Determination of Groundwater Level using Advanced Machine Learning Methods

Amirhossein Najafabadipour

Department of Mining Engineering, Shahid Bahonar University of Kerman, Kerman, Iran, najafabadipour@eng.uk.ac.ir, ORCID 0000-0003-3411-8122

Abstract

Increasing the depth of mining causes the mine pit to be situated below the water table. This leads to groundwater flowing into the pit, which reduces effectiveness, raises expenses, and compromises safety. Estimating the groundwater level proves to be an important method for overseeing groundwater resources in the mining zone. In this research, 10 temporal features and 12 spatial features serve as inputs for four powerful machine learning models to determine the groundwater level. The developed MARS model accurately determines groundwater level data, achieving an average absolute relative error value of less than 0.16%. Feature selection highlights that electrical resistivity and sediment depth have the highest effects on groundwater level fluctuations. Utilizing a leveraged method to identify outliers suggests that 98% of the data is legitimate. Finally, the analysis of the trend showed that with the intensification of drainage, the groundwater level decreases. The greatest decrease in the groundwater level is related to piezometer number five in the northeastern part of the mining pit (52 m). This increase in drainage is due to the placement of pumping wells in the north and east of the pit.

Keywords: Groundwater level, drainage, machine learning, feature selection, outliers

Introduction

The depletion of surface mineral deposits has caused an expansion of open-pit mining operations into deeper levels. As these operations reach greater depths, often surpassing the groundwater table, excavations below become necessary, causing water to migrate toward the mining sites(Brawner 1986). Consequently, the establishment of an effective drainage system becomes imperative. Accurate predictions regarding groundwater level fluctuations play a pivotal role in designing such a system (Singh and Atkins 1985). Effective drainage management not only minimizes equipment failures but also enhances pit slope stability, reduces reliance on explosives, and fosters improved safety conditions (Najafabadipour et al. 2022b).

Novel machine learning approaches, reliant on nonlinear dependencies, offer promising avenues for forecasting groundwater levels and addressing the complexities of subsurface conditions, even in the absence of an in-depth understanding of fundamental physical parameters (Najafabadipour et al. 2022a). The primary objective of this research is to leverage spatial and temporal data to develop predictive models for fluctuations in groundwater levels using innovative machine learning techniques. To fulfill this objective, four machine learning models were developed from 3534 data points sourced from the Gohar Zamin iron ore mine, encompassing both spatial and temporal features. This research aims to explore the effects of input features on groundwater levels and conduct a detailed analysis of outlier data. Additionally, it seeks to investigate the trends in drainage within the developed models.

Material and Methods

Study Area

The Gol_e_Gohar Iron Ore mine, one of the pivotal districts in the Middle Eastern mining industry, encompasses six separate anomalies and harbors a deposit of about 1.2×10^9 t, spanning an area of 4 km by 10 km (Fig. 1). In Gohar Zamin iron ore mine, groundwater enters the pit, and water seeps through the alluvium on the pit's steps. Water pumping wells encircle the Gohar Zamin iron ore mine, specifically around anomaly no. 1, recognized as a drainage zone. A substantial portion of the Gol_e_Gohar region is blanketed by young erosion and alluvial deposits (Najafabadipour et al. 2023). Rock outcrops predominantly consist of mica-schist, gneiss, and amphibolite metamorphic rocks, within which the mineral magnetite is observable. From a tectonic perspective, the studied area resides in the Sanandaj-Sirjan zone, characterized as a compressional type oriented in a SW-NE direction. The presence of permeable units and fractures within the stone formations comprising the pit wall influences groundwater flow.

Data Collection

In this study, 3534 data points from six piezometers have been used to determine the groundwater level. 10 temporal features (year, month, day, discharge (L/s), drainage (m^3/day), temperature (), wind speed (m/s), relative humidity (%), evaporation (m/day), and rainfall (m/day)) and 12 spatial features (latitude (m), longitude (m), effective porosity of bedrock and sediments, hydraulic conductivity of bedrock and sediments (m/day), depth of sediments (m), the electrical resistivity of bedrock and sediments (ohm.m), bedrock and surface level (m), and fault) have been used to develop four novel machine learning methods. This input features was obtained from the Gohar Zamin iron ore mine.

