Prediction of Groundwater Inflow Rate Using Non-Linear Multiple Regression and ANFIS Models: A Case Study of Amirkabir Tunnel in Iran

Mohammad Reza MOGHADDASI[a]*, Abbas Aghajani BAZZAZI[a], Ali AALIANVARI[a]

[a]Department of Mining Engineering, University of Kashan, Kashan, Iran

*Corresponding Author: mr.moghaddasy@gmail.com

ABSTRACT: The prediction of groundwater inflow rate (GIR) into a tunnel is one of the serious challenges during the design, construction and exploitation of tunnels. GIR could lead to undesirable effects on excavation process such as decrease in rock mass stability, make extra pressure on permanent and temporary stability system, destructive effects on geomechanical condition of rock and finally physical and economical dangers. In this paper, an adaptive neuro fuzzy inference system (ANFIS) has been presented for anticipating the GIR into AmirKabir tunnel, Iran. For this purpose, a total number of 110 datasets including most influential parameters on GIR were inquired and used to construct the GIR predictive model. To illustrate superiority of ANFIS model, a non-linear multiple regression (NLMR) model was also developed for anticipating of GIR. In order to assess the performance of the developed models, coefficient of determination (R^2), root mean square error (RMSE) and variance account for (VAF) were calculated. The results of this research indicate the higher reliability of ANFIS compared to NLMR model for GIR prediction.

KEYWORDS: Groundwater inflow rate, ANFIS, Non-linear multiple regression, Amir Kabir Tunnel

1. INTRODUCTION

Due to the absence of simple, accurate equations or models that can be readily applied to hard-rock tunnels and extend range of rock mass permeability in fractured rocks that normally repeats again and again over the lengths of long tunnels, estimation of groundwater inflow rate into tunnel is very difficult. The inflow rate estimation is required to size the pumping system, and treatment the plant facilities for construction planning and cost assessment. An estimate of the excavation-induced drawdown of the initial groundwater level is required to evaluate the potential of environmental impacts. The groundwater inflow during excavation can cause serious instability of tunnel roof and walls as well as ground settlement due to ground losses or consolidation of soft overburden deposits. It can also cause flooding inside the tunnel, construction difficult, and ultimately abandonment of tunnel. Due to impossibility of recognition and exact determination of all effective factors on groundwater flow into tunnels, especially during drilling operation in rock medium, exact prediction of groundwater flow into drilled tunnels is difficult. So, analytical methods and equations, because of their simplifications and practical theories, have many applications in calculation of groundwater infiltration into the tunnels. The most important researches about calculation of the rate of groundwater flow into tunnels are studies of Goodman et al (1965), Freeze and Cherry (1979), Heuer (1995), Lei (1999), Karlsrud (2001), El-Tani (2003) and Aalianvari (2014). Analytical methods considering the parameters such as rock mass permeability, water table height above tunnel axis, tunnel radius, etc, and estimate the groundwater inflow rate into tunnels. In this paper based on the rock mass parameters of AmirKabir tunnel such as GSI, Permeability, Average Shear S.P.C, Overburden, Lugeon, water head above tunnel and Average Shear S.P.F by the Non-linear multiple regression and ANFIS method, the rate of groundwater inflow into tunnel has been predicted.

2. THEORY AND METHODS OF MODELING

2.1 Non-linear multiple regression

Regression analysis is a statistical process for estimating the relationships among variables. It frequently makes useful forecast (Allen & Fildes, 2001). Regression-based prediction is most effective when dealing with the small numbers of variables, and large amounts of the reliable and valid data (Armstrong, 2012). It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors') it is called simple regression and multiple regression, respectively. Multiple regression technique can be used to obtain the best-fit equation when there is more than one input variable. In general, the objective of such techniques is to estimate a relationship between input and output parameters. (Shirani Faradonbeh et al. 2015). Nonlinear multiple regression (NLMR) is a method of finding a nonlinear model of the relationship between the dependent variable (output parameters) and a set of independent variables (input parameters). Unlike the traditional multiple linear regression, which is restricted to estimating linear models, NLMR can estimate models with arbitrary relationships between independent and dependent variables.

