How to cite: Demirci, M., Taşar, B., Kaya, Y.Z., Gemici, E. (2021) Monthly Groundwater Level Modeling Using Data Mining Approaches. *2021 "Air and Water – Components of the Environment" Conference Proceedings*, Cluj-Napoca, Romania, p. 75-86, DOI: 10.24193/AWC2021_07.

MONTHLY GROUNDWATER LEVEL MODELING USING DATA MINING APPROACHES

***Mustafa DEMIRCI*[1], *Bestami TAŞAR* [1]*, *Yunus Ziya KAYA* [2], *Ercan GEMİCİ* [3]**

DOI: 10.24193/AWC2021_07

ABSTRACT. Determination of the fluctuations in groundwater level (GWL) in terms of planning and operating their resources is important**.** In Turkey, many basins are experiencing problems in terms of the potential of groundwater. Increasing water demand, adverse conditions created by climate change and lack of planning related to underground water management in the basin have increased these problems. As a field of application, it was applied for General Directorate of State Hydraulic Works (DSI) well of Hatay province in Turkey. In the study, GWL predictions were evaluated using data mining approaches such as Radial Basis Neural Network (RBNN) and Support Vector Machines (SVM) methods. Monthly data sets between 2002 and 2015, including hydrological parameters predict the GWL used. According to comparison results, it was observed that the data mining models gave good results for observation in test phase.

Keywords: Ground water level, Prediction, Neural Network, Support Vector Machines, Data mining.

1. INTRODUCTION

Groundwater is one of the important fresh water resources in the world. It is frequently used in agricultural irrigation, industry and other basic human needs. However, resources are threatened by intense population growth, global warming, excessive and unconscious use of resources. The existence of these threats made the planned use of resources compulsory. In order to prepare an effective resource utilization plan, it is necessary to carefully monitor the changes in the groundwater level and calculate the future conditions of the resources accurately.

Although it is possible to calculate the groundwater level by physically based methods, they have practical limits. The existence of many parameters affecting the groundwater level and their complex effects, as well as the cost and difficulty of large-scale and accurate measurements of these parameters, increased the importance of artificial intelligence methods as an alternative to physical-based

[1] Iskenderun Techn. University, Iskenderun/Hatay – TURKEY, e-mail: mustafa.demirci@iste.edu.tr , bestami.tasar@iste.edu.tr *,
[2] Osmaniye Korkut Ata University, Osmaniye – TURKEY, e-mail: yzkoku@outlook.com
[3] Bartin University, Bartin – TURKEY, e-mail: egemici@bartin.edu.tr

methods. The precipitation-evaporation relationship, the interaction between groundwater and surface waters and the quantity, storage and nutritional potentials of the modeling studies should be determined accurately. In the estimation of these parameters, the determination or prediction of the groundwater in the region is important in determining the other parameters of the hydrological cycle.

Data mining methods collect information about the samples, make generalizations and then make decisions about the samples by using the information they have learned compared to the samples they have never seen before. Recently, artificial intelligence methods have begun to be frequently used in modeling the suspended sediment (Taşar et al. (2017)), dam reservoir level (Demirci et al. 2018; Üneş et al., 2019a), density flow plunging (Üneş et al., 2015), dam reservoir volume (Üneş et al. 2013; Üneş et al., 2019b), coastal sediment Demirci and Akoz (2013), evapotranspiration (Kaya and Tasar 2019; Üneş et al. 2018) and rainfall-runoff (Taşar et al. (2019).

Many investigations have been also conducted to predict groundwater level fluctuations with using data driven techniques. Demirci et al (2017), estimated groundwater level using Multi Linear Regression (MLR) and Artificial Neural Networks (ANN) models in Amik Plain, Turkey. Heesung et al.(2011), studied two nonlinear time-series models for predicting the groundwater level (GWL) fluctuations using ANNs and SVM. Kaya et al. (2018) investigated the groundwater level (GWL) using artificial neural networks (ANN), M5tree (M5T) approaches in Turkey. Demirci et al. (2019) studied the groundwater level of Reyhanli region using multi-linear regression (MLR), adaptive neural fuzzy inference system (ANFIS) methods.