Modeling

A multilayer perceptron optimized with batch training (MLP-B) is a type of neural network with multiple layers, each neuron receiving inputs from the previous layer without connections among neurons within the same layer (Hebb 2005). Cascade forward optimized with gradient descent (CF-GDA) refers to a neural network structure where information flows strictly from input to output layers, with no connections or feedback loops within layers, enabling sequential processing of data through the network (Fahlman 1990). Multivariate adaptive regression splines (MARS) is a regression technique that fits piecewise linear regression models to data, identifying breakpoints in predictors and creating simple models for different segments to improve prediction accuracy (Friedman 1991). A random subspace ensemble (RS) is a method where subsets of features are randomly selected, and multiple models are trained independently on these subsets to improve prediction accuracy and robustness in machine learning (Tin Kam 1998).

Results and Discussion

Model development

Feature selection holds significance due to its effects on the uncertainty of the model's output, which arises from the variety of existing features. This process serves multiple purposes: assessing the correlation between model inputs and outputs and streamlining the developed model by eliminating inputs that lack influence on its output. Evaluating

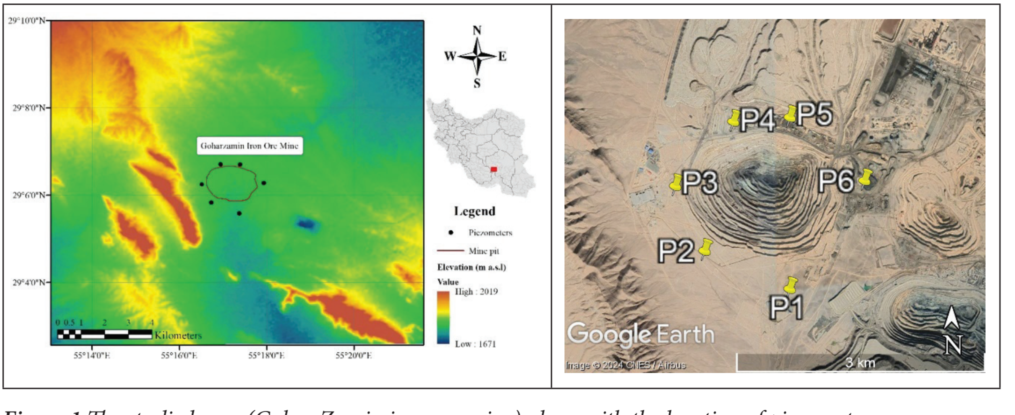

***Figure 1** The studied area (Gohar Zamin iron ore mine) along with the location of piezometers*

the influence of a model's input features on its output can be achieved through a reliable method known as relevancy factor analysis. Due to the potential for excessive model complexity stemming from numerous features, relevancy factor analysis was employed to select those with the most relevant effect on groundwater level (Najafabadipour et al. 2022a). Fig. 2 displays the relevancy factor values for the input features. This signifies that altering the value of each input influences the groundwater level. Among the 22 input features, month, day, discharge, temperature, wind speed, rainfall, and relative humidity exhibit the least effect on groundwater level. The groundwater level in the complex mining area remains largely unaffected by rainfall due to the predominance of fossil water. Consequently, rainfall does not substantially alter the groundwater level, while variations in the electrical resistivity of sediments, influenced by water presence, exert the most relevant influence on groundwater levels. Consequently, 15 other features were selected for machine learning model development. In the pre-processing step, unnecessary data were identified and cleaned, and the input dataset was prepared for further analysis. After the cleaning process, four distinct modeling methods were used: MLP-B, CF-GDA, MARS, and RS. In all these models, 70% of the data were used for training, while the remaining 30% were assigned for testing. Random distributions were applied to partition the data into testing and training sets, preventing localized data accumulation.