2.2 Adaptive Neuro Fuzzy Inference System

An adaptive neuro fuzzy inference system (ANFIS) is a combination of adaptive network and fuzzy logic. This technique was developed in the early 1990s (Jang and Shing, 1991). ANFIS is mixing both adaptive network and fuzzy logic principles, since it has potential to capture the benefits of both in a single framework. This system, using Takagi and sugeno's fuzzy if-then rules and have learning capability to approximate nonlinear functions (Abraham, 2005). Because of that, ANFIS is considered to be a universal estimator. Basically, a fuzzy inference system is composed of five function blocks (Figure 1):

(i) A rule base containing a number of fuzzy if-then rules.
(ii) A database which defines the membership function of the fuzzy sets used in the fuzzy rules.
(iii) A decision-making unit which perform the inference operation on the rules.
(iv) A fuzzification inference which transforms the crisp inputs into degree of match with linguistic values.
(v) A defuzzification inference which transforms the fuzzy results of the inference into a crisp output.

The typical ANFIS architecture with two inputs x and y, and one output is shown in Figure. 2.

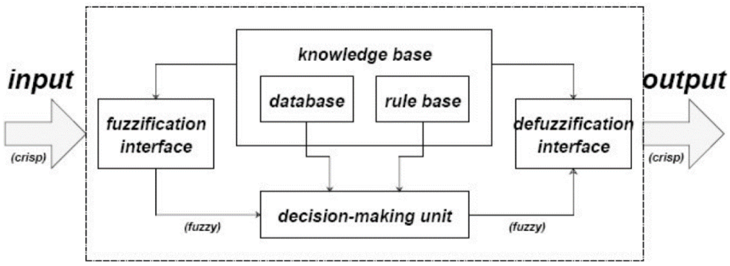

Figure 1. Fuzzy inference system (Jang and Shing, 1991)

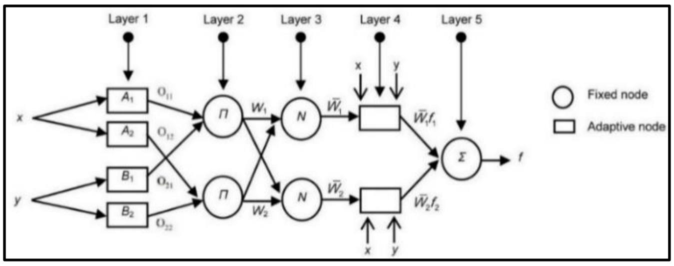

Figure 2. A typical ANFIS architecture

The entire ANFIS architecture consist of five layers that each layer output is input for the next layer. Every node in layer 1 is an adaptive node with a node function that may be a Gaussian membership function or any membership function. Every node in layer 2 is a fixed node, representing the firing strength of each rule. Every node in layer 3 is a fixed node, representing the normalized firing strength of each rule. Every node in layer 4 is an adaptive node with a function. The single node in layer 5 is a fixed node, indicating the overall output as the summation of all incoming signals. A detailed discussion on ANFIS is described by Jang (Jang, 1993).

3. Case Study

The case study was carried out in Amirkabir tunnel, located in the north west of Tehran, Iran. That is designed and being operated to transfer water from Amirkabir dam to Tehran. One of the difficulties in this project is groundwater inflow into the tunnel while doing the excavation operations (Figure. 3).