Emamgholizadeh et al. (2014), studied the potential of artificial neural network (ANN) and adaptive neural fuzzy inference (ANFIS) for the groundwater levels (GWL) predictions. Nourani et al. (2015), used feed-forward neural network (FFNN), Auto Regressive Moving Average (ARIMAX) models for GWL forecasting of the plain of Ardabil, Northwestern Iran. Zare and Koch (2018) used new hybrid Wavelet-ANFIS model with several combinations of inputs and mother wavelets to simulate and predict GWL-fluctuations in the Miandarband plain, Iran. The results showed that all model approaches could be used with acceptable accuracy.

The aim of this study is to investigate the monthly ground water level (GWL) fluctuation estimation based on Radial Basis Neural Network (RBNN) and Support Vector Machines (SVM) models performance.

2. DATA AND METHODS

In this paper, Radial Basis Neural Network (RBNN) and Support Vector Machines (SVM) methods were used to obtain GWL estimation. Data belongs to GWL well in Hatay region. 136 monthly (101 for train and 35 for test data) parameters such as relative humidity (RH), precipitation (P), groundwater level data were used for GWL estimation. Monthly GWL fluctuations which used in the study are given in Fig 1.

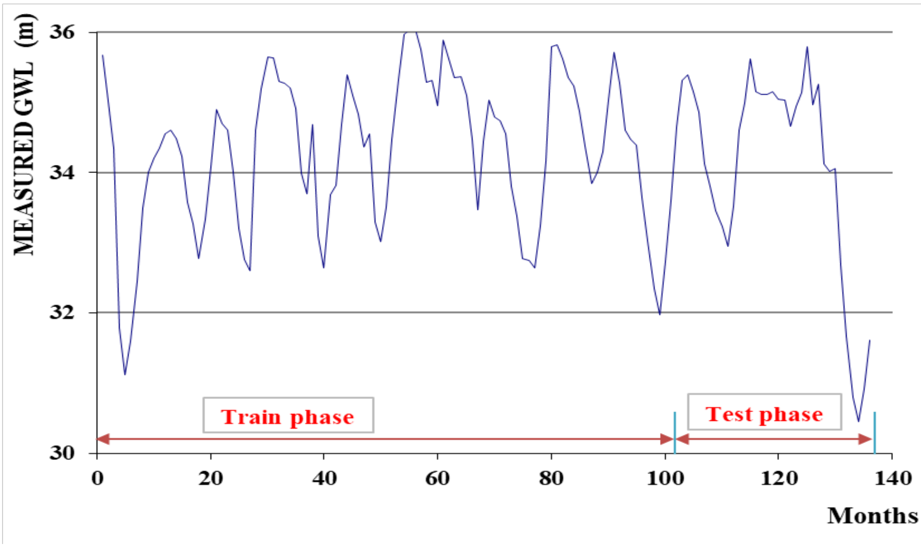

Fig. 1. ***Monthly GWL (m) fluctuations between 2002-2015 years***

2.1. Radial Basis Neural Network Method

Radial basis neural network was first presented into the ANN literature by Broomhead and Lowe in (1988). The RBNN models consist of two layers whose output nodes form a linear combination of the basis functions. The learning scheme of RBNN is fundamentally different from that of the feed-forward ANN. The RBNN assumes a radially symmetric function, typically the Gaussian function, for its hidden units.

Detailed information about RBNN theory can be found in the works of Haykin (1998)

2.2. Support Vector Machines Method

The SVM has become a relatively novel and promising estimator in data-driven research fields, of which basic concept and theory have been introduced by Vapnik (1998).

The generalization ability of the SVM is considered to be better than ANN, in the sense that it is based on the structural risk minimization rather than the empirical risk minimization of ANN.

