	db4	Training	0/99	0/34	0/28	0/97	0/97	0/53	0/43	0/94	0/96	0/63	0/45	0/91
		test	0/98	0/40	0/29	0/95	0/94	0/54	0/38	0/89	0/88	0/81	0/60	0/74
Comb. 2	db4	Training	0/96	0/60	0/49	0/92	0/93	0/78	0/63	0/87	0/87	1/07	0/87	0/75
		test	0/97	0/45	0/37	0/93	0/94	0/63	0/53	0/85	0/86	0/81	0/61	0/74
Comb. 3	db4	Training	0/97	0/51	0/41	0/95	0/93	0/79	0/65	0/86	0/86	1/14	0/95	0/72
		test	0/98	0/45	0/35	0/93	0/90	0/86	0/67	0/73	0/84	0/94	0/78	0/65
Comb. 4	db4	Training	0/98	0/55	0/43	0/94	0/96	0/61	0/36	0/92	0/94	0/71	0/58	0/89
		test	0/98	0/40	0/34	0/95	0/89	0/82	0/61	0/75	0/89	0/77	0/47	0/76
Comb. 5	db4	Training	0/98	0/39	0/31	0/97	0/96	0/67	0/50	0/90	0/94	0/74	0/55	0/88
		test	0/98	0/44	0/32	0/94	0/91	0/79	0/58	0/77	0/90	0/78	0/56	0/76

Table 4. The scenarios' outcomes employ the W-ANN methods to predict GWL_{t+1}, GWL_{t+2}, and GWL_{t+3}.

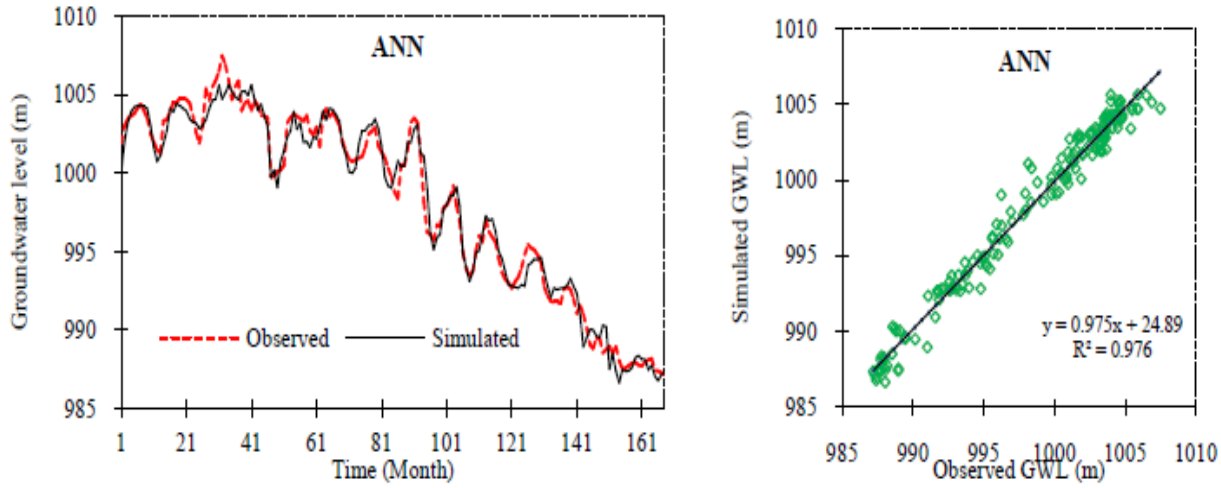


Fig. 2. The predicted and observed GWLemploying the ANN method in the Ht+1 for combination 1.

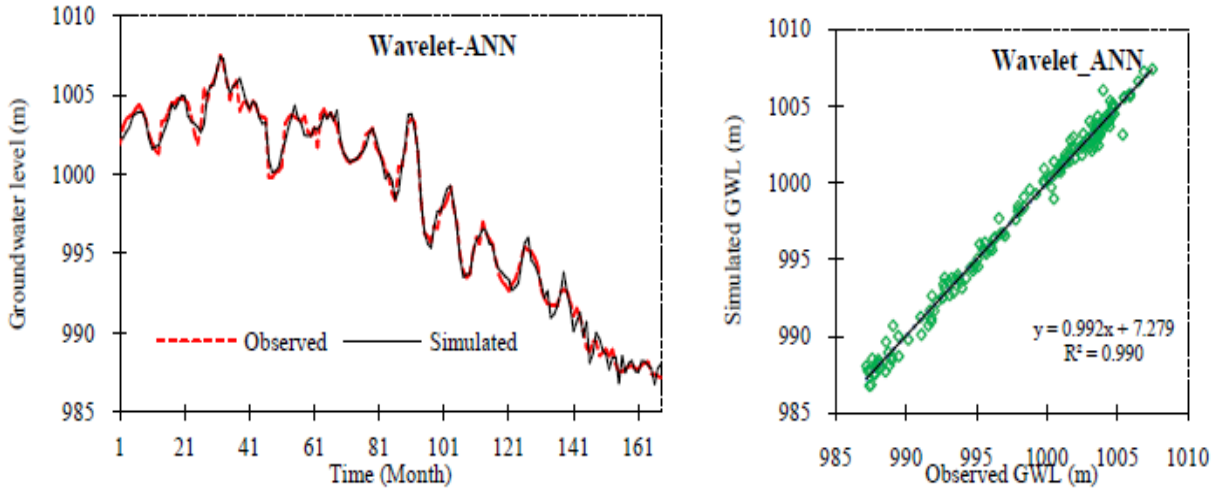


Fig. 3. The predicted and observed GWL employing the W-ANN method in the Ht+1 for combination 1.

