Environment Protection Engineering

Vol. 50 2024 No. 1

DOI: 10.37190/epe240106

PRACHI SINGH (ORCID: 0000-0003-4460-0537)[1]
VIKAS MAHOR (ORCID: 0000-0002-1304-4964)[2]
NATRAYAN LAKSHMAIYA (ORCID: 0000-0003-1001-3422)[3]
KANDUKURI SHANKER (ORCID: 0000-0002-5982-8983)[4]
SEENIAPPAN KALIAPPAN (ORCID: 0000-0002-5021-8759)[5]
MUTHIAH MUTHUKANNAN (ORCID: 0000-0003-1912-3513)[6]
GOPINATH RAJENDRAN (ORCID: 0000-0003-1820-7088)[7]

# PREDICTION OF GROUNDWATER CONTAMINATION IN AN OPEN LANDFILL AREA USING A NOVEL HYBRID CLUSTERING-BASED AI MODEL

Groundwater is a vital resource that provides drinking water to over half of the world's population. However, groundwater contamination has become a serious issue due to human activities such as industrialization, agriculture, and improper waste disposal. The impacts of groundwater contamination can be severe, including health risks, environmental damage, and economic losses. A list of unknown groundwater contamination sources has been developed for the Wang-Tien landfill using a groundwater modeling system (GMS). Further, AI-based models have been developed which accurately predict the contamination from the sources at this site. A serious complication with most previous studies using artificial neural networks (ANN) for contamination source identification has been the large size of the neural networks. We have designed the ANN models which use three different ways of presenting inputs that are categorized by hierarchical *K*-means clustering. Such an implementation reduces the overall complexity of the model along with high accuracy. The predictive capability of developed models was assessed using performance indices and compared with the ANN models. The results show that

[1]Department of Civil Engineering, Madhav Institute of Technology and Science, Gwalior 474011, India.

[2]Department of Electronics Engineering, Madhav Institute of Technology and Science, Gwalior 474011, India, corresponding author, email address: vikas@mitsgwalior.in

[3]Department of Mechanical Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India.

[4]Civil Engineering Department, Kamala Institute of Technology and Science. Singapur, Karimnagar, Telangana, India.

[5]Department of Mechatronics Engineering, KCG College of Technology, Karapakkam Chennai 600097, Tamil Nadu, India.

[6]Department of Civil Engineering, KCG College of Technology, Karapakkam, Chennai 600097, India.

[7]QIS College of Engineering and Technology, Vengumukpalem, Ongole-523272, AP, India.

the hybrid model of hierarchical $K$-means clustering and ANN model (HCA-ANN) is a highly accurate model for identifying pollution sources in contaminated water.

## 1. INTRODUCTION

Groundwater contamination is a serious environmental issue in many urban areas, posing a significant threat to public health and the environment. Groundwater contamination prediction is a crucial component of groundwater management, as it helps to identify potential sources of contamination and take appropriate measures to prevent or mitigate the impacts of contamination. In recent years, there has been significant research in the field of groundwater contamination prediction, aimed at developing reliable models and tools for assessing the vulnerability and risk of groundwater contamination.

Several studies have focused on developing predictive models that use a combination of statistical and machine-learning techniques to identify the factors that contribute to groundwater contamination [1–4]. These models consider various parameters such as land use, hydrogeological characteristics, and environmental factors, to identify areas that are most vulnerable to contamination. Other studies have used geospatial analysis and remote sensing techniques to map the distribution of groundwater contamination, providing valuable insights into the sources and extent of contamination.

The use of advanced analytical techniques such as isotopic and geochemical analyses has also shown promise in predicting groundwater contamination. These techniques help to identify the sources and pathways of contaminants, enabling better management and remediation strategies.

However, despite the progress made in the field of groundwater contamination prediction, several challenges remain. The complex nature of groundwater systems, variability in contaminant behavior, and the lack of reliable data remain major obstacles to developing accurate predictive models. Furthermore, the effectiveness of these models in real-world applications remains to be tested.

While significant progress has been made in the field of groundwater contamination prediction, more research is needed to develop accurate and reliable predictive models. In this research, a novel hybrid model of hierarchical $K$-means clustering and ANN model (HCA-ANN) is utilized to reduce the complexity of the AI model and increase the accuracy of the prediction model.

Atmadja et. al. [5] divided the problem of identifying groundwater contamination sources into two categories. Firstly, to reconstruct the source release history, we need to locate a known source location. In the second problem, the location of the source is determined using the concentration and time of the release. The researchers also examined the methods that had been developed for identifying the release history and source location, including optimization, probabilistic, and geo-statistical simulation, and analytical approaches based on regression. Liu and Ball [6] divided this issue into two categories when it came to recovering the release history. The release history is assumed to be a function first and then reformulated as an optimization problem. An estimation method using a gradient type and a non-gradient type is used for estimating the release function's best-fit parameters. Comparing the simulated and sampling concentrations, the full estimation approach reconstructs the release history [7].

Groundwater transport simulation models based on linear programming and multiple regressions were used [8] to estimate the source information. As a result of a modified finite element model and limited data from monitoring wells. A method of estimating source location and time history in heterogeneous sites was proposed in [7] using particle methods. Bagtzoglou and Hossain [9] investigated different types of source information estimation problems using finite difference methods. By using this method, they approximated the groundwater flow equation and the transport equation in two dimensions.

An artificial neural network (ANN) was used to identify unknown pollution sources by Singh et al. [10]. For identifying unknown sources of water, the GA-based simulation optimization technique was used by Sun [11]. They studied the application of ANN methods based on varying data availability and errors in concentration measurements. To identify unknown sources of groundwater pollution, they used a simulation optimization technique based on GA. They evaluated the performance of the designed methodology based on the characteristics of the sources.

Mahinthakumar and Sayeed [12] solved the source identification problem of groundwater using the hybrid genetic algorithm local search (GA-LS) method. As a result of GA, this group obtained results that then served as the basis for the local search, to identify the global optimum. They also studied various hybrid GA-LS optimization techniques to determine the contamination source information. In steady-state flow fields, single and multiple-source releases in 3D heterogeneous flows were investigated. Results showed that the source location identification problem is much more challenging than the release history recovery problem.

Yeh et al. [13] investigated sources using a backward probability model. SATS-GWT is an approach designed for estimating water contamination source information, including the location of the source, concentration of the waste released, and the time for which the waste is released, they used a combination of simulated annealing (SA), tabu search (TS), and the MODFLOW-GWT groundwater flow and solute transport model. MODFLOW-GWT simulated sampling concentrations at monitoring points based on a known source and concentration during release. Zou et al. [14] proposed a neural network-based generative algorithm that overcomes the computational limitations of inverse modeling by replacing the water quality model with an efficient NN-based functional model. Bagtzoglu and Hossain examined ANN applications for site characterization in 2009 based on a geo-statistical approach to identify contaminants in a 2D aquifer.