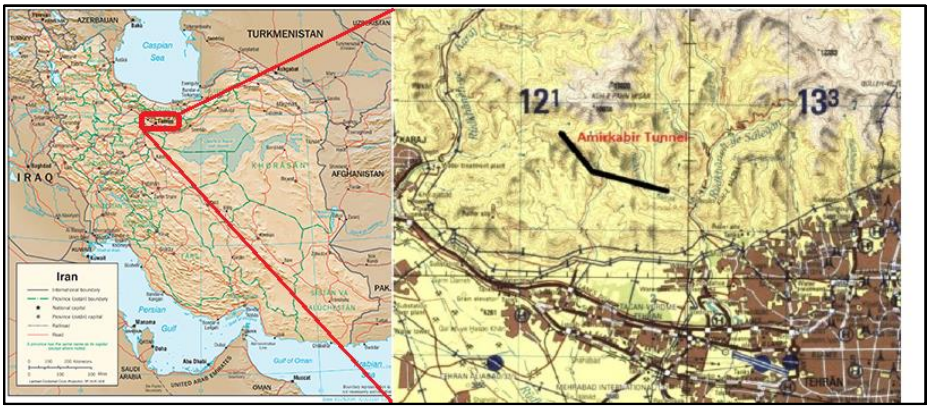

Figure 3. Location of Amirkabir Tunnel

In geological studies, the tunnel intercept 14 geological units, which generally encompass various sedimentary-volcanic sets of Karaj formation. Its petrology contains layers of tuff, sandstone, fine-grained conglomerate, siltstones, lava and agglomerate. In this study, the groundwater inflow to tunnel from kilometers 3.1 to 14.1 has been investigated. Tunnel is divided to 9 engineering geological sections: Gta2 (sandstone and tuff layers), Gta3 (sandstone layers, tuff, and micro conglomerate), Gta4-1 (sandstone, tuff), Gta4-2 (tuff, in sandstone sections and micro conglomerate), Sts1 (tuff, siltstone, layers of sandstone and micro conglomerate), Sts2-1 (tuff, limestone), Sts2-2 (tuff, limestone, shale and siltstone), Tsh-1 (Sandstone, Shale, Silt stone) and Cz (tuff, sandstone, and micro conglomerate). Geological profile along the tunnel is shown in Figure 4.

3.1 Data Collection and Preprocessing

In this research, a database including 110 datasets was gathered from tunneling operation of Amirkabir tunnel. For modeling the GIR, seven parameters were considered as the input parameters. Descriptive statistics distribution of the input and output parameters and their respective symbols are illustrated in Table 1. At the initial stage of this modeling procedure, to simplify the design procedure, the prepared database was normalized to within the range 0-1 by following equation:

$$X_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Where x and X_{norm} are the measured data and normalized data, respectively. X_{min} and X_{max} are the minimum and maximum values of the x. The ratio of training to test datasets used in this study for all the predictive models is 80 to 20 percent. This means that 88 datasets randomly selected and employed to train the predictive models and 22 datasets were randomly selected and employed to test the predictability of the GIR and also check the performance of the developed predictive models.

Table 1. The descriptive statistics of input and output parameters

Parameters	Description	Unit	Symbol	Minimum	Maximum
Inputs	GSI	-	GSI	33.5	77.5
	Permeability	m/s	PR	$5*10^{-8}$	$2.35*10^{-6}$
	Average Shear S.P.C	Mpa	S.P.C	0.484	6.518
	Overburden	m	OB	65	660
	Lugeon	-	LU	0.5	23.45
	Water head	m	WH	55	535
	Average Shear S.P.F	Mpa	S.P.F	24.7	57.6
Output	Groundwater inflow rate	lit/sec/m	GIR	0.009	0.024

