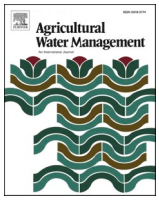
Agricultural Water Management 293 (2024) 108690

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat



Machine learning driven forecasts of agricultural water quality from rainfall ionic characteristics in Central Europe

Safwan Mohammed [a,b,*], Sana Arshad [c], Bashar Bashir [d], Attila Vad [b], Abdullah Alsalman [d], Endre Harsányi [a,b]

[a] Institute of Land Use, Engineering and Precision Farming Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 4032 Debrecen, Hungary
[b] Institutes for Agricultural Research and Educational Farm, University of Debrecen, Böszörményi 138, 4032 Debrecen, Hungary
[c] Department of Geography, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[d] Department of Civil Engineering, College of Engineering, King Saud University, P.O.Box 800, Riyadh 11421, Saudi Arabia

ARTICLE INFO

Handling Editor - J.E. Fernández

Keywords:
Rainwater chemistry
Sodium adsorption ratio
Multilayer perceptron
Agriculture water optimization
Hungary

ABSTRACT

Sodium hazard poses a critical threat to agricultural production globally and regionally which has been previously predicted from ground or surface water. Monitoring rainwater quality in this context is ignored but essential for agricultural water management in central Europe. Our study focused to predict sodium adsorption ratio (SAR) from 1985 to 2021 from ten ionic species of rainwater (pH, EC, Cl^-, SO_4^{-2}, NO_3^-, NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+}) employing four machine learning (random forest (RF), gaussian process regression (GU), random subspace (RSS), and artificial neural network-multilayer perceptron (ANN-MLP)) methods at three stations K-puszta (KP), Farkasfa (FAK), and Nyirjes (NYR) of Hungary, central Europe. Exploratory data analysis was performed using the Mann-Kendall test, Pearson correlation, and principal component analysis (PCA). Rainwater composition revealed the highest percentage of SO_4^{-2} ions i.e., 21 to 31%, followed by 10 to 15% of Na^+ ions. Mann-Kendall test revealed a significant ($p < 0.05$) increasing trend of Na^+ ions and SAR portraying it a serious hazard limiting agricultural production. Machine learning results from 10 model runs of all algorithms for SAR prediction at KP station proved the efficacy of ANN-MLP as superior with RMSE range of 0.02 to 0.05, followed by RF with RMSE of 0.14 to 0.19 in scenario 2 (SC-2) (Na^+, Mg^{2+}, Ca^{2+}). Validation of the best-selected algorithm (ANN-MLP) and scenario (SC-2) also predicted the SAR with a low RMSE of 0.08 and 0.05 at both FAK and NYR stations, respectively. Hence, the efficiency of ANN-MLP in forecasting SAR from rainwater proves it to be a meticulous tool for enhancing agricultural water management practices in Central Europe and enhancing resource efficiency and crop production in the future.

1. Introduction

Rainfall is a significant sinker of various atmospheric matters, and its chemistry reflects the microphysical properties of the atmosphere such as atmospheric pollutants and aerosol deposition (Ge et al., 2021; Keresztesi et al., 2020a). Several factors of climate change including GHGs emissions due to anthropogenic activities like land cover/land use changes, urbanization, fossil fuels burning, mining, and industrial processes release large amounts of air pollutants such as SOx, NOx, etc. into the atmosphere (Redington et al., 2009; Zeng et al., 2022). Rainfall scavenges the soluble atmospheric components, acts as a collector, and transports of these pollutants back towards the earth's surface (Das et al., 2005; Lü et al., 2017). Therefore, ionic chemistry of rainwater is linked to several terrestrial, atmospheric, and sea salt source (Xu et al., 2015).

Low pH acidic rains with high SO_4^{-2} and NO_3^- are reported in the Americas, Europe (Al-Momani et al., 1995; Alastuey et al., 1999), and parts of China (Facchini Cerqueira et al., 2014; Hontoria et al., 2003; Rodhe et al., 2002; Zeng et al., 2022; Zhou et al., 2019) transferring the pollutants load from atmosphere to terrestrial ecosystems.

The first global assessment of rainwater chemistry released by the World meteorological organization concluded that sulfate and nitrate

* Corresponding author at: Institute of Land Use, Engineering and Precision Farming Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 4032 Debrecen, Hungary.
E-mail address: safwan@agr.unideb.hu (S. Mohammed).

https://doi.org/10.1016/j.agwat.2024.108690
Received 30 August 2023; Received in revised form 1 January 2024; Accepted 12 January 2024
Available online 19 January 2024

emissions were prominent acidic wet and dry depositions in North America and Europe which significantly declined in the early 1990s (Whelpdale et al., 1997). Other than SO_4^{-2} and NO_3^-, rainwater chemistry also constitutes alkaline metallic elements namely, Na^+, Mg^{2+}, and Ca^{2+} linked with sea salt, crustal dust, and other natural and anthropogenic sources (Vet et al., 2014). For instance, Keresztesi et al. (2019) reported that in Europe the average (2000 – 2017) pH of rainwater was 4.80 associated with a high concentration of sulfates, and chloride and moderate concentration of neutral cations i.e., magnesium, calcium, and sodium, the significant ingredients of sodium absorption ratio (SAR). Numerous other water quality indicators such as permeability index (PI), percent sodium (%Na), Kelly ratio (KR) (Kushwaha et al., 2023), and Mg^{2+} / Ca^{2+} (Yuan et al., 2023) are also being used to assess the agricultural water quality. SAR is not only particular to address sodium hazard but also considerate the ratio of Ca^{2+} and Mg^{2+} ions making it the most significant indicator in agriculture management (Docheshmeh Gorgij et al., 2023; Sattari et al., 2020).