Evaluation of the Precision and Validity of the Developed Models

AARE and R^2 statistical parameters are shown in table 1. The ideal predictive system, denoted by a point with an R^2 value of 1 and AARE value of 0, serves as a benchmark (Najafabadipour and Kamali 2019). For the developed MARS smart model in this study, the AARE and R^2 values were calculated at 0.158 and 0.994, respectively. The results of table 1 has been used to rank the developed models based on their accuracy below:

$$\text{MARS} > \text{MLP-B} > \text{CF-GDA} > \text{RS}$$

Table 1 *Statistical parameters of all developed models*

Models	MLP-B	Cf-GDA	MARS	RS
AARE (%)	0.188	0.213	0.158	0.223
R^2	0.9917	0.9894	0.9936	0.9863

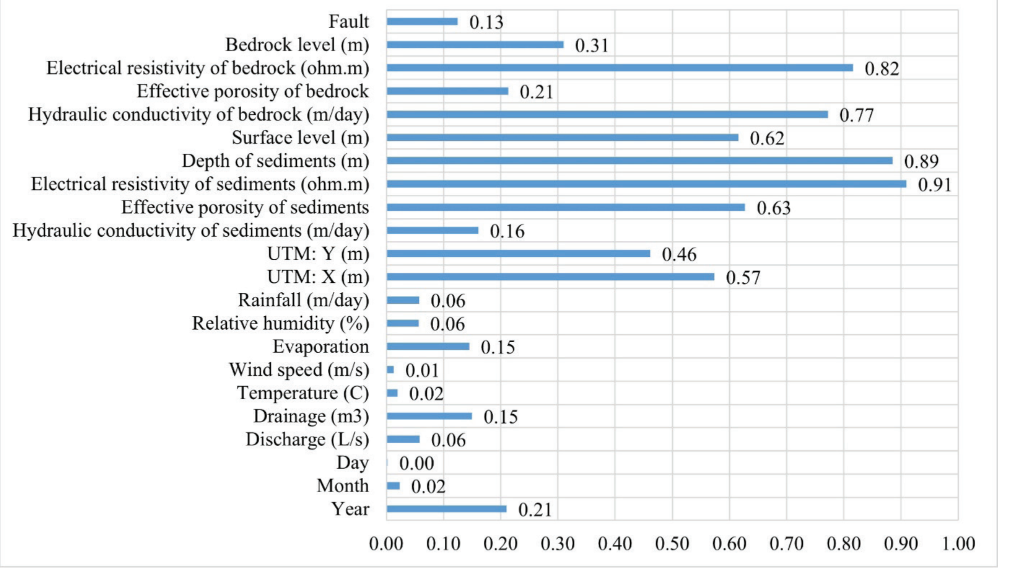

Figure 2 *Feature selection for temporal and spatial input features that influences groundwater level*

For a graphical comparison of models, fig. 3 displays the cumulative frequency of average absolute relative error (AARE) for all developed models. According to fig. 3, approximately 90% of the groundwater level determination made by the MARS model exhibits an AARE of less than 0.4%. These findings from the figure reinforce the success of the developed MARS method in the determination of groundwater levels compared to the other developed methods.

To validate the confirmation of the MARS model's accuracy, fig. 4 showcases a graph plotting real data against the determined groundwater level. The clustered arrangement of data points closely following the unit slope signifies the MARS model's high predictive accuracy.

Applicability Domain of the Model

Outliers frequently occur within a comprehensive set of real data. As such a dataset can affect the precision and reliability of models, identifying and addressing these outliers is essential in model development (Hemmati-Sarapardeh et al. 2020). The leverage approach stands as a potent tool for identifying and eliminating outliers, encompassing the computation of the model's deviation from real data. Fig. 5 illustrates outliers resulting from the MARS model for the groundwater level determination dataset. The majority of determined points fall within the feasible domain of the developed model ($0 \leq$ hat ≤ 0.013 and $-3 \leq R \leq 3$), demonstrating the high reliability and statistical validity of the MARS model. Approximately 98% of these points reside within the acceptable model range, while the remaining outliers, although present, can be disregarded considering the substantial volume of data used for model development. Points falling beyond $R < -3$ or $R > 3$ are categorized as 'Bad High Leverage,' regardless of their hat value with hat* (Hoaglin and Welsch 1978). While these data points might be well determined, their existence outside the model's acceptable range can be attributed to the sheer magnitude of the dataset.