The main process of SVM model building consists of selecting support vectors which support the model structure and determining their weights. Fig. 2 shows the SVM2 models schematic representations in this paper.

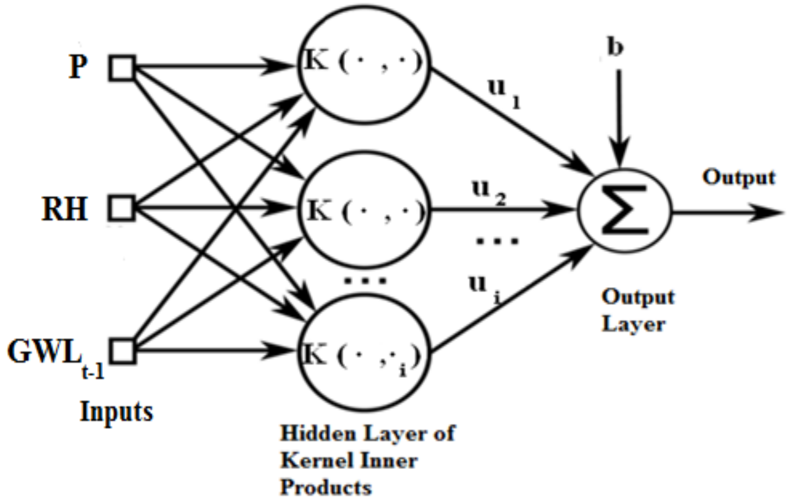

Fig. 2. ***Schematic structure of SVM2 model in this study***

SVM provides to define how to draw this boundary between variables group. In SVM, the Kernel method greatly increases machine learning in nonlinear data. The process of an SVM estimator (y) can be expressed as:

$$y = (K_{xi} \cdot W_{jk}) + b \tag{1}$$

where the Kernel function is K_{xi}, b is bias term of SVM network and W_{jk} is called as the weight vector. K_x and W show Lagrange multipliers. K_{xi} is a nonlinear function that maps the input vectors into a high-dimensional feature space. The inner product of the inputs is calculated by using kernel functions. Lagrange multipliers show the weights the non-linear Poly Kernel functions used in this study. Details about SVM can be found in Vapnik (1999) and Haykin (1999).

3. RESULTS AND DISCUSSIONS

In this study, the results of Radial Basis Neural Network (RBNN1, RBNN2) and Support Vector Machines (SVM1, SVM2) were compared according to the following statistical criteria. In the RBNN1 and SVM1 models, monthly groundwater level time series (GWL_{t-1}) were used for ground water level (GWL) modeling. In addition, in the RBNN2 and SVM2 models, GWL modeling was performed with relative humidity (RH), precipitation (P), groundwater level time series (GWL_{t-1}).

In this study, 101 of the monthly relative humidity (RH), precipitation (P), groundwater level data were used for training and the 35 data were used for testing. In the modeling, Statistical criteria such as Correlation coefficient (R), Mean

Absolute Error (MAE) and Mean Square Error (MSE) were calculated and the results were interpreted by two evaluations for each model.
The R measures the strength of the correlation between the predicted and real values. MAE and MSE measure the accuracy by continuously calculating the mean size of the errors in the estimation without taking into account the aspects of the variables.

MAE and MSE are used to diagnose the possibility of errors. MAE and MSE calculations were determined according to below equations:

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|GWL_{MEASURE} - GWL_{predicted}\right| \quad (2)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(GWL_{MEASURE} - GWL_{predicted}\right)^2 \quad (3)$$

MAE, MSE and R statistics are calculated for comparison of methods used. RBNN and SVM results are given in Table 1.