Nayak et al. [15] forecasted groundwater levels using ANN. The approach was applied to an unconfined aquifer system located in Godavari Delta, India. The data selection was done by applying statistical analysis of available datasets. The groundwater levels in two wells were predicted using many ANN models. The sensitivity analysis of the developed approach was also performed. The results showed the novelty of ANN models in simulating groundwater levels.

Behzad et al. [16] forecasted water levels of aquifer systems using SVM and ANN . The forecasting abilities of both approaches were also compared. Both models provided almost similar results in all the cases. However, SVM outstripped the results of ANN predicting a longer duration. This suggests that SVM can be effectively applied for future GWL prediction when the data availability is limited. Jalalkamali et al. [17] applied a neuro-fuzzy approach to predict the monthly frequency of groundwater levels in Kerman, Iran.

Lee et al. [18] mapped groundwater quality in Boryeong, Korea using GIS, ANN, and SVM approaches. The ANN model was developed using the BP algorithm while a polynomial kernel was applied to develop the SVM approach. Both models showed almost similar accuracy as ANN showed 83.57% and SVM 80.83%. This shows that both ANN and SVM approaches can be effectively utilized for groundwater-related studies.

ANN- and SVM-based AI models have great potential for water contamination source detection. The models have shown high accuracy in detecting contamination sources, and they can be used as a tool for effective management of water resources. However, further research is needed to optimize the models and to integrate them with other data-driven techniques for improved performance. Hierarchical *K*-means clustering followed by ANN models has been extensively used for prediction tasks in various domains [19]. As per the extensive literature survey, this method is not utilized for the prediction of water contamination sources. In this research, the generated water contamination data is divided into clusters using a hierarchical *K*-means clustering algorithm and further, it is utilized for training and testing the ANN model. This multi-stage approach reduces the complexity of the ANN model and also increases the accuracy of the prediction.

## 2. AI MODEL DESIGN METHOD FOR WATER CONTAMINATION PREDICTION

This section briefs about the standard methodology utilized in AI-based model development for water contamination source identification. The methodology adopted for contamination source identification sequentially is as follows (Fig. 1):

- data generation/generation of patterns,
- simulation of spatial and temporal responses,
- breakthrough curve characterization,
- AI model building,
- performance evaluation of developed models.

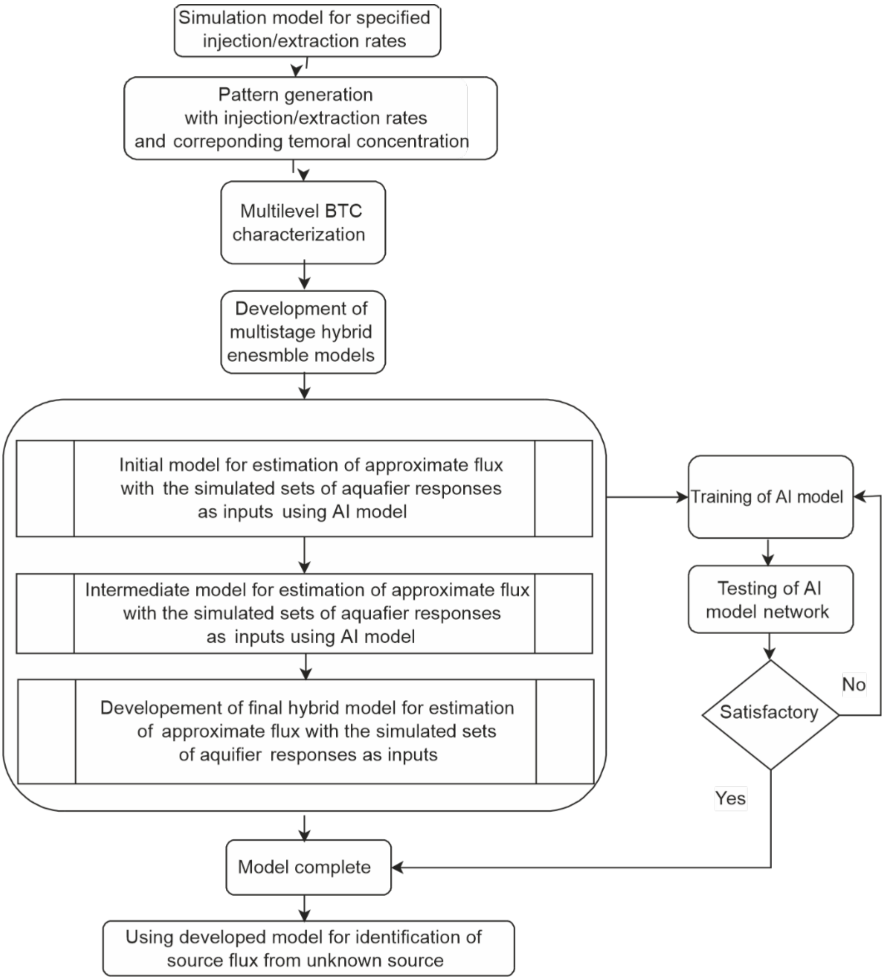


Fig. 1. Methodology of model development

To understand and mitigate the effects of water contamination, it is important to generate realistic and representative data. One approach to generating such data is by using Latin hypercube sampling in MATLAB [20]. Latin hypercube sampling is a statistical sampling technique that generates samples that are evenly spaced along each dimension of the input variables. In the context of water contamination, the input variables could include pH, temperature, concentration of contaminants, etc.

To generate water contamination data using Latin hypercube sampling in MATLAB, the first step is to define the input variables and specify the range of values for each variable. Next, the number of samples is determined and the Latin hypercube sample matrix is generated in MATLAB. This sample matrix is then mapped to the input variable ranges. Once the Latin hypercube sample matrix has been generated, the water contamination data can be simulated for each sample using a water contamination model. This model could be a mathematical model or a simulation model that describes the behavior of water contaminants under different conditions.

Finally, the results can be visualized using plots or other graphical representations. Groundwater Management Systems (GMS) [21] simulate the spatial and temporal responses based on these data. These responses produce a breakthrough curve (BTC) at each location. Each breakthrough curve is characterized using multistage model development [22]. These models are then evaluated using different performance indices.

Groundwater simulation involves specifying equations that describe the process of flow, defining the boundaries of the aquifer, and setting up the aquifer's initial conditions. Simulation of groundwater flow and contamination transport in the groundwater system has been carried out using the Groundwater Management System (version 7.1, 2011) developed by Aquaveo [21].