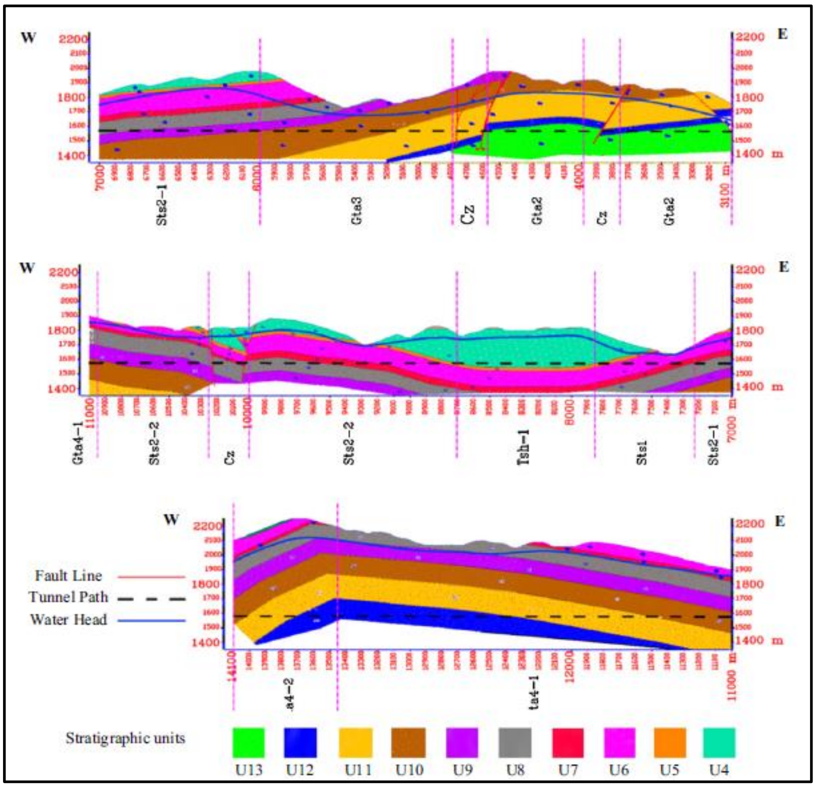

Figure 4. Geological profile along Amirkabir tunnel

4. RESULTS AND DISCUSSION

4.1 Non-Linear Multiple Regression Model

Non-linear multiple regression analysis was carried out on groundwater inflow rate (GIR) as output and gsi (GSI), permeability (PR), average shear S.P.C (S.P.C), overburden (OB), Lugeon (LU), water head (WH) and average shear S.P.F (S.P.F) as input parameters.

NLMR model to predict GIR is given below:

$$GIR = 0.177 - 0.928\,GSI + 2.1\,GSI^2 - 1.026\,GSI^3 + 0.487\,SPC - 1.247\,SPC^2 + 0.715\,SPC^3 + 1.219\,OB - 2.207\,OB^2 + 1.052\,OB^3 - 0.048\,LU + 0.102\,LU^2 - 0.062\,LU^3 - 0.28\,WH + 0.622\,WH^2 - 0.336\,WH^3 + 4.447\,SPF - 10.804\,SPF^2 + 6.203\,SPF^3 \quad (2)$$

The NLMR modeling was carried out by IBM SPSS Ver.20 Statistics software and the software excluded the permeability in modeling of GIR, because this parameter has negligible effect on the GIR.

In this paper, the accuracy level of all developed models is assessed by evaluating three of the most well-known performance indices called correlation coefficient (R^2), root mean square (RMSE) and variance account for (VAF).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_{imeas} - x_{ipred})^2}{\sum_{i=1}^{n}(x_{imeas} - \bar{x})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n}[(x_{imeas} - x_{ipred})^2]} \quad (4)$$

$$VAF = \left[1 - \frac{\mathrm{var}(x_{imeas} - x_{ipred})}{\mathrm{var}(x_{imeas})}\right] \times 100 \quad (5)$$

Where x_{imeas}, $\bar{x}$, x_{ipred} and n are i^{th} actual value, mean value of the x, i^{th} predicted value by the models and number of datasets, respectively. It is well known that higher value of R^2 and VAF and also lower value of RMSE, indicates the superiority of the predictive model. If R^2 is one, RMSE is zero and VAF is 100 (%), the model will be perfect. The R^2, RMSE and VAF for NLMR model are given in Table 2. Figure 5 illustrate the relationship between actual and predicted values obtained by the Equation. (2) In training and testing stages.