Table 1. ***Statistical results of RBNN and SVM models***

MODELS	RBNN1	SVM1	RBNN2	SVM2
INPUTS	GWL_{t-1}	GWL_{t-1}	P, RH, GWL_{t-1}	P, RH, GWL_{t-1}
MAE(m)	0.5800	0.4144	0.4635	0.4048
MSE(m^2)	0.9064	0.3067	0.3414	0.2503
R	0.8794	0.9237	0.9167	0.9435

P: Precipitation (mm), RH: Relative Humidity (%), GWL: Groundwater level (m)

The most appropriate result among the models where data is used, as shown in Table 1, is given by SVM2 model with the lowest error rates and highest correlation coefficient.

Distribution and scatter graphs of RBNN1 model are shown in Figure 3 and 4, respectively.

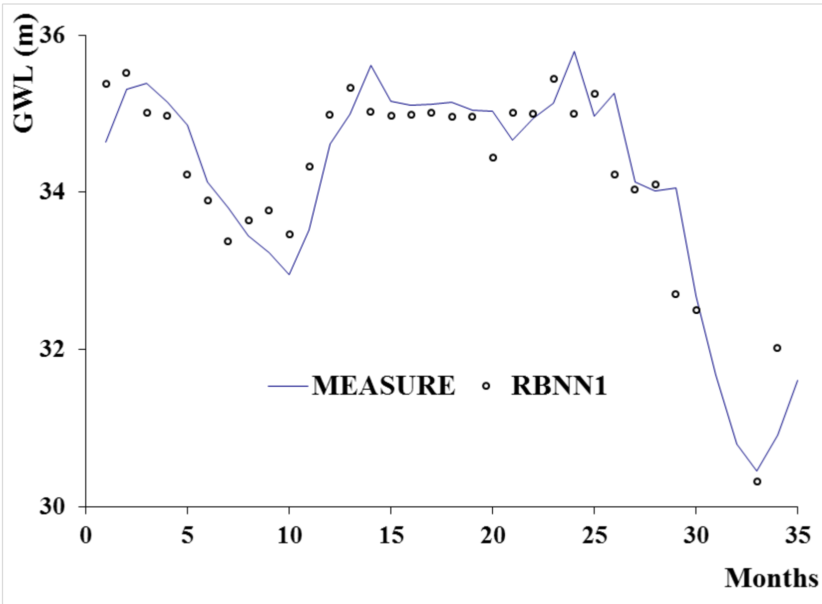

Fig. 3. ***RBNN1 model distribution charts for GWL test data***

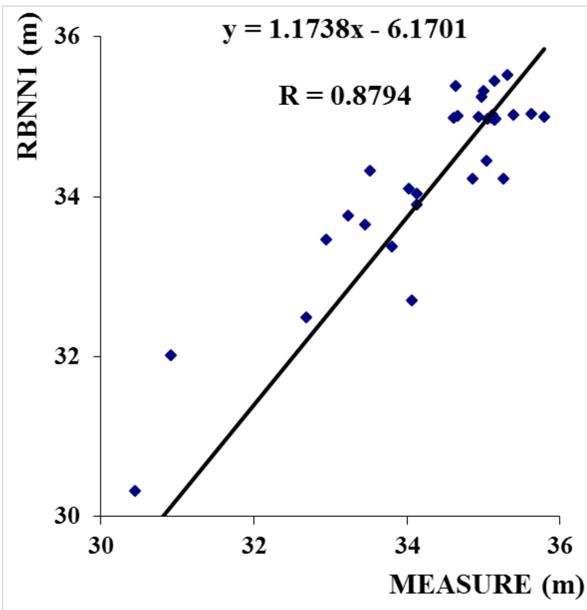

Fig. 4. ***RBNN1 model scatter charts for GWL test data***

Figure 3 and 4. show the performance of RBNN1 model. Correlation coefficient for RBNN1 model is 0.8794. Distribution and scatter graphs of SVM1 model are shown in Figure 5 and 6 below, respectively.