Using Pinder and Bredehoeft [23], a nonhomogeneous anisotropic and saturated aquifer can be represented by the following equation as a steady-state, three-dimensional flow:

$$\frac{\partial}{\partial x}\left(K_x\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_z\frac{\partial h}{\partial z}\right)\pm Q=S_s\frac{\partial h}{\partial t} \tag{1}$$

$x$, $y$, and $z$ are the directions in which the hydraulic conductivity coefficients are measured, $h$ stands for the potentiometric head, sources and sinks are represented by $Q$ per unit volume, and time is represented by $t$.

A transient three-dimensional equation can be written as follows [23] for transient solute transport in a saturated aquifer:

$$\frac{\partial C}{\partial t}=\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C}{\partial x}\right)-\frac{\partial}{\partial x_i}\left(v_iC\right)+\frac{q_s}{\theta}C_s+\sum_{k=1}^{N}Q_h \tag{2}$$

$C$ *is* the contaminant concentration, $q_s$ being a specific discharge of fluid, $Q_h$ volume of source/sink, the distance $x_i$ along the respective axis is $D_{ij}$, and the seepage is $v_i$, sources and sinks are represented by $C_s$ and porosity by $\theta$.

## 3. STUDY AREA

Located in the Yong-Kang Municipality of Tainan, the Wang-Tien landfill site is the subject of the study [24] (Fig. 2). A total of 773 970 m$^3$ of solid waste accumulated in the landfill from 1992 until 2002 when it was decommissioned. In the study area, the longitudes are 23°2′29″ N and 23°2′36″ N, and the latitudes are 120°16′1″ E and 120°16′22″ E. The total area is 39 333 m$^2$ and the longitudes are 23°2′29″ N and 23°2′36″ N [24]. With an elevation range of 8.75–25 m above mean sea level and a slope of 0.1% from southeast to northwest, the study area is situated above mean sea level. Currently, the landfill is surrounded by industrial areas of Yong-Kang and Hsu-Hsian Creek, which drains into Yan-Shuei Creek. The site consists of an area 175 m North-South and 225 m East-West.


Fig. 2. Study area of Wang-Tien landfill site [24]

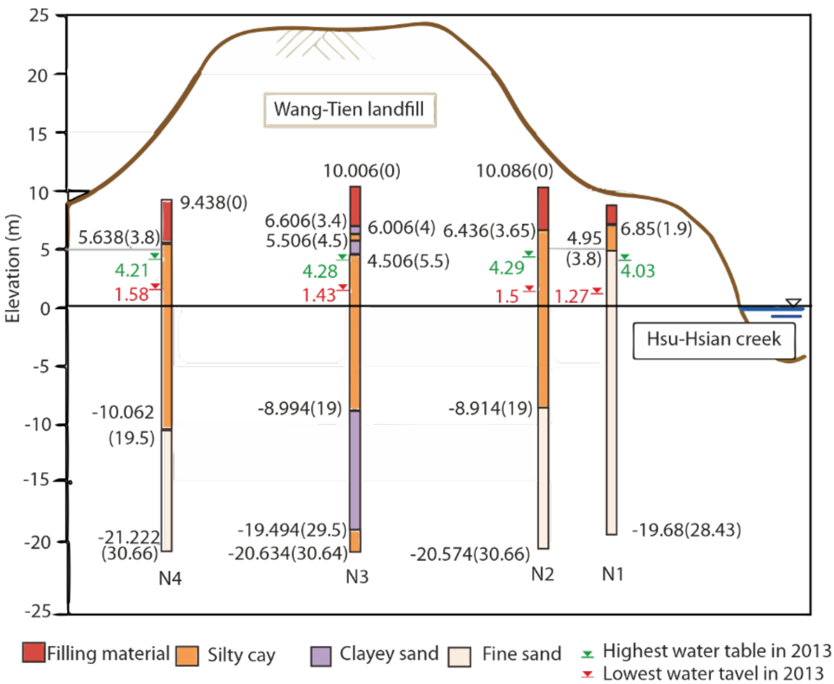


Fig. 3. Layer details of Wang-Tien landfill site [24];
the depth below the ground surface is given in brackets in m

The details of elevation and material of layers are shown in Fig. 3. The average precipitation in the Tainan area is 1828.4 mm/year and the average evaporation is 1476 mm/year in the field area. The monthly variation in recharge in 2013 is shown in Fig. 4. Model parameters for flow and contaminant transport modeling are tabulated in Table 1. The magnitude of source fluxes at different locations and time duration is presented in Table 2.

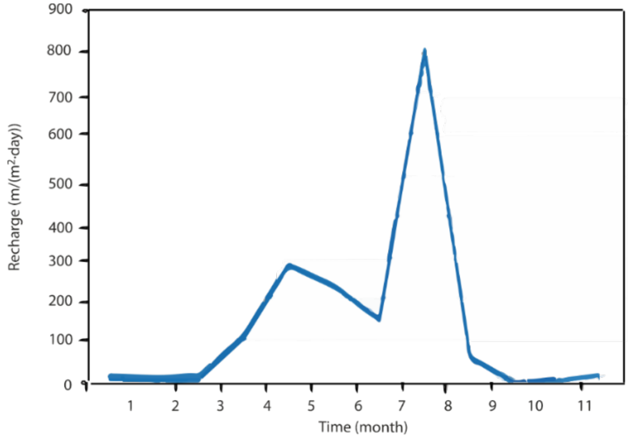


Fig. 4. Transient variation of recharge in 2013

Table 1

Model Parameters for the Wang-Tien Landfill Site [24]

| Model parameters | Value |
|---|---|
| Longitudinal dispersivity $\alpha_L$, m | 2.5 |
| Transverse dispersivity $\alpha_{TH}$, $\alpha_{TV}$, m | 0.5 |
| Filling material porosity $\theta$ | 0.35 |
| silty clay layer | 0.28 |
| fine sand layer | 0.3 |
| clayey sand layer | 0.04 |
| Hydraulic conductivity for filling material layer, m/s | $1.2610^{-4}$ |
| silty clay layer | $1.26\times10^{-5}$ |
| fine sand layer | $7.17\ 10^{-4}$ |
| clayey sand layer | $7.17\times10^{-7}$ |

Table 2

Source flux from different sources

| Time Step | Source [g/s] | | | |
|---|---|---|---|---|
| | S1 | S2 | S3 | S4 |
| 1 | 15.49 | 21.21 | 10.60 | 8.73 |
| 2 | 2.2 | 1.50 | 3.20 | 0.88 |

## 4. ANN MODELS FOR IDENTIFYING CONTAMINATION SOURCES

To identify the potential pollution source using breakthrough curves, the pollution source is assumed to be active for a constant duration injecting a conservative pollutant at a constant rate. Each BTC contains 20 time intervals, which are then used for making different models. A total of 250 input-output patterns were generated, of which 175 were used for the training and 75 for testing. Four ANN models were generated using these data to predict the source magnitude. The outputs are eight corresponding source flux values as the target in each pattern. The different models are described later in the paper.