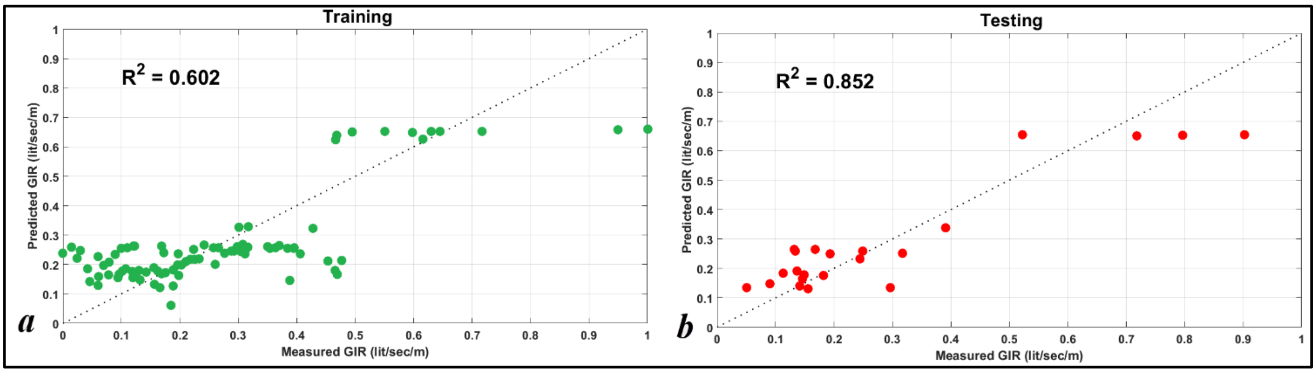

Figure 5. Correlation coefficient for NLMR model: *a*: Training *b*: Testing

4.2 ANFIS Model

In this research, the Gaussian membership function is employed for the input variables and ANFIS learning algorithm proposed by Jang, called hybrid learning algorithm that is a combination of the least-squares estimate (LSE) method and the back-propagation gradient descent method, is used to fine-tuning the parameters of Sugeno-type fuzzy inference system (Esmaeili et al., 2014). The Hybrid learning algorithm is one of the advantage of ANFIS that can estimate the premise and consequent parameters (Jang, 1993). Hybrid learning algorithm have the ability that can decrease the complexity of algorithm and increasing the learning efficiency simultaneously (Singh et al., 2005). To generate fuzzy inference system (FIS) structure, fuzzy c-means clustering method (FCM) was used. FCM clustering algorithm was developed by Duda and Hart (Duda and Hart, 1973), and improved by Bezdek (Bezdek, 1981). Clustering is a procedure that divides the data into groups named clusters, or homogeneous classes that parameters in the same cluster are as similar as possible and parameters in different clusters are as dissimilar as possible. Unlike non-fuzzy clustering methods, that partitions the data into crisp clusters, the fuzzy clustering methods and FCM method allows each data point to belong to multiple clusters with varying degrees of membership. Several models with different structure and seven inputs and one output were constructed and trained. For the best model determination, RMSE was calculated for this models.

Figure 6 reveals the proposed ANFIS model structure for GIR estimation which has eight membership function for each inputs and eight rules. Figure 7 shows the relationship between actual and predicted values by ANFIS model in training and testing stages. The obtained values of R^2, RMSE and VAF, given in Table 2, reveals high prediction performance.