Overall, agricultural practices in central and northern Europe are supplemented by irrigation to optimize crop production specially in dry summers and water scarce months (Wriedt et al., 2009). However, rainwater interaction with irrigation enhances the vulnerability of sodicity hazard which plays a vital role in agricultural water management (Suarez et al., 2008). Soil rich in sodium disperses the clay particles and clog soil pores, reducing its permeability. Hence the altered soil structure lowers the hydraulic conductivity impeding the water movement through soil profile (Gautam et al., 2023).

Consequently, high SAR values reduce hydraulic conductivity, clay dispersion, soil crusting, aggregate stability, and irrigation effectiveness (Gharaibeh et al., 2021; Klopp and Daigh, 2020). Ultimately, SAR affected soil leads to poor crop performance and yield reduction (Minhas et al., 2019). Sodicity is a worldwide hazard in varied climatic conditions (Mohanavelu et al., 2021). For instance, Daliakopoulos et al. (2016) provided a review of salinity and sodicity in Europe and reported 3.5 – 5 million ha irrigation-induced soil degradation due to sodium hazard. (Suarez et al., 2006). Another study by Çankaya et al. (2023) also reported the reduced soil infiltration due to SAR hazard in irrigation. In the similar context, Birkás and Dekemati (2023) also highlighted the soil degradation as a hindering factor of crop production in a longer run. Another recent study by Koseoglu-Imer et al. (2023) also emphasized the importance of examining recent agricultural water challenges and creative technical solutions from future perspective in EU countries. Hence, in the assessment of agricultural water quality, the inclusion of rain hazards is quite essential especially when rain events occur throughout the crop season. For the purpose, statistical methods are variously adopted in ground, river, and rainwater quality research specifically for experimental data obtained from the field observations (Li et al., 2023). Statistical modeling includes meticulous analytical techniques like classical regressions, factor analysis, or hypothesis testing and validations which are complex and require substantial field observations (Khan et al., 2022). Further, such kind of modeling deals with linearly distributed data and is unable to detect the non-linear predictors of water quality indicators (Najah Ahmed et al., 2019). The adoption of artificial intelligence AI-based methods provides an effective alternate approach to dealing with several non-linear and complex relationships (Najah Ahmed et al., 2019). Studies have implemented several machine learning techniques like M5P, bagging, random forest (RF), multiple linear regression (MLR), support vector regression (SVR) (Nong et al., 2023; Nouraki et al., 2021; Sepahvand et al., 2021), artificial neural network (ANN) (Rahnama et al., 2020) for predicting SAR and other water quality indicators in multi environmental conditions. For instance, Mustafa et al. (2021) reviewed several ANN algorithms including feed-forward backpropagation (FFBP), radial basis function (RBF), cascade forward back propagation (CFBP), ensemble ANN, etc. to be useful for monitoring and predicting water quality indicators. In other hydrological applications, Lu and Ma (2020) used RF and extreme gradient boosting (XGBoost) to predict the polluted river water quality in the USA. Another study by Wagh et al. (2016) predicted groundwater quality from 13 physiochemical parameters using ANN methods for irrigation purposes. El Bilali et al. (2021) founded RF and Adaboost ML models to be more effective over support vector machine (SVM), and ANN of six irrigation water quality indicators including SAR.

A review of recent studies for predicting water quality indicators for irrigation water management presented in Table 1 shows utilization of robust ML algorithms and resulted high accuracy by ANN. Moreover, previous research also implies that most of the recent studies have been conducted on ground or surface water sample observations in developing countries. No recent research is found predicting water quality indicators like SAR from rainwater chemicals specifically in central Europe. In this context, our study aims to predict the SAR from 10 rain water quality parameters namely, pH, EC, Cl^-, SO_4^{-2}, NO_3^-, NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+} at three selected stations in Hungary, central Europe.

Our study effectively employed four ML algorithms namely random forest (RF), Gaussian regression (GR), random subset (RS), and multilayer perceptron (ANN-MLP) to find the best ML algorithm and predictors for agricultural water management.