The multistage models viz., initial (single-stage), intermediate stage (second stage), and final stage (third stage), are developed for the identification of groundwater sources. A total of eight different models have been developed using ANN and hybrid (ANN-SVM) approach. For the nomenclature of models, I-models indicate single-stage, II-models and III-models indicate the intermediate (second) stage, and IV-models indicate final (third) stage models. In I-models, the inputs are the twenty simulated temporal concentrations. The predicted outputs from the I-model, i.e., Ninety-six source fluxes predicted from all the well models, are utilized as the input for the II-model. Similarly, the twelve average values of each source flux predicted from the I-model from different wells are utilized as the input for the III-model. The source fluxes predicted from the II-model are the inputs into the IV--model. The source fluxes that cause these values of concentration in observation wells are the outputs of the models.

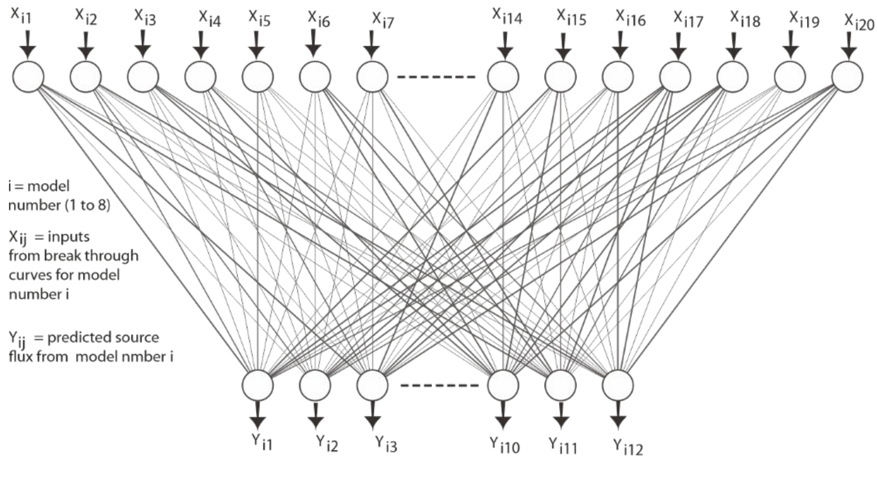


Fig. 5. ANN-I model architecture

Figure 5 illustrates the ANN-I model architecture. This model consists of twenty simulated temporal concentration measurements ($X_{i1}$–$X_{i20}$) (complete BTC) as input in each well, $i$ corresponds to the number of models, i.e., 8. The output of the model is twelve corresponding source fluxes as target values per pattern ($y_{i1}$–$y_{i12}$) for each set of inputs, output $Y_{ij}$ obtained from one set of input $X_{ij}$ is illustrated in Fig. 5.

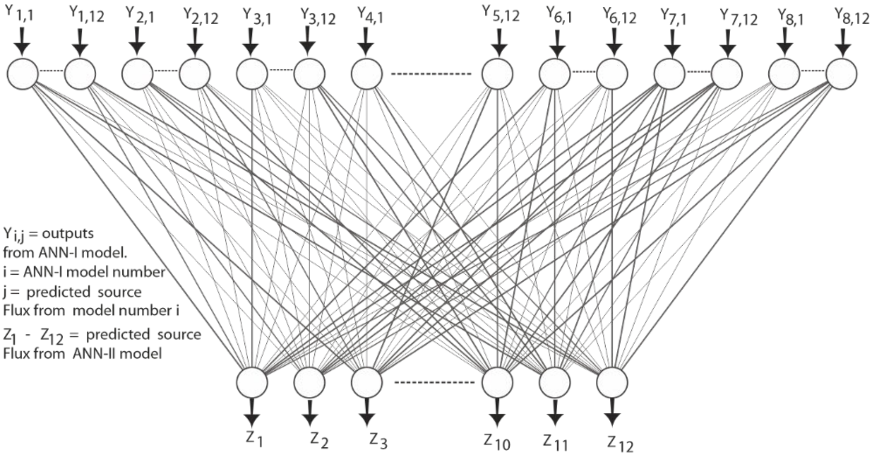


Fig. 6. ANN-II model architecture

Figure 6 illustrates the ANN-II model architecture. The model consists of 96 input values which are predicted as $Y_{ij}$ (by applying it to each observation well, i.e., $i = 8$ and $j = 12$ corresponds to the output obtained from ANN-I model for each input pattern). Hence, $j = 12$ inputs are applied as inputs in each ANN-II model, for 8 such models leading to 96 inputs. Twelve corresponding source fluxes as the target in ($Z_1$–$Z_{12}$) are obtained for each pattern $i$.

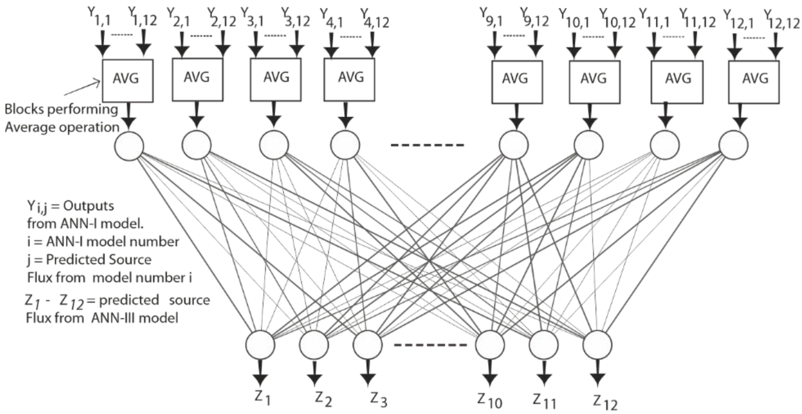


Fig. 7. ANN-III model architecture

Similarly, the ANN-III model (Fig. 7) consists of twelve average values of source fluxes as inputs which are predicted from the ANN-I model (by applying it to each observation well), and twelve corresponding source flux values as the target in each pattern. The same approach is used to design the ANN-IV model which consists of twelve input values predicted from ANN-II models, and twelve corresponding source flux values as the target in each pattern.