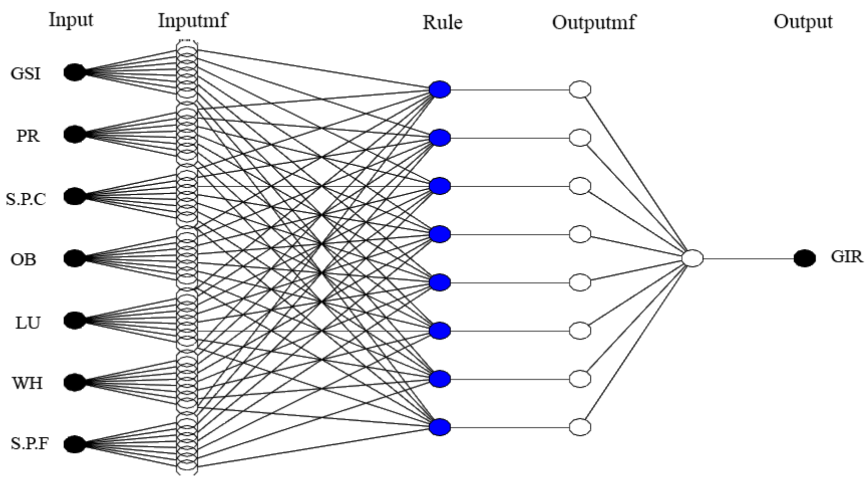

Figure 6. ANFIS model structure for prediction of GIR

4.3 Comparison Between the Models

As mentioned in section 4.1, a comparison based on performance prediction is calculated between the NLMR and ANFIS models (Table 2). It is notable that the ANFIS model can predict GIR with high accuracy level compared to NLMR model. The concordance between the actual and predicted values of GIR obtained by NLMR and ANFIS models for the training and testing data is shown in Figures 8 and 9, respectively.

Table 2. Performance indices for predictive models

Model	Training			Testing		
	R^2	VAF (%)	RMSE	R^2	VAF (%)	RMSE
NLMR	0.602	60.02	0.122	0.852	85.23	0.096
ANFIS	0.984	98.46	0.024	0.971	97.17	0.041