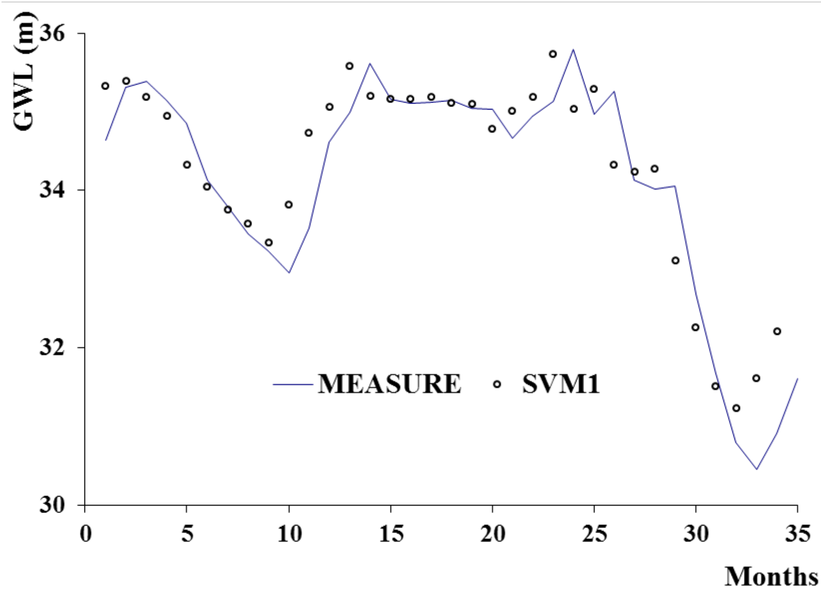

Fig. 5. ***SVM1 model distribution charts for GWL test data***

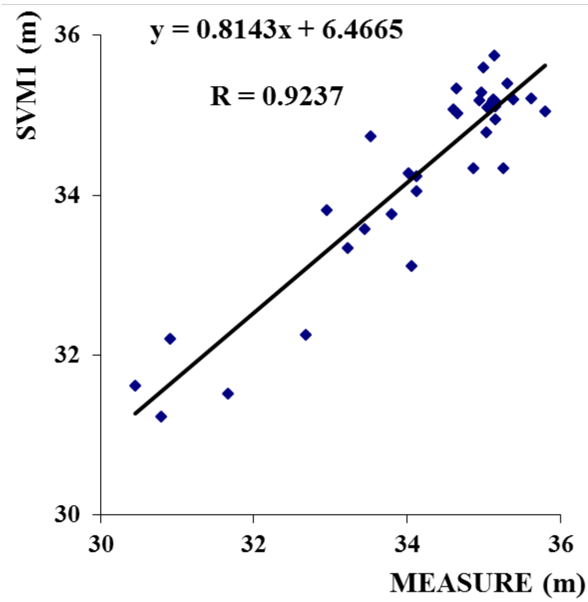

Fig. 6. ***SVM1 model scatter charts for GWL test data***

Figure 5 and 6. show the performance of SVM1 model. Correlation coefficient for SVM1 model is 0.9237. When RBNN1 and SVM1 models were examined, it was observed that RBNN1 and SVM1 model gave good results. Distribution and scatter graphs of RBNN2 model are shown in Figure 7 and 8, respectively.