REFERENCES

[1] Adamowski, J., &Karapataki, C. (2010). Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. Journal of Hydrologic Engineering, 15(10), 729-743. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000245

[2] Al Atawneh, D., Cartwright, N., & Bertone, E. (2021). Climate change and its impact on the projected values of groundwater recharge: A review. Journal of Hydrology, 601, 126602.

[3] Almuhaylan, M. R., Ghumman, A. R., Al-Salamah, I. S., Ahmad, A., Ghazaw, Y. M., Haider, H., &Shafiquzzaman, M. (2020). Evaluating the Impacts of Pumping on Aquifer Depletion in Arid Regions Using MODFLOW, ANFIS and ANN. Water, 12(8), 2297. https://doi.org/10.3390/w12082297

[4] Azizpour, A., Izadbakhsh, M. A., Shabanlou, S., Yosefvand, F., &Rajabi, A. (2022). Simulation of time-series groundwater parameters using a hybrid metaheuristic neuro-fuzzy model. Environmental Science and Pollution Research, 1-17. https://doi.org/10.1007/s11356-021-17879-4

[5] Chakraborty, S., Maity, P. K., & Das, S. (2020). Investigation, simulation, identification and prediction of groundwater levels in coastal areas of Purba Midnapur, India, using MODFLOW. Environment, Development and Sustainability, 22(4), 3805-3837. https://doi.org/10.1007/s10668-019-00344-1

[6] Gordu, F., &Nachabe, M. H. (2021). A physically constrained wavelet-aided statistical model for multi-decadal groundwater dynamics predictions. Hydrological Processes, 35(8), e14308. https://doi.org/10.1002/hyp.14308

[7] Graf, R., Zhu, S., & Sivakumar, B. (2019). Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. Journal of Hydrology, 578, 124115.

[8] Haykin, S., (2004). Neural Networks: A Comprehensive Foundation. Prentice Hall, New Jersey.

[9] Hssaisoune, M., Bouchaou, L., Sifeddine, A., Bouimetarhan, I., &Chehbouni, A. (2020). Moroccan groundwater resources and evolution with global climate changes. Geosciences, 10(2), 81.

[10] Jyrkama, M. I., & Sykes, J. F. (2016). Impact of climate change on groundwater. In The Handbook of Groundwater Engineering (pp. 709-738). CRC Press.

[11] Karimi, H. S., Natarajan, B., Ramsey, C. L., Henson, J., Tedder, J. L., & Kemper, E. (2019). Comparison of learning-based wastewater flow prediction methodologies for smart sewer management. Journal of Hydrology, 577, 123977. https://doi.org/10.1016/j.jhydrol.2019.123977

[12] Liu, J., Gu, J., Li, H., & Carlson, K. H. (2020). Machine learning and transport simulations for groundwater anomaly detection. Journal of Computational and Applied Mathematics, 380, 112982. https://doi.org/10.1016/j.cam.2020.112982

[13] Moeeni, H., Bonakdari, H., &Ebtehaj, I. (2017). Integrated SARIMA with neuro-fuzzy systems and neural networks for monthly inflow prediction. Water Resources Management, 31(7), 2141-2156. https://doi.org/10.1007/s11269-017-1632-7

[14] Moghaddam, H. K., Kivi, Z. R., Bahreinimotlagh, M., & Alizadeh, M. J. (2019). Developing comparative mathematic models, BN and ANN for forecasting of groundwater levels. Groundwater for Sustainable Development, 9, 100237. https://doi.org/10.1016/j.gsd.2019.100237

[15] Monjezi, M., Mohamadi, H. A., Barati, B., & Khandelwal, M. (2014). Application of soft computing in predicting rock fragmentation to reduce environmental blasting side effects. Arabian Journal of Geosciences, 7(2), 505-511.

[16] Nourani, V., Baghanam, A. H., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. Journal of Hydrology, 514, 358-377. https://doi.org/10.1016/j.jhydrol.2014.03.057

[17] Rezaei, M., Mousavi, S. F., Moridi, A., Gordji, M. E., &Karami, H. (2021). A new hybrid framework based on integration of optimization algorithms and numerical method for estimating monthly groundwater level. Arabian Journal of Geosciences, 14(11), 1-15. https://doi.org/10.1007/s12517-021-07349-z

[18] 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.

[19] Roshni, T., Jha, M. K., &Drisya, J. (2020). Neural network modeling for groundwater-level forecasting in coastal aquifers. Neural Computing and Applications, 1-18. https://doi.org/10.1007/s00521-020-04722-z

[20] Su, Z., Wu, J., He, X., & Elumalai, V. (2020). Temporal changes of groundwater quality within the groundwater depression cone and prediction of confined groundwater salinity using Grey Markov model in Yinchuan area of northwest China. Exposure and Health, 12(3), 447-468. https://doi.org/10.1007/s12403-020-00355-8

[21] Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K. W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. Journal of Hydrology, 569, 387-408. https://doi.org/10.1016/j.jhydrol.2018.11.069

[22] Zhu, S., Ptak, M., Yaseen, Z. M., Dai, J., & Sivakumar, B. (2020). Forecasting surface water temperature in lakes: A comparison of approaches. Journal of Hydrology, 585, 124809. https://doi.org/10.1016/j.jhydrol.2020.124809

★★★