Trend Analysis of the Developed Models

Trend analysis within a hydrogeological time series can serve as a valuable method for examining changes in groundwater levels. Consequently, since the drainage time series substantially effects groundwater levels, the models depicted in fig. 6 investigate the anticipated trends in groundwater levels with

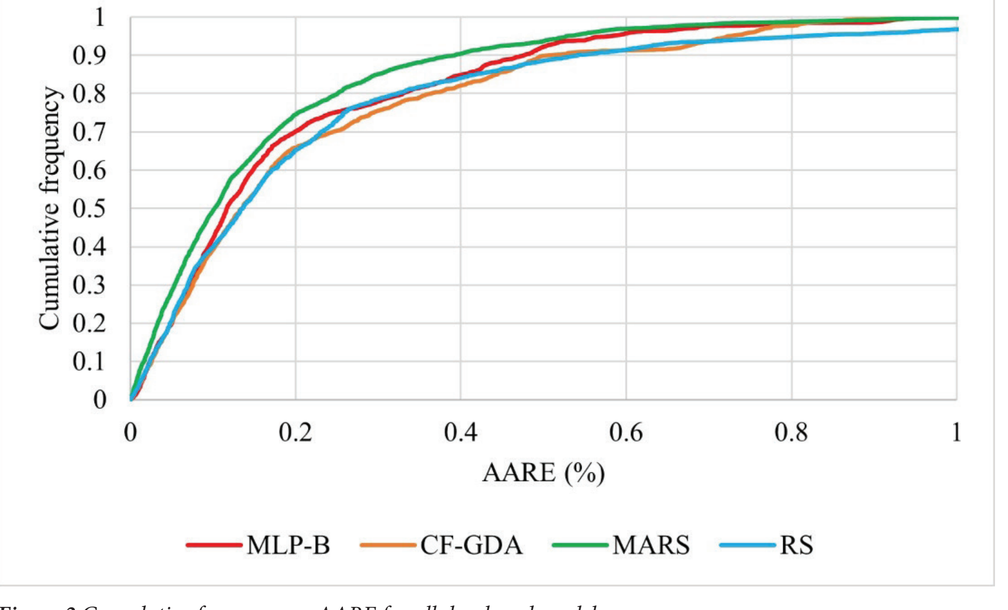

***Figure 3** Cumulative frequency vs AARE for all developed models*

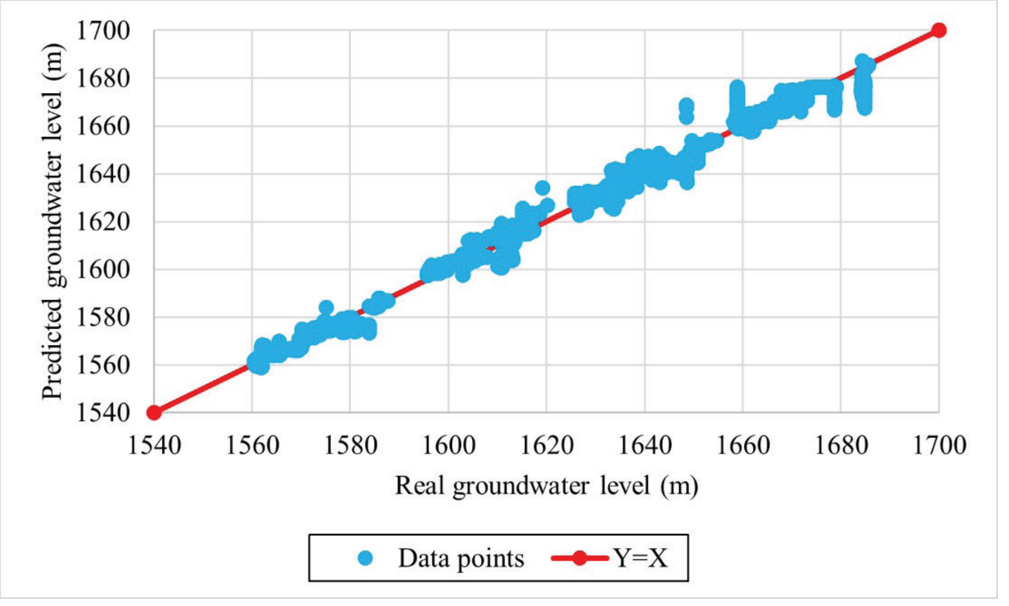

Figure 4 *Cross plots of the MARS model for the groundwater level*

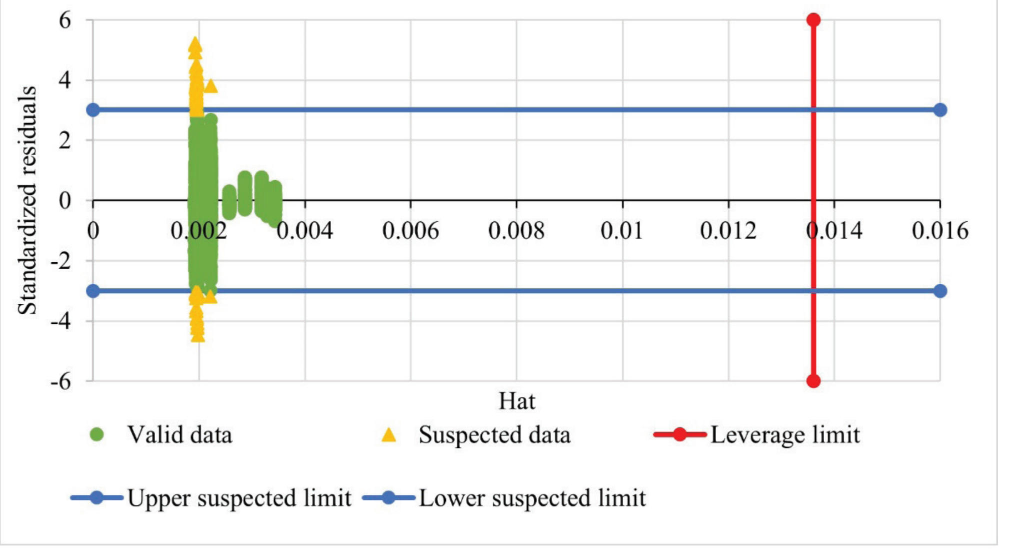

Figure 5 *Outlier data for the resulting outputs by the MARS model*

alterations in drainage. As illustrated, the groundwater level diminishes as drainage intensifies in this figure. The highest decrease in the groundwater level is related to piezometer number five in the northeast part of the mining pit (52 m) and then piezometer 6 in the east of the mining pit (42 m). This increase in the amount of drainage is due to the placement of pumping wells in the north and east of the pit. Each constructed model is capable of capturing the anticipated trend associated with variations in drainage. Additionally, this illustration demonstrates the precision of the suggested MARS model when contrasted with other models.