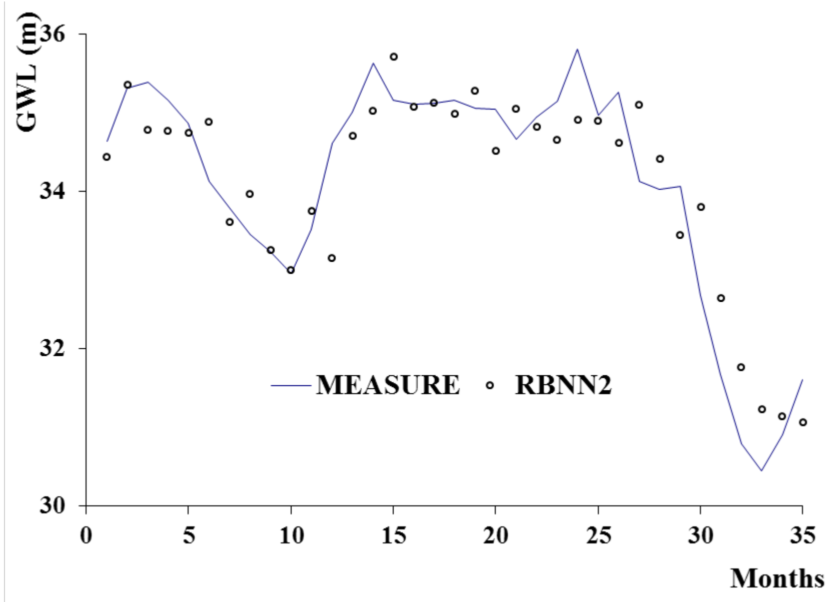

Fig. 7. ***RBNN2 model distribution charts for GWL test data***

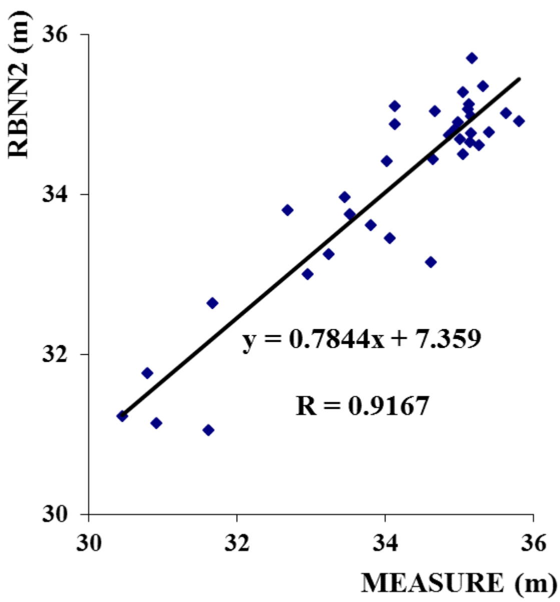

Fig. 8. ***RBNN2 model scatter charts for GWL test data***

Figure 7 and 8 show the performance of RBNN2 model. Correlation coefficient for RBNN2 model is 0.9167. Among RBNN models, RBNN2 model (R = 0.9167) have better performance. Distribution and scatter graphs of SVM2 model are shown in Figure 9 and 10 below, respectively.