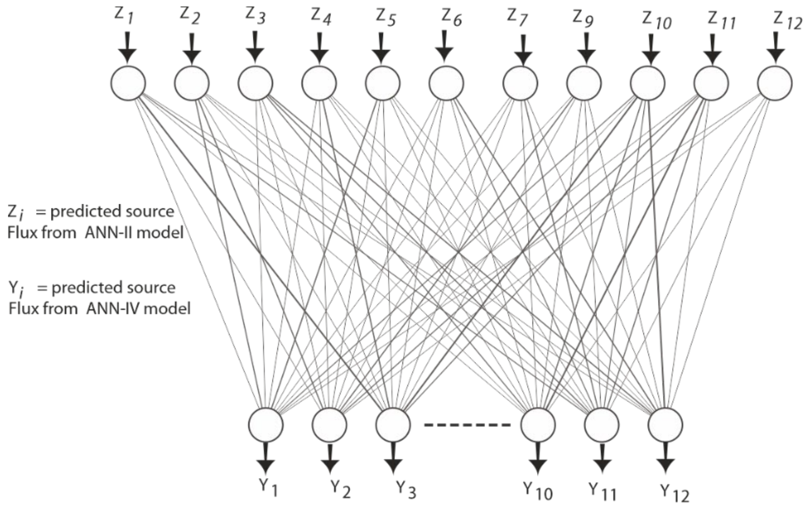

Fig. 8. ANN-IV model architecture

The type-1 model consists of 20 simulated temporal concentration measurements (complete BTC) as input in each well (Fig. 5).

The type-2 model consists of 32 input values which are predicted from the type-1 models (by applying it to each observation well) (Fig. 6).

The type-3 model consists of 8 average values of source fluxes as inputs which are predicted from the type-1 model (by applying it to each observation well) (Fig. 7).

The type-4 model consists of 8 input values which are predicted from the type-2 model (Fig. 8).

## 5. NOVEL HCA-ANN MODEL FOR PREDICTION OF GROUNDWATER CONTAMINATION

This section describes a new hybrid model, i.e., the HCA-ANN model, based on Hierarchal *K*-means clustering Algorithm (HCA) and ANN. The training dataset was first partitioned into k clusters by HCA. We then developed ANN models based on the acquired clusters. The CLARA method was applied to individual k clusters on training dataset [25]. Total Within Sum of Square (TWSS) was calculated for various values of *k* as shown in Fig. 9.

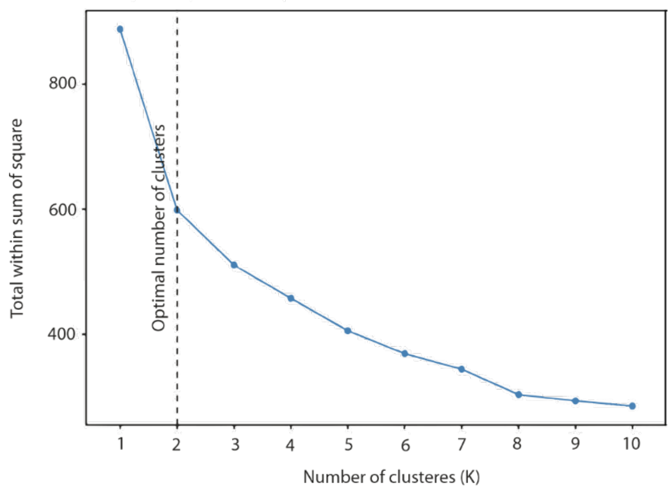


Fig. 9. Total within the sum of square (TWSS) vs. number of clusters graph for determining the optimal value of the number of required clusters ($K$)

It can be observed that the TWSS decreases sharply in the range of 1–2 of $K$, and the TWSS decreases gradually from 2 (Fig. 9). This result recommends that the entire data set of resource flux should be divided into 2 clusters. The first cluster contains 94 observations. The second cluster consists of 81 observations, as shown in Fig. 10.

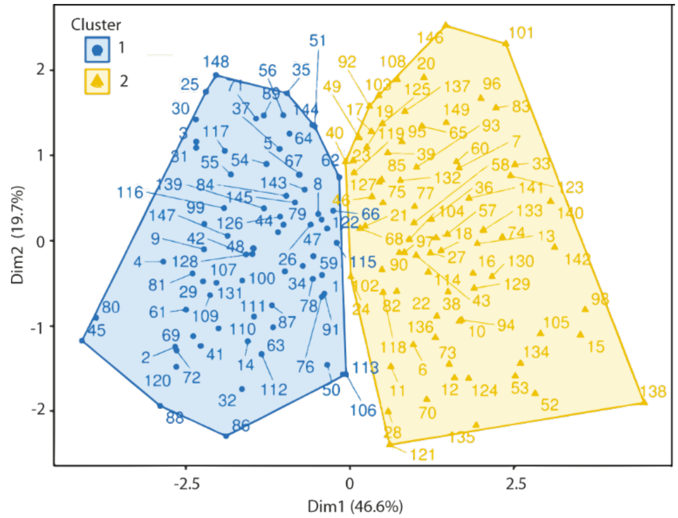


Fig. 10. Division of the given dataset in two clusters as per the HCA algorithm

By HCA clustering data with homologous characteristics were grouped into clusters. Those are represented by two principal components, Dim 1 and Dim 2. These two principal components account for 19.7 + 46.6 = 66.3% of the variance in the data set. After separating the clusters with the HCA algorithm, an ANN model was considered and developed based on the result clusters. For each cluster, best-fit ANN models were developed to estimate the flux generated by the contamination. The block diagram of the proposed HCA-ANN model for flux prediction is shown in Fig. 11.