2. Material and method

2.1. Study area

The study area for current research is located in central Europe between latitudes 45° 55N - 48° 60N and longitudes 16° 10E – 22° 50E, called Hungary. The country is dominated by lowlands i.e., the great Hungarian Plain in the east and hilly Carpathian range in the north (Harsányi et al., 2023). Central Danube River is dividing the eastern Hungarian plains from the Transdanubian hills in the west. Hydrographically, the Danube and Tisza are recognized drainage basins of the country (Pinke et al., 2020). The old 40 years meteorological record of the country reveals the mean annual temperature of the country to be 11 – 12 °C and 600 mm mean annual precipitation (Kern et al., 2018). The major land cover/ land use of the country includes forests in the northern and western parts, scattered shrubs, and grasslands along the hilly areas, and cultivated cropland area in the plains with wheat, maize, sunflower, and rapeseed as main cereal crops (Harsányi et al., 2023; Kern et al., 2018). Agriculture of the country is mainly dependent upon rainfall supplemented by irrigation in rain-scarce months of the year (Bussay et al., 2015; Pinke et al., 2020). The three selected stations for the current study include K-puszta (KP) (46° 58' N, 19° 33' E, altitude, 127 m), Farkasfa (FAK) (46° 54' N, 16° 18' E, altitude, 312 m), and Nyirjes (NYR) (47° 52′ N, 19° 57' E, altitude, 702 m) (Fig. 1). The KP station belong to the flat and plain lowlands with agriculture as a dominant land use whereas FAK and NYR belongs to western and northern highlands. The topography and land cover characteristics of the three stations represent the varied agroclimatic and regional environmental conditions (Fig. 1). Furthermore, the selected stations provide a rich, diversified, and reliable historical achieve of rainwater ionic composition from Hungarian Central Statistical Office (ksh.hu). Hence, it makes these stations good representatives at a regional level to robustly analyze and predict the SAR from rainwater ions using machine learning algorithms.

2.2. Data used (predictors and response variables)

Available annual rainwater characteristics for 10 parameters were collected online from the Hungarian Central Statistical Office (https://www.ksh.hu/stadat_files/kor/en/kor0055.html) for the three stations; KP (1985–2021), FAK (1997–2021), and NYR (1997–2021) across Hungary. The rainfall parameters (explanatory variables or predictors) used to forecast the response variable i.e., sodium hazard (SAR) includes: pH, EC (μS/cm), Cl^- (mg/l), SO_4^{-2}(mg/l), NO_3^- (mg/l), NH_4^+ (mg/l), Na^+ (mg/l), K^+ (mg/l), Mg^{+2} (mg/l), and Ca^{+2} (mg/l).

Table 1
A review of recent studies for irrigation water quality prediction using ML methods.

WQ indicator	Water source	Objective	Predictors	Region	ML-method	Outcome	Reference
SAR, ESP, %Na, RSC, PI, KR, MAR,	Surface water	Assessment of irrigation water quality	EC, pH, CO_3^{-2}, HCO_3^{-1}, Cl^-, NO_3, NH_4^+, K^+, Na^+, Ca^{+2}, Mg^{+2}	Morocco	MLR, ANN, SVR, kNN, DT SGD, AdaBoost	Except SVRm and kNN all performed well & AdaBoost outperformed	(El Bilali and Taleb, 2020)
SAR	River surface	Predict SAR from 15 WQ params for agriculture monitoring	Na^+, EC, SO_4^{-2}, TDS, Ca^{+2}, Cl^-, hardness, Mg^+, CO_3^{-2}, HCO_3^{-1}, T, PRCP, pH, B	Turkey	SVR, KNN, RF, Random tree, Hoeffiding tree, REPTree,	SVR with poly kernel provided the best SAR estimation	(Sattari et al., 2020)
TDS, potential salinity (PS), SAR, MAR, ESP	Groundwater	Predicting water quality for irrigation	T, pH, EC	Tunisia	RF, SVR, ANN, AdaBoost	AdaBoost followed by RF performed better for all WQ indicators	(Trabelsi and Bel Hadj Ali, 2022)
SAR, %Na, RSC, Mg hazard, PI, KR,	Groundwater	Assessment of irrigation water quality	K^+, Na^+, Ca^{+2}, Mg^{+2}, CO_3^{-2}, HCO_3^{-1}	India	LSTM, MLR, ANN,	ANN performed better in two scenarios	(Kouadri et al., 2022)
SAR, %NA, KR, RSC	Groundwater	Assessment of sodium hazards in irrigation water quality	K^+, Na^+, Ca^{+2}, Mg^{+2}, CO_3^{-2}, HCO_3^{-1}	India	ANN	ANN performed as better suited for all WQ indicators	(Gautam et al., 2023)
WQI, SAR, PI, KR, %Na,	Groundwater	Assessment of groundwater quality for irrigation	pH, EC, K^+, Na^+, Ca^{+2}, Mg^{2+}, CO_3^{-2}, HCO_3^{-1}, SO_4^{-2}, Cl^-,	India	Bagging, RF, REPTree, RSS, M5P, AdR, ANN,	ANN and M5P outperformed others in predicting WQ indicators	(Kushwaha et al., 2023)
SAR, PI, SSP, PS, KR, MAR,	Both surface & groundwater	Assessment of irrigation water quality	EC, pH, HCO_3^{-1} Cl^-, NO_3, Ca^{2+}, Mg^+, Na^+, K^+,	Nigeria	ANN-MLP, MLR,	ANN-MLP performed better	(Omeka, 2023)
SAR	Groundwater	Monitoring irrigation water quality	EC, K^+, Na^+, Ca^+, Mg^{2+}, CO_3^{-2}, HCO_3^{-1} Cl^-, SO_4^{-2}	Iran	LSTM	SAR for 2020 is forecasted from SAR input from 2002-2019 in the LSTM model	(Docheshmeh Gorgij et al., 2023)