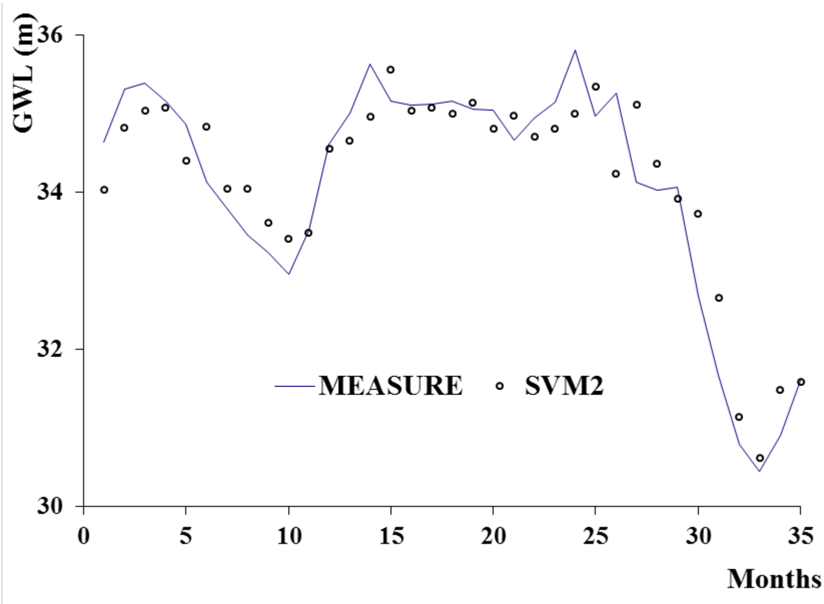

Fig. 9. ***SVM2 model distribution charts for GWL test data***

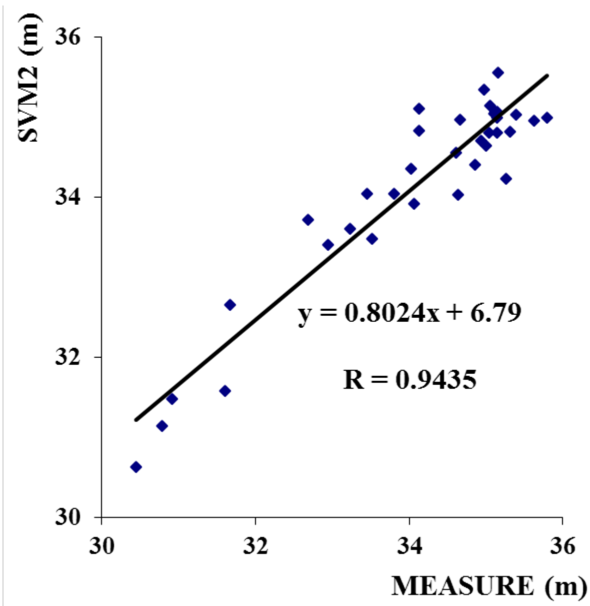

Fig. 10. ***SVM2 model scatter charts for GWL test data***

Figure 9 and 10 show the performance of SVM2 model. Correlation coefficient for SVM2 model is 0.9435. When RBNN2 and SVM2 models were examined, it was observed that RBNN2 and SVM2 model gave better results than the other models.

According to Table 1 and distribution-scatter charts, it is observed that RBNN and SVM models have good results for the test data. The good results can be expressed by a high coefficient of correlation (R) and a low error amount (MSE, MAE). Accordingly, the best estimation is given by the SVM2 model with the highest value of correlation (R = 0.9435) and the lowest error value - MSE (0.25 m^2) and MAE (0.4048 m). In addition, comparing the two RBNN models highlights that RBNN2 models is better (R = 0.9167, MSE = 0.3414 m^2 and MAE = 0.4638 m). As a result of this study, the use RBNN and SVM methods for modeling the relationship between GWL (which is one of the data mining methods) can be presented as an alternative to traditional methods.

4. CONCLUSIONS

In this study, Radial Basis Neural Network (RBNN) and Support Vector Machines (SVM) methods were used to obtain the groundwater level (GWL) estimation.

Monthly groundwater level time series was performed for the model SVM1 and RBNN1. In addition, in the model SVM2 and RBNN2, GWL modeling was performed with monthly relative humidity, precipitation, groundwater level time series. SVM model results are compared with the measured GWL quantity and the results of the RBNN method.

As a result, the low amount of error (MAE, MSE) ratios and high correlation (R) provided the desired performance in RBNN and SVM data mining methods. The reason for the high correlation of the RBNN and SVM data mining methods were that determine GWL.

Radial Basis Neural Network (RBNN) and Support Vector Machine (SVM) have been found to be a model that can be applied in the estimation of the GWL occurring with different rainfall, evaporation, humidity conditions in the studies which water planning is required and in determining the water level changes. As a final result, it is understood that data mining can be used for hydrological modelling which is necessary for water resources management and planning future requirements.

REFERENCES

1. Broomhead, D.S., Lowe, D. (1988): Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks, London.
2. Demirci, M., Üneş, F., Körlü, S. (2019). Modeling of groundwater level using artificial intelligence techniques: A case study of Reyhanli region in Turkey. Applied Ecology and Environmental Research, 17(2), 2651-2663
3. Demirci, M., Unes, F., Kaya, Y. Z., Tasar, B., Varcin, H. (2018). Modeling of dam reservoir volume using adaptive neuro fuzzy method. Aerul si Apa. Componente ale Mediului, 145-152.
4. Demirci, M., Aköz, M. S. (2013). Investigation of bar parameters occurred by cross-shore sediment transport. International Journal of Naval Architecture and Ocean Engineering, 5(2), 277-286.