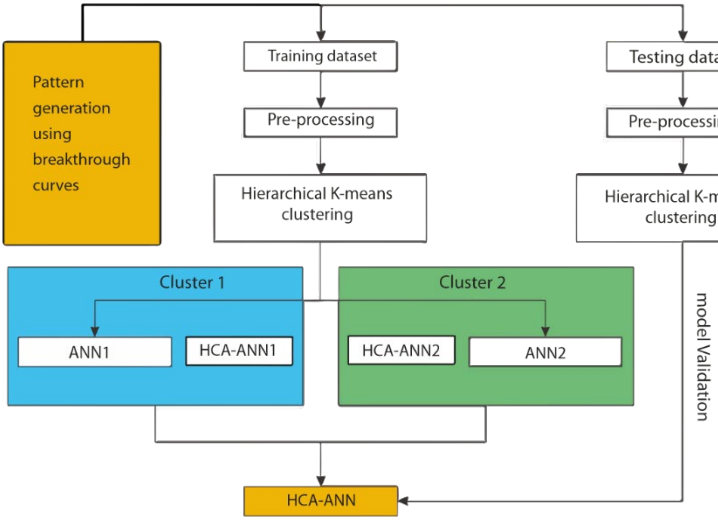


Fig. 11. Flowchart of the proposed HCA-ANN model to predict flux in this study

Table 3

Performance indices of ANN models

| Model | APE [%] | | Correlation coefficient $R$ | | Normalised error $NE$ | | PBIAS [%] | | MENASH | | RMSE [%] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test |
| ANN-I | 28.36 | 26.8 | 0.5 | 0.85 | 0.51 | 0.66 | –0.01 | 0.793 | 0.25 | 0.34 | 2.15 | 1.99 |
| ANN-II | 23.26 | 22.87 | 0.56 | 0.81 | 0.35 | 0.46 | 1.45 | 4.94 | 0.4 | 0.59 | 1.78 | 1.85 |
| ANN-III | 23.66 | 19.08 | 0.55 | 0.86 | 0.33 | 0.32 | –0.24 | –0.094 | 0.65 | 0.63 | 2.01 | 1.49 |
| ANN-IV | 25.7 | 22.37 | 0.49 | 0.84 | 0.23 | 0.29 | 0.04 | 0.586 | 0.31 | 0.51 | 2.07 | 1.69 |

## 6. RESULTS AND DISCUSSION

Figure 11 shows the pictorial representation of the concentration changes with time, based on a ten-year simulated plume for the study area. For 180 days, it can be seen that the contour concentration has been evolving. Further, Fig. 12b–i indicates the contour concentration is moving towards Hsu-Hsian Creek, which is potentially contaminating usable water.

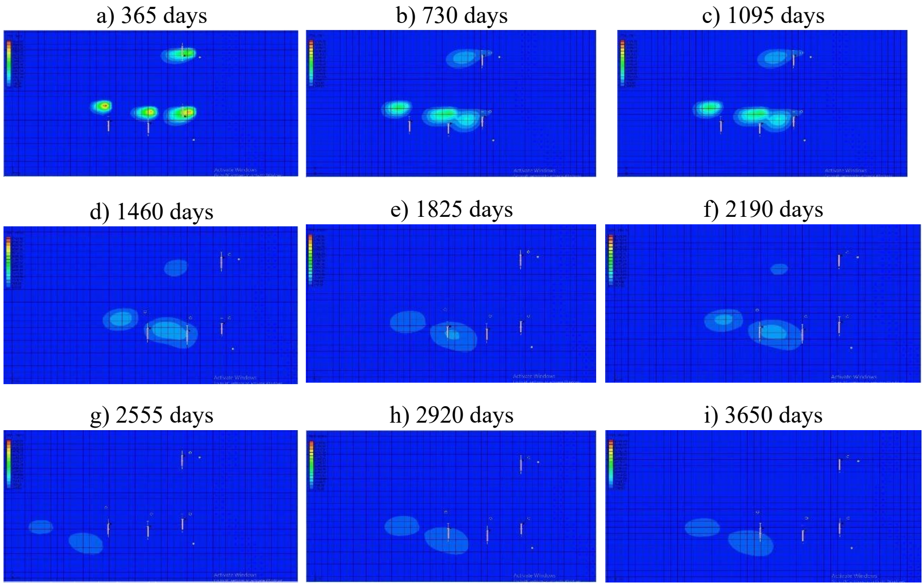


Fig. 12. Contaminant transport contours at different time periods;
contamination levels: blue – lowest to red – highest)

We evaluate the developed models for the demonstrative identification problem, where the magnitude of source fluxes is unknown. The proposed HCA-ANN model performed comparatively better in training and testing phase. Further, this model is applied to identify the magnitude of contaminant sources. The actual and predicted source flux are compared in Fig. 13 with the respective normalized error (*NE*) values.

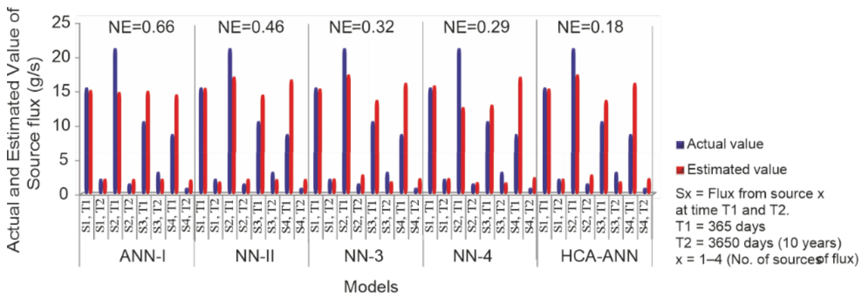


Fig. 13. Actual and predicted source flux of different models
with their respective normalized error (*NE*) values

The proposed HCA-ANN model provided the best performance for predicting contamination flux in this study (Fig. 13). The results indicate that the HCA-ANN algorithm seems more robust than the ANN algorithm alone. In Figure 14, measured and predicted values of the proposed HCA-ANN model are compared. The HCA-ANN model has the best convergence and hence it can predict the source flux with more accuracy as compared with other models.