2.3. Calculation of sodium adsorption ratio (SAR)

Several indicators of water quality assessment are used for irrigation purposes including TDS, SAR, MAR, KR, %Na, PS, WQI, etc. (Alsubih et al., 2022; Mokhtar et al., 2022; Wagh et al., 2016). Currently, we aim to predict and forecast the sodium hazard in rainwater which is better described as "Sodium adsorption ratio (SAR)" from the perspective of agricultural water monitoring. It is calculated by:

$$\boldsymbol{SAR} = \frac{\boldsymbol{Na^+}}{\sqrt{\frac{\boldsymbol{Mg^{2+}+Ca^{2+}}}{\boldsymbol{2}}}} \tag{1}$$

Sodium hazard directly relates to the salinity and sodicity in the soil and agricultural water with a SAR value greater than 10 reduces hydraulic conductivity affecting the permeability of the soil and is not recommended for plant growth keeping in count the Gypsum in the soil (Zaman et al., 2018).

2.4. Exploratory analysis (Man-Whitney, Man-Kendall trend, Pearson Correlation, Principal Component Analysis)

Before ML application for SAR prediction, all rainwater ionic species (predictors) and SAR (response variable) are explored to find the trend and interrelationship. First, the descriptive statistics (minimum, maximum, mean, standard deviation, skewness, Kurtosis) of all predictors and response variables were carried out. Then, the non-parametric Mann-Whitney U test is applied to examine the statistical difference in the distribution and median of independent groups (rainwater ions at three independent stations) with non-normal distribution (MacFarland and Yates, 2016).

Moreover, a monotonic trend of all variables for three stations is identified by employing Mann-Kendall (MK) trend analysis (Mann, 1945) and Sen's slope (SS) estimator (Sen, 1968). The MK is a non-parametric trend test that is widely used to identify the significance of trends for several hydro-meteorological variables (Güçlü, 2020). The output of MK test Kendall value (Tau) at 95 to 99% level of significance (*P < 0.05, **P < 0.01) for rejecting the null hypothesis (*H0*) stating no significant trend in time series in the opponent of alternate hypothesis (*H1*) stating a significant trend exists in time series (Kendall, 1948). It is presented by:

$$\boldsymbol{S} = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \boldsymbol{singn}(\boldsymbol{x_j} - \boldsymbol{x_k}) \tag{2}$$

Afterward, the slope of the trend is identified by Sen's slope estimator with negative slope values revealing decreasing and positive slope values revealing an increasing trend of time series presented as

$$\boldsymbol{q_i} = \frac{\boldsymbol{x_j} - \boldsymbol{x_k}}{\boldsymbol{j} - \boldsymbol{k}} \tag{3}$$

Further to identify the breakpoints in the test, another parametric Buishand range (BR) test is applied to the time series of all variables (Buishand, 1982). Rejecting the null hypothesis, it assumes that a step-wise change or break in the mean is present in the time series with the value K (the year or month) as trend changing point (Costa and Soares, 2009). Significant shift $\boldsymbol{R}$ in time series is computed from the normalization of rescaling adjusted partial sums ($\boldsymbol{S_k}$) (Costa and Soares, 2009).

Furthermore, Pearson correlation is a widely applied method to explore the significant (*P < 0.05, **P < 0.01) negative and positive linear relationship between all ionic species of rainwater and SAR presented in Eq. 4 (Keresztesi et al., 2020b; Nasiruddin Khan and Sarwar, 2014; Vlastos et al., 2019).

$$\boldsymbol{r} = \frac{\sum_{i=1}^{n}(\boldsymbol{x_j} - \boldsymbol{x})(\boldsymbol{y_j} - \bar{\boldsymbol{y}})}{\sqrt{\sum(\boldsymbol{x_j} - \boldsymbol{x})^2 \sum(\boldsymbol{y_j} - \bar{\boldsymbol{y}})^2}} \tag{4}$$

Another significantly applied multivariate statistical technique to examine the qualitative relationships between rainwater ionic species is Principal Component Analysis (PCA) (Cao et al., 2009). PCA is an unsupervised dimensionality reduction technique which simplifies the