5. Demirci, M., Unes, F., Kaya, Y. Z., Mamak, M., Tasar, B., Ispir, E. (2017, March). Estimation of groundwater level using artificial neural networks: a case study of Hatay-Turkey. In 10th International Conference "Environmental Engineering ".
6. Emamgholizadeh, S., Moslemi, K., Karami, G. (2014): Prediction the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS).- Water Resources Management 28(15): 5433-5446.
7. Haykin, S. (1998): Neural Networks: A Comprehensive Foundation. 2ndedition. Macmillan, New York.
8. Haykin, S. (1999). Multilayer perceptrons. Neural networks: a comprehensive foundation, 2, 135-155.
9. Heesung, Y., Jun, S.C., Yunjung, H., Bae, G.O., Kang, K.L.. (2011) : A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer.- Journal of Hydrology 396: 128-138.
10. Kaya, Y. Z., Üneş, F., Demirci, M., Taşar, B., & Varçin, H. (2018). Groundwater level prediction using artificial neural network and M5 tree models. Aerul si Apa. Componente ale Mediului, 195-201.
11. Kaya, Y.Z., Taşar, B. (2019) Evapotranspiration Calculation for South Carolina, USA and Creation Different ANFIS Models for ET Estimation. 2019 "Air and Water – Components of the Environment" Conference Proceedings, Cluj-Napoca, Romania, p. 217-224.
12. Nourani, V., Alami, M. T., Vousoughi F. D. (2015): Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. -Journal of Hydrology 524: 255-269.
13. Taşar, B., Kaya, Y.Z., Varçin, H., Üneş, F., Demirci, M., 2017. Forecasting of Suspended Sediment in Rivers Using Artificial Neural Networks Approach. Int. J. Adv. Eng. Res. Sci. 4, 79–84.
14. Taşar, B., Unes, F., Varcin, H. (2019) Prediction of the Rainfall – Runoff Relationship Using NeuroFuzzy and Support Vector Machines. 2019 "Air and Water – Components of the Environment" Conference Proceedings, Cluj-Napoca, Romania, p. 237-246
15. Üneş, F., Yıldırım, S., Cigizoğlu, H.K., Coşkun H. (2013). Estimation of Dam Reservoir Volume Fluctuations Using Neural Network and Support Vector Regression. Journal of Engg. Research, 1(3), 53-74.
16. Üneş, F., Joksimovic, D., & Kisi, O. (2015). Plunging flow depth estimation in a stratified dam reservoir using neuro-fuzzy technique. Water resources management, 29(9), 3055-3077.
17. Üneş F., Doğan S., Taşar B., Kaya Y., Demirci M. (2018), The Evaluation and Comparison of Daily Reference Evapotranspiration with ANN and Empirical Methods. Natural and Engineering Sciences, 3(3), Supplement, 54-64.
18. Üneş, F., Demirci, M., Taşar, B., Kaya, Y. Z., & Varçin, H. (2019a). Estimating Dam Reservoir Level Fluctuations Using Data-Driven Techniques. Polish Journal of Environmental Studies, 28(5).
19. Unes, F., Demirci, M., Tasar, B., Kaya, Y. Z., & Varçin, H. (2019b). Modeling of dam reservoir volume using generalized regression neural network, support vector machines and M5 decision tree models. *Applied Ecology And Environmental Research*, 17(3): 7043-7055.
20. Vapnik, V.N. (1998): .Statistical Learning Theory. Wiley, New York, USA

21. Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE transactions on neural networks, 10(5), 988-999.
22. 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 Research18: 63-76.