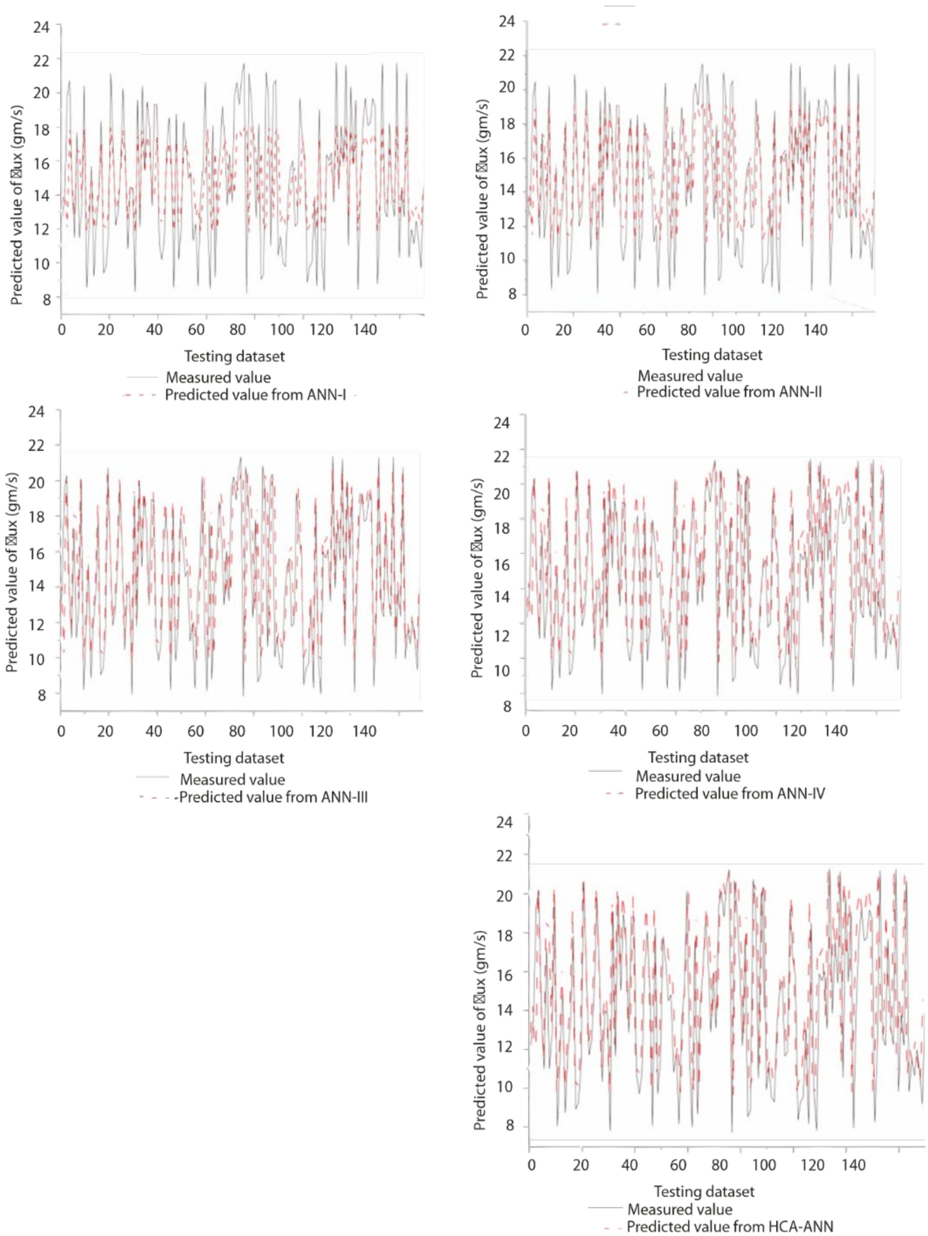


Fig. 14. The accuracy of the models on the testing dataset

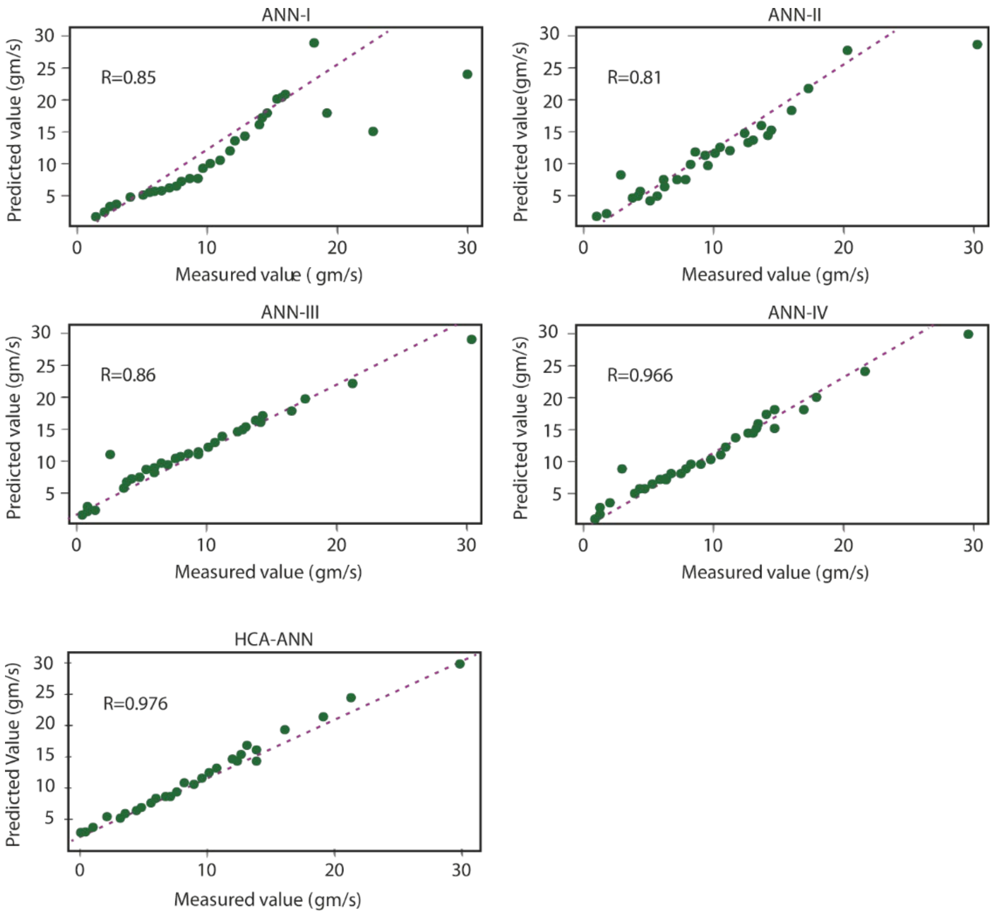


Fig. 15. Correlation of measured flux value and predicted flux values for different models

The measured and predicted values of the different models and the proposed HCA-ANN model have been plotted in Fig. 15. The best convergence on the regression line is obtained with HCA-ANN with $R = 0.976$. This clearly explains the superior performance of the proposed HCA-ANN model compared with the other models.

## 7. CONCLUSIONS

The soft computing-based contamination source identification method has been developed. The applicability of the developed method has been demonstrated for various types of models using different methods of representing inputs. Multilevel models are developed using ANN. Further, the hierarchical *K*-means clustering algorithm is employed to cluster the dataset, and the hybrid predicting models were developed by using

ANN and hierarchical *K*-means clustering. These models were applied to a complex real area. The transport pattern of the contamination is also obtained by using GMS simulation models. 250 input-output patterns were generated in each scenario, out of which 175 were used for the training and 75 for testing. Each breakthrough curve (BTC) contains 20 time intervals (each of 6 months duration), which are then used for making different models. All four models are developed using ANN. These 175 input patterns are clustered into two different clusters. Further, best best-performing ANN algorithm was identified and used on these two developed clusters. The predictive effectiveness of developed models was evaluated based on performance indices. The results show that the HCA-ANN model approach can accurately predict the groundwater contamination sources.