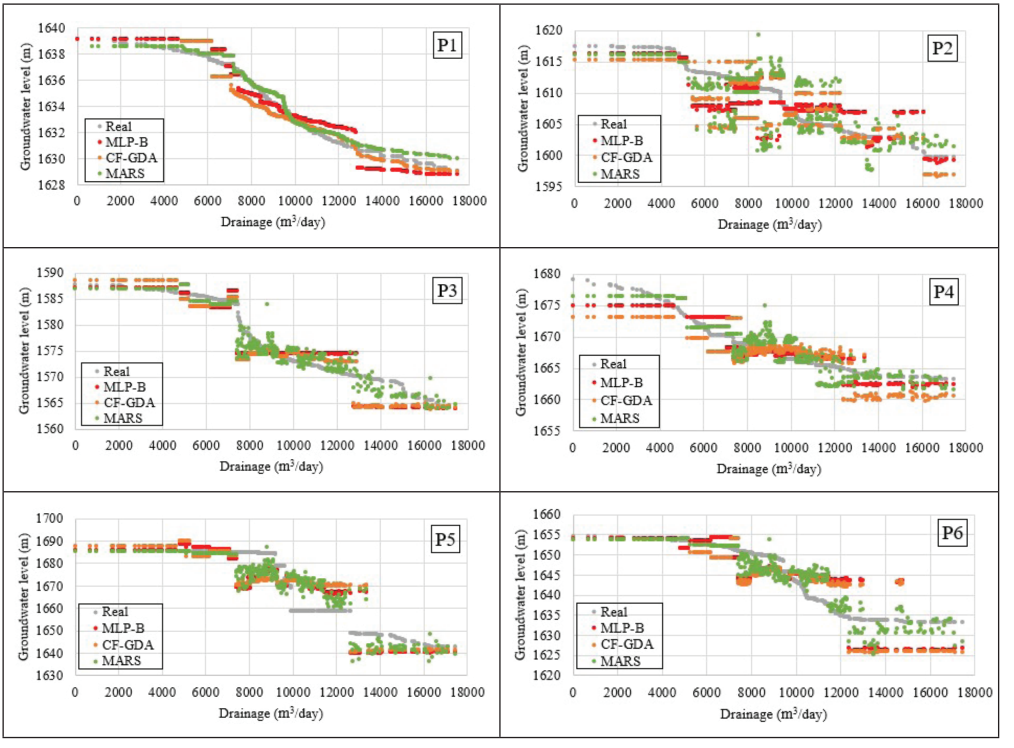

Figure 6 Comparing the groundwater level variation with drainage for the MLP-B, CF-GDA, and MARS models

Conclusion

In this research, four novel machine learning models were used to determine the groundwater level in the Gohar Zamin iron ore mine. The results of statistical and graphical error analysis showed the high accuracy of the MARS model compared to other data development models. Expanding the quantity of input data is recommended to enhance the precision of machine learning techniques. Future research could investigate the combination of new models to increase forecast accuracy. The results of this study can be used in different fields to increase the accuracy of any determination model. The capability of the MARS model to determine new data can be relevant in areas where data scarcity is a common challenge.

References

Brawner C (1986) Groundwater and coal mining. Mining Science and Technology 3(3):187-198

Fahlman SE (1990) The cascade-correlation learning algorithm. Neural Information Proccesing Systems:524-532

Friedman JH (1991) Multivariate Adaptive Regression Splines. The Annals of Statistics 19(1):1-67, 67

Hebb DO (2005) The organization of behavior: A neuropsychological theory. Psychology press

Hemmati-Sarapardeh A, Larestani A, Nait Amar M, Hajirezaie S (2020) Chapter 1-Introduction. Applications of Artificial Intelligence Techniques in the Petroleum Industry:1-22

Hoaglin DC, Welsch RE (1978) The Hat Matrix in Regression and ANOVA. The American Statistician 32(1):17-22. https://doi.org/10.1080/00031305.1978.10479237

Najafabadipour A, Kamali G (2019) Exploring and Modelling the Qanat Gallery Path Using Electrical Resistivity Tomography Data. 2019(1):1-5. https://doi.org/https://doi.org/10.3997/2214-4609.201902532

Najafabadipour A, Kamali G, Nezamabadi-pour H (2022a) Application of Artificial Intelligence Techniques for the Determination of Groundwater Level Using Spatio–Temporal Parameters. ACS Omega 7(12):10751-10764. https://doi.org/10.1021/acsomega.2c00536

Najafabadipour A, Kamali G, Nezamabadi-pour H (2022b) The Innovative Combination of Time Series Analysis Methods for the Forecasting of Groundwater Fluctuations. Water Resources 49(2):283-291. https://doi.org/10.1134/S0097807822020026

Najafabadipour A, Kamali G, Nezamabadi-pour H (2023) Numerical modeling, groundwater management, and evaluation of optimal water pumping rate in Gohar Zamin Iron Ore Mine (Sirjan-Iran). Hydrogeology 8(1):17-33. https://doi.org/10.22034/hydro.2023.49719.1256

Singh R, Atkins A (1985) Analytical techniques for the estimation of mine water inflow. International Journal of Mining Engineering 3:65-77

Tin Kam H (1998) The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8):832-844. https://doi.org/10.1109/34.709601