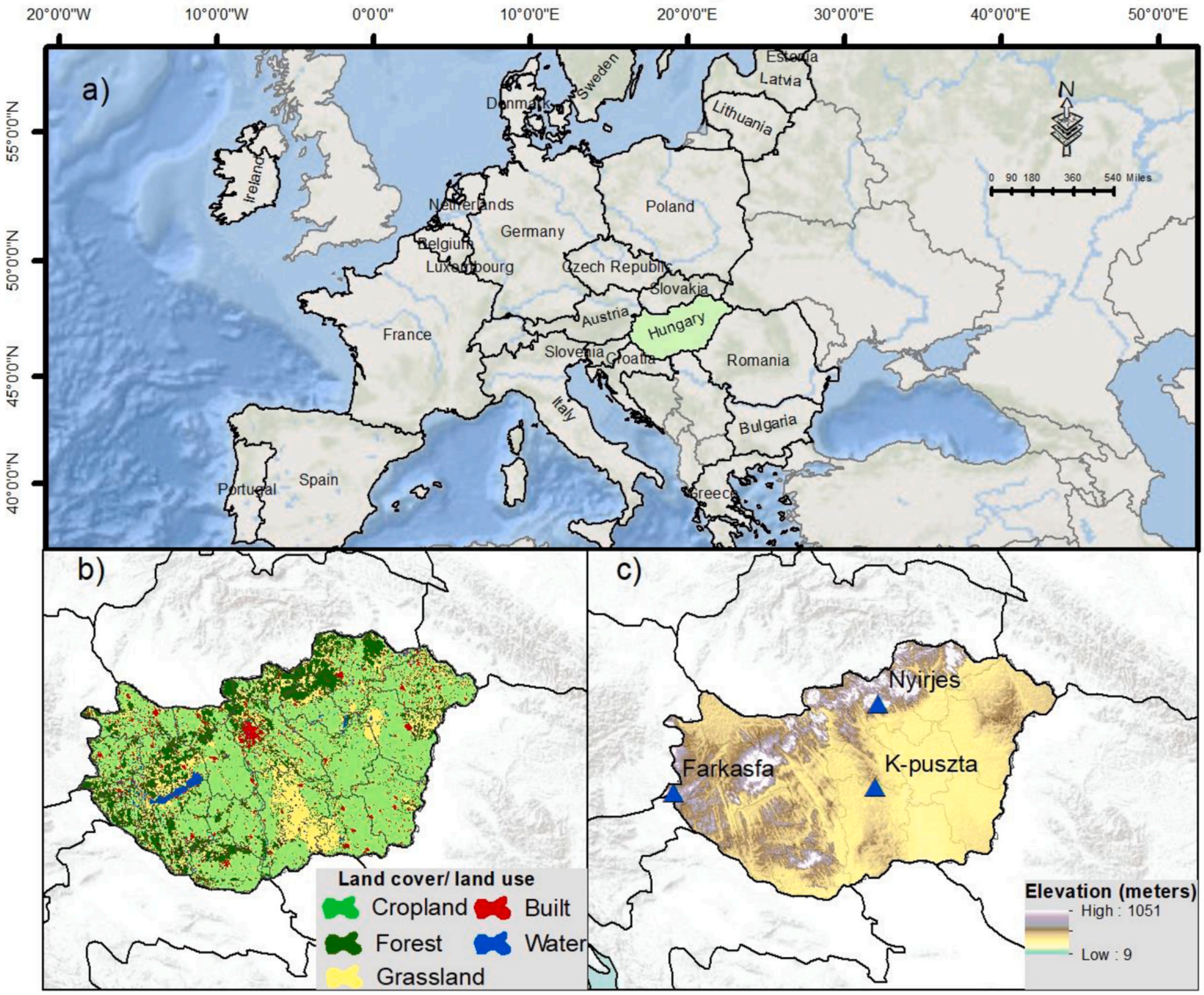


Fig. 1. a) Location of the study area (Hungary) in central Europe b) Land cover/land use of Hungary acquired from MODIS/061/MCD12Q1c) Elevation of Hungary in meters acquired from USGS/SRTMGL1_003.

complex datasets identifying important variations and patterns within the data (Kouadri et al., 2022). Currently, varimax rotation method is adopted to identify the loading factors in two major Principal components explaining the maximum variability of rainwater chemicals presented in loading biplot.

2.5. Machine learning models

2.5.1. Random subspace (RSS)

The random subspace is a decision tree-based and ensemble machine learning method which works on the principle of a random subset of feature selection in a high-dimensional feature space (Lai et al., 2006; Tin Kam, 1998). RSS is a common random selection approach well adopted in small training samples and the data is highly dimensional. Hence, it enhances or improves the performance capability of "weak classifiers" (Saha et al., 2023). Introducing randomness in feature selection helps to achieve good accuracy at the training stage with further improvement as it grows complex. The algorithm consists of systematically constructing multiple trees by pseudorandom selection of subsets in feature space (Skurichina and Duin, 2002). The details of currently adopted parameters of random subspace are presented in Table 2.

Table 2
Parameters selection for adopted machine learning models.

Algorithm	Parameters
Random subspace (RSS)	weka.classifiers.meta. Random SubSpace, classifier: REPTree, iterations = 10 subspace size = 0.5, seed = 5,10,15,….45.
Random Forest (RF)	weka.classifiers.trees.RandomForest, ntrees I = 100, batch-size = 100, computevarimp = TRUE, max tree depth = 0, variance V = 0.001, seed S = 5,10,15…45,
Gaussian process regression (GPR)	weka.classifiers.functions.GaussianProcesses, batch size = 100, kernel = polynomial, do not check capabilities = FALSE, seed S = 5, 10, 15, …….45.
Artificial Neural Network-Multi-layer Perceptron (ANN-MLP)	Batch size = 100, Learning rate L = 0.3, momentum M = 0.2, Activation function = sigmoid, hidden layers = a, Num of epochs to train = 500, Regularization = weight decay, seed S = 5,10, 15,…45.