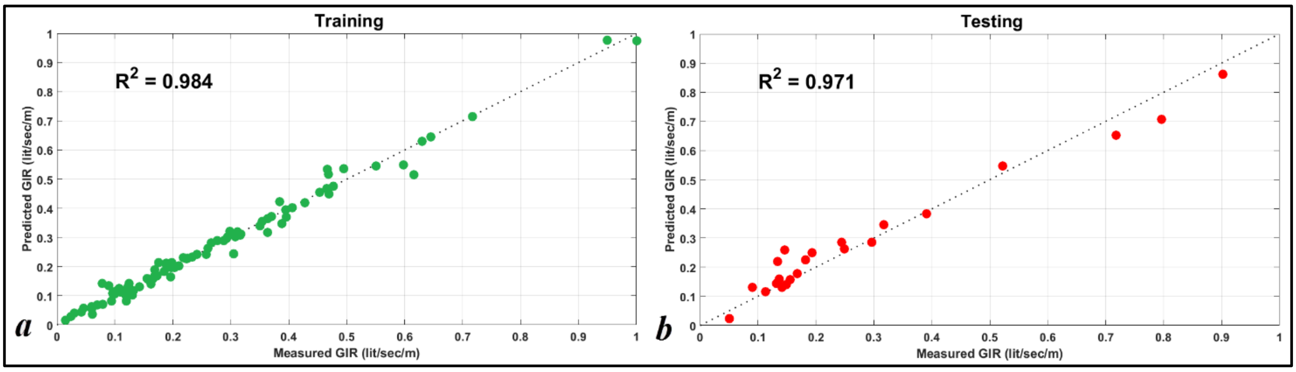

Figure 7. Correlation coefficient for ANFIS model: *a*: Training *b*: Testing

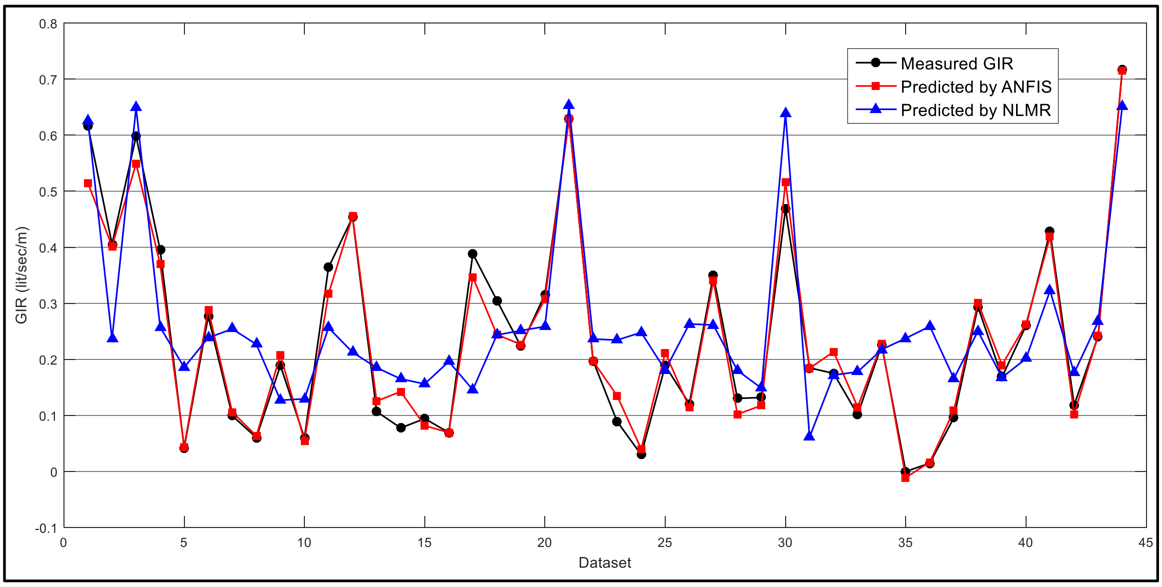

Figure 8. Comparison of measured and predicted GIR by different models for 44 datasets randomly selected of training stage