## REFERENCES

[1] MICHALEC B., CUPAK A., *Assessment of quality of water and sediments in small reservoirs in Southern Poland. A case study*, Environ. Eng. Res., 2022, 27 (2), 200660. DOI: 10.4491/eer.2020.660.

[2] ADDA A., HANINI S., BEZARI S., LAIDI M., ABBAS M., *Modeling and optimization of small-scale NF/RO seawater desalination using the artificial neural network (ANN)*, Environ. Eng. Res., 2022, 27 (2), 200383. DOI: 10.4491/eer.2020.383.

[3] KABSCH-KORBUTOWICZ M., KUTYLOWSKA M., *The possibilities of modelling the membrane separation processes using artificial neural networks*, Environ. Prot. Eng., 2008, 34 (1), 15–36.

[4] BLAGOJEVIĆ M., PAPIĆ M., VUJIČIĆ M., SUCUROVIC M., *Artificial neural network model for predicting air pollution. Case study of the Moravica District, Serbia*, Environ. Prot. Eng., 2018, 44 (1), 131–139. DOI: 10.37190/epe180110.

[5] ATMADJA J., BAGTZOGLOU A.C., *State of art report on mathematical methods for groundwater pollution source identification*, Environ. Forens., 2001, 2 (3), 205–214. DOI: 10.1006/enfo.2001.0055.

[6] LIU C., BALL W.P., *Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware*, Water Res. Res., 1999, 35 (7), 1975–1985. DOI: 10.1029 /1999WR900092.

[7] ABABOU R., BAGTZOGLOU A.C., MALLET A., *Anti-diffusion and source identification with the RAW scheme: A particle-based censored random walk*, J. Environ. Fluid Mech., 2010, 10 (1–2), 41–76. DOI: 10.1007 /s10652-009-9153-4.

[8] GORELICK S.M., EVANS B., REMSON I., *Identifying sources of groundwater pollution. An optimization approach*, Water Res. Res., 1983, 19, 779–790. DOI: 10.1029/WR019i003p00779.

[9] BAGTZOGLOU A.C., HOSSAIN F., *Radial basis function neural network for hydrologic inversion: An appraisal with classical and spatiotemporal geostatistical techniques in the context of site characterization*, Stoch. Environ. Res. Risk Assess., 2009, 23 (7), 933–945. DOI: 10.1007/s00477-008-0262-2.

[10] SRIVASTAVA D., SINGH R.M., *Breakthrough curves characterization and identification of an unknown pollution source in groundwater system using an artificial neural network (ANN)*, Environ. Forens., 2014, 15 (2), 175–189. DOI: 10.1080/15275922.2014.890142.

[11] SUN A.Y., *A robust geostatistical approach to contaminant source identification*, Water Res. Res., 2007, 43, 1–12. DOI: 10.1029/2006WR005106.

[12] MAHINTHAKUMAR G., SAYEED M., *Hybrid genetic algorithm. Local search methods for solving groundwater source identification inverse problems*, J. Water Res. Plan. Manage., 2005, 131 (1), 45–57. DOI: 10.1061 /(ASCE)0733-9496 (2005)131:1(45).

[13] YEH H.D., CHANG T.H., LIN Y.C., *Groundwater contaminant source identification by a hybrid heuristic approach*, Water Res. Res., 2007, 43 (9), W09420. DOI: 10.1029/2005WR004731.

[14] ZHOU S., WU Y., NI Z., ZHOU X., WEN H., ZOU Y., *Dorefa-net: Training low bitwidth convolutional neural networks with low bitwidth gradients*, 2016, arXiv preprint arXiv:1606.06160.

[15] NAYAK P.C., RAO Y.S., SUDHEER K.P., *Groundwater level forecasting in a shallow aquifer using artificial neural network approach*, Water Res. Manage., 2006, 20, 77–90. DOI: 10.1007/s11269-006-4007-z.

[16] BEHZAD M., ASGHARI K., COPPOLA Jr. E.A., *Comparative study of SVMs and ANNs in aquifer water level prediction*, J. Comp. Civ. Eng., 2010, 24 (5), 408–413. DOI: 10.1061/ (ASCE)CP.1943-5487. 0000043.

[17] JALALKAMALI A., SEDGHI H., MANSHOURI M., *Monthly groundwater level prediction using ANN and neuro-fuzzy models: a case study on Kerman plain, Iran*, J. Hydroinf., 2011, 13 (4), 867–876. DOI: 10.2166 / hydro.2010.034.

[18] LEE J.Y., YI M.J., YOO Y.K., AHN K.H., KIM G.B.,WON J.H., *A review of the national groundwater monitoring network in Korea*, Hydrol. Proc., 2007, 21 (7), 907–919. DOI: 10.1002/hyp.6282.

[19] PING W.L., PHUAN A.T.L., JIAN X., *Hierarchical fast learning artificial neural network: progressive learning in high dimensional spaces*, 2004, International Report, https://neuro.bstu.by/ai/To-dom /My_research/failed%201%20subitem/For-research/D-mining/Iris-kmean/InternalReport2004.pdf

[20] SHEIKHOLESLAMI R., RAZAVI S., *Progressive Latin hypercube sampling. An efficient approach for robust sampling-based analysis of environmental models*, Environ. Model. Softw., 2017, 93, 109–126. DOI: 10.1016/j.envsoft.2017.03.010.

[21] RINK K., KALBACHER T., KOLDITZ O., *Visual data exploration for hydrological analysis*, Environ, Earth Sci., 2012, 65, 1395–1403. DOI: 10.1007/s12665-011-1230-6.

[22] SINGH P., SINGH R.M., *Identification of pollution sources using artificial neural network (ANN) and multilevel breakthrough curve (BTC) characterization*, Environ. Forens., 2019, 20 (3), 219–227. DOI: 10.1080/15275922.2019.1629548.

[23] BREDEHOEFT J.D., PINDER G.F., *Mass transport in flowing groundwater*, Water Res. Res., 1973, 9 (1), 194–210. DOI: 10.1029/WR009i001p00194.

[24] CHEN, C.S., TU C.H., CHEN S.J., CHEN C.C., *Simulation of groundwater contaminant transport at a decommissioned landfill site. Case study, Tainan City, Taiwan*, Int. J. Environ. Res. Public Health, 2016, 13 (5), 467. DOI: 10.3390/ijerph13050467.

[25] LI S., SARI Y.A., KUMRAL M., *Optimization of mining. Mineral processing integration using unsupervised machine learning algorithms*, Nat. Res. Res., 2020, 29, 3035–3046. DOI: 10.1007/s11053-020-09628-0.