2.5.2. Random forest (RF)

The random forest is another supervised and non-parametric ensemble ML tree-based model based on several decision trees. Individual trees are built from bootstrap subsample selection of training data where each subsample develops a decision tree to give a prediction of the response variable (Breiman, 1996, 2001). Finally, the ensemble or average of all decision trees is computed to take one final prediction (Strobl et al., 2008). Random forest is widely applied in water quality prediction research (Alnahit et al., 2022; Avila et al., 2018; Chen et al., 2020). The detailed parameters of the applied RF model in our research are presented in Table 2.

2.5.3. Gaussian process regression (GU)

The gaussian process regression (GU) belongs to the Bayesian non-linear, non-parametric ML algorithm for high-dimensional space problems and is based on Gaussian probability distribution (Meng et al., 2021). It is a kind of supervised or semi-supervised learning method to solve various probabilistic classification or regression problems (Liu et al., 2021; Wang et al., 2021). The Gaussian process capability to deal with noise and missing data with small-sample problems makes it a good choice for other supervised ML methods. GPR in varied hydrological research proves its applicability to achieve more accurate predictions in complex environments (Liu et al., 2022; Shadrin et al., 2021; Wan et al., 2022; Zare Farjoudi and Alizadeh, 2021). The detailed parameter selection adopted for GPR in our research is presented in Table 2.

2.5.4. Artificial neural network multi-layer perceptron (ANN-MLP)

Feedforward multilayer perceptron (MLP) is a well-established artificial neural network (ANN) architecture based on a backpropagation (BP) algorithm to solve non-linear variable relationships (Lek et al., 1996). The basic architecture of MLP is described by three layers: an input layer, one or more hidden and an output layer which is interconnected by weighted nodes or neurons. The BP algorithm adjusts the weights between neurons to minimize the output predicted error. In this case, only one hidden layer is used, and nodes are all sigmoid as it provides satisfactory results. Overall, neurons in hidden layers are user-defined to achieve accuracy in results while the output layer consists of only one neuron corresponding to the predicted value (Hornik et al., 1989). The weighted sum of input Net_j is computed from $W_{i,j}$, the weight between ith and jth neuron, the output is created using sigmoid function (Najah et al., 2013).

$$Net_j = \sum_{i=1}^{I} W_{i,j} + \theta_j \tag{4}$$

The detailed parameters of the currently applied MLP algorithm are provided in Table 2.

2.6. Machine learning methodology

Currently, SAR is predicted from ten rainwater ionic species in two developed scenarios (Table 3). To implement the ML algorithms at three stations, KP is chosen as a standard station to implement the four selected ML algorithms for predicting SAR in two scenarios i.e., SC-1, and SC-2. The whole dataset of KP station is randomly split into 70% training and 30% testing. Also, the k (10)-fold was used for cross-validation (CV) approach. To achieve a higher accuracy, each ML model was run 10 times using a seed selection approach from S = 1, 5, 10, 10, 15 …. 45 for each scenario at the training, testing, and CV stage getting in total 120 model outputs. On each run, all ML models were evaluated based on root mean squared error (RMSE) and mean absolute error (MAE). Further, to select the best model performance (SC and ML-algorithm), the RMSE and MAE are analyzed, and the models with the lowest RMSE from all algorithms and scenarios were chosen as the best and most effective in predicting SAR. To verify the performance of the selected ML- algorithm/ SC with seed value is further implemented to predict SAR on the remaining two (FAK, and NYR) stations. The whole ML methods were implemented in the WEKA environment already proven in various data mining and ML research (Ćalasan et al., 2020; Elbeltagi et al., 2023; Harsányi et al., 2023; Yadav et al., 2014).

Table 3
Applied Scenarios with input attributes in ML algorithms.

SC	ML algorithms	Input attributes	output
SC-1	1. ANN-MLP 2. RF 3. RSS 4. GU	pH, EC, Cl^-, SO_4^{-2}, NO_3^-, NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+}	SAR
SC-2	1. ANN-MLP 2. RF 3. RSS 4. GU	Na^+, Mg^{2+}, Ca^{2+}	SAR

2.7. Performance evaluation

The performance of ML algorithms is evaluated by employing root mean square error (RMSE) and mean absolute error (MAE), which are used in various ML studies (Ahaninjan and Egdernezhad, 2020; Alnahit et al., 2022; Barzegar et al., 2020). Currently, the RMSE refers to the square root of the average of squared distances between actual and predicted SAR values. MAE computes the error difference between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (SAR_{act} - \widehat{SAR}_{pred})^2} \tag{5}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |SAR_{Prd} - SAR_{act}| \tag{6}$$