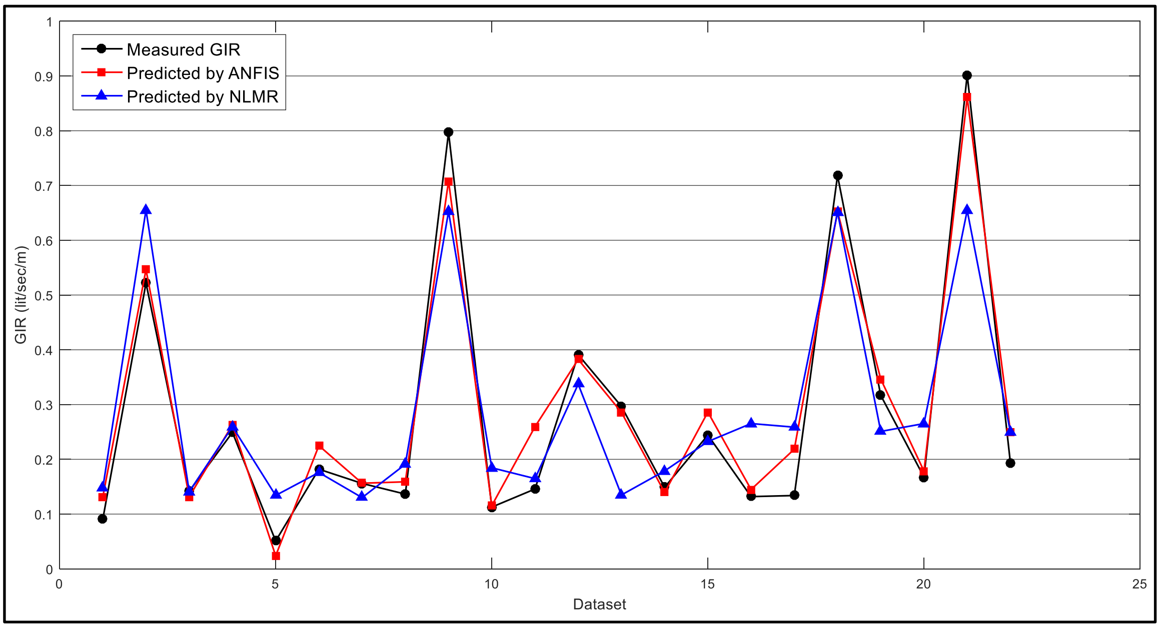

Figure 9. Comparison of measured and predicted GIR by different models for all testing stage datasets

5. CONCLUSIONS

In this paper, non-linear multiple regression and ANFIS models were applied to predict the groundwater inflow rate (GIR) caused by tunneling operation. For GIR modelling, 110 datasets including seven most effective parameters were collected from Amirkabir tunnel of Iran. Results of comparison between the NLMR and ANFIS models based on three performance indices; R^2, RMSE and VAF indicates that ANFIS model with R^2=0.984, RMSE=0.024 and VAF (%) =98.46 for training stage and R^2=0.971, RMSE=0.041 and VAF (%) =97.17 for testing stage, can predict the GIR with high level of accuracy than the NLMR model. It is important to note that developed models in this paper are specific to Amirkabir tunnel region and application of these models in other regions needs some modifications based on their conditions.

REFERENCES

Aalianvari A., 2014. Optimum depth of grout curtain around pumped storage power cavern based on geological conditions. Bulletin of Engineering Geology and the Environment, vol. 73, 775-780.

Abraham A., 2005. Adaptation of fuzzy inference system using neural learning. In Fuzzy systems engineering, Springer Berlin Heidelberg, 53-83 p.

Allen PG., & Fildes R., 2001. Econometric forecasting. In Principles of forecasting, Springer US, 303-362 p.

Armstrong JS., 2012. Illusions in Regression Analysis. International Journal of Forecasting, vol. 28, 689–694.

Bezdek JC., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business Media, 256p.

Duda R., & Hart P., 1973. Pattern classification and science analysis, Wiley-Interscience, 482. P.

El Tani M., 2003. Circular tunnel in a semi-infinite aquifer. Tunnelling and underground space technology, vol.18, 49-55.

Esmaeili M., Osanloo M., Rashidinejad F., Aghajani Bazzazi A., & Taji M., 2014. Multiple regression, ANN and ANFIS models for prediction of backbreak in the open pit blasting. Engineering with computers, vol. 30, 549-558.

Freeze RA., & Cherry JA., 1979. Groundwater, Printice-Hall Inc, 604p.

Goodman RE., Moye DG., Van Schalkwyk A., & Javandel I., 1964. Ground water inflows during tunnel driving. College of Engineering, University of California.

Heuer RE., 1995. Estimating rock tunnel water inflow. Proceedings of the Rapid Excavation and Tunneling Conference, 41-60.

Jang J., & Shing R., 1991. Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm. Proceedings of the 9th National Conference on Artificial Intelligence, USA, 762–767.

Jang JS., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics, vol. 23, 665-685.

Karlsrud K., 2001. Water control when tunneling under urban areas in the Olso region. NFF publication, vol. 12, 27–33.

Lei S., 1999. An Analytical Solution for Steady Flow into a Tunnel. Ground water, vol.37, 23-26.

Shirani-Faradonbeh R., Monjezi M., & Armaghani DJ., 2015. Genetic programing and non-linear multiple regression techniques to predict back break in blasting operation, Engineering with Computers, vol. 32, 123-133.

Singh TN., Kanchan R., Verma AK., & Saigal K., 2005. A comparative study of ANN and neuro-fuzzy for the prediction of dynamic constant of rock mass. Journal of Earth System Science, vol.114, 75-86.