3. Results

3.1. Exploratory data analysis of rainfall characteristics at three stations (Descriptive statistics and trend analysis)

The difference between the minimum and maximum value of rainwater ions reveals a good range specifically in EC, SO_4^{-2}, NH_4^+, Na^+, Ca^{2+}, and SAR. The skewness ranges from −0.7 to 1.5 at KP, −0.9 to 3.69 at FAK, and −1 to +1 at NYR. Similarly, kurtosis range from −1.6 to 2.6 at KP, −1 to 14.9 at FAK, and −0.6 to 1.29 at NYR station, demonstrating a non-normal data distribution of ionic rainwater species (Table 4). The further exploration of the data employing the Mann-Whitney U test demonstrates a significant statistical difference in the median for pH, Cl^-, NH_4^+, Na^+, K^+, and Ca^{2+}(Fig. 2). Moreover, the average composition of rainwater ions for the time series of all stations (Fig. 2) exhibit SO_4^{-2} with the highest 31.6%, 21.6%, and 29.1% at three (KP, FAK, and NYR) stations followed by NO_3^- as 22.6%, 20.9%, and 22.8% respectively. Cl^- is another important anion with the highest 24.2% at FAK, 12.2% at NYR, and 10.1% at KP station. The third abundant component is Na^+ ions with 10.4% at KP, 13.8% at FAK, and 14.6% at NYR station. Other than these, K^+, and Mg^{2+} are the least abundantly found cations in the long-term rainwater composition at three stations (Fig. 2). The sequence of ionic species for the three stations follows as $SO_4^{-2} > NO_3^- > Ca^{2+} > Na^+ > Cl^- > NH_4^+ > K^+ > Mg^{2+}$ (KP), $Cl^- > SO_4^{-2} > NO_3^- > Na^+ > Ca^{2+} > NH_4^+ > K^+ > Mg^{2+}$ (FAK), and $SO_4^{-2} > NO_3^- > Na^+ > Cl^- > Ca^{2+} > NH_4^+ > K^+ > Mg^{2+}$ (NYR).

Monotonic trend analysis through Mann-Kendall and Sen's slope test applied to all ionic elements of rainfall at selected stations provided meaningful results (Table 5). pH of rainwater at KP station showed a highly significant ($P < 0.001$) increasing trend with tau = 0.4 and SS = 0.02 over the time of 37 years (1985–2021). Contrary, the remaining two stations FAK and NYR showed a decreasing trend with negative Kendall and SS values. EC of rainwater revealed a significant

Table 4
Descriptive statistics of rainwater ions and SAR.

Station	Stats/var	pH	EC	Cl^-	$SO4^{-2}$	NO_3^-	NH_4^+	Na^+	K^+	Mg^{2+}	Ca^{2+}	SAR
KP	Min	4.6	14.1	0.42	1.17	1.12	0.31	0.28	0.09	0.07	0.30	0.22
	Max	6.06	33.2	1.37	7.55	2.90	1.29	1.53	0.61	0.60	3.22	2.58
	Mean	5.5	22.1	0.92	2.89	2.07	0.75	0.94	0.23	0.22	1.10	1.27
	Std.Dev	0.45	5.69	0.22	1.43	0.49	0.22	0.34	0.10	0.12	0.69	0.67
	Skewness	-0.7	0.48	-0.07	1.27	-0.3	0.22	-0.1	1.4	0.86	1.5	0.08
	Kurtosis	-0.87	-1.04	-0.51	1.81	-0.78	-0.22	-1.25	2.66	0.45	2.11	-1.60
FAK	Min	5.11	14.18	**0.62**	1.22	1.28	0.41	0.49	0.15	0.09	0.42	0.55
	Max	6.21	44.76	**19.2**	4.20	3.67	1.78	2.27	0.90	2.27	1.42	3.32
	Mean	5.81	22.2	2.47	2.20	2.13	0.62	1.40	0.31	0.31	0.73	1.98
	Std.Dev	0.23	8.02	4.16	0.71	0.63	0.30	0.40	0.18	0.47	0.28	0.80
	Skewness	-0.9	1.86	3.6	0.91	0.85	2.92	-0.51	1.86	3.69	1.12	-0.4
	Kurtosis	2.7	3.3	13.7	1.4	0.2	9.6	1.0	3.6	14.9	0.7	-1.1
NYR	Min	5.05	14.5	0.4	1.34	1.11	0.34	0.43	0.14	0.08	0.11	0.49
	Max	5.9	26.1	1.49	4.10	2.80	0.82	1.74	0.37	0.49	1.72	3.29
	Mean	5.6	19.4	1.0	2.3	1.87	0.52	1.19	0.24	0.20	0.77	1.6
	Std.Dev	0.22	3.10	0.27	0.66	0.41	0.12	0.36	0.06	0.10	0.37	0.67
	Skewness	-1.1	0.45	0.05	0.61	0.48	0.28	-0.96	0.58	1.06	1	-0.12
	Kurtosis	0.86	-0.37	-0.38	0.37	0.18	-0.53	-0.06	-0.64	0.90	1.29	0.05

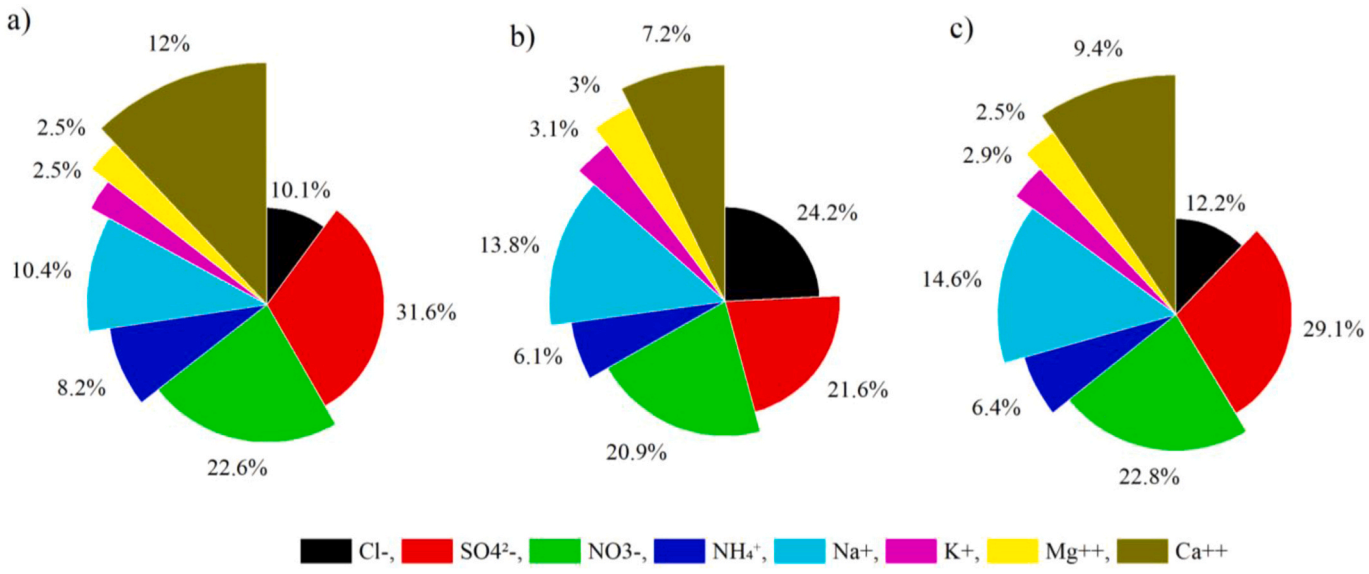


Fig. 2. average composition (%) of rainwater ions during the study period at three stations a) KP b) FAK c) NYR.

Table 5
Man-Kendall trend and Sen's slope estimator and Buishand range test for all ionic species at three stations.

Stations	K-puszta (KP)			Farkasfa (FAK)			Nyirjes (NYR)		
WQ / Trend	Tau	Sen's Slope SS	BR	Tau	Sen's Slope SS	BR	Tau	Sen's Slope SS	BR
pH	0.41***	0.02	1995	-0.01	-0.002	1997	-0.3*	-0.01	2008
EC	-0.63***	- 0.41	2003	-0.72***	-0.67	2008	-0.46***	-0.26	2009
Cl^-	0.23*	0.008	1996	0.03	0.002	2010	0.37**	0.02	2006
SO_4^{-2}	-0.56***	-0.08	2003	-0.65***	-0.08	2006	-0.48***	-0.05	2004
NO_3^-	-0.58***	-0.03	2009	-0.54***	-0.06	2006	-0.13	-0.01	2009
NH_4^+	-0.42**	-0.01	2000	-0.33*	-0.01	2003	0.08	0.002	2002
Na^+	**0.50*** **	**0.02**	**2001**	0.10	0.007	**2002**	**0.29***	**0.02**	**2002**
K^+	-0.55***	-0.006	1996	-0.47**	-0.01	2006	0.12	0.001	2003
Mg^{2+}	-0.63***	-0.007	2004	-0.67***	-0.01	2003	-0.67***	-0.009	2010
Ca^{2+}	-0.67***	-0.03	2003	-0.58***	-0.02	2003	-0.58***	-0.02	2008
SAR	**0.74*** **	**0.05**	**2002**	**0.46****	**0.08**	**2003**	**0.67*** **	**0.07**	**2005**

**P < 0.001
**P < 0.01
*P < 0.05,

($p < 0.001$) decreasing trend with a tau range from 0.4 to 0.7 and SS = (−0.26) – (−0.67) at NYR, KP, and FAK stations.

The other anions i.e., SO_4^{-2} and NO_3^- also revealed a significant ($p < 0.001$) decreasing trend with SS = −0.05 to −0.08, −0.01 to −0.06. The three important cations relating to sodium hazard namely K^+, Mg^{2+}, and Ca^{2+} also revealed a significant ($P < 0.001$, $p < 0.01$) decreasing trend at three stations with negative Sen's slope. The Na^+ and SAR revealed a significant ($p < 0.05$) increasing trend at three stations with a positive SS range from 0.02 to 0.07. Trend prediction of all rainwater ions at three stations is also supported by Fig. 3 clearly showing an average increasing trend of pH and Na^+, decreasing trend of EC, SO_4^{-2}, NO_3^-, K^+, Mg^{2+}, and Ca^{2+}, and no significant trend for Cl^- ions. The year 2010 identified exceptionally high concentration of Cl^- ions.

Furthermore, the Buishand range (BR) trend-changing detection test identified the breaking points in the trend as the K value equivalates to the specific year in the time series (Table 5 and Fig. 3). For example, the year 1995 is found to be trend-breaking for the rainwater pH at KP station which means higher pH after 1995 caused an increasing trend over here. It is also evidently visible in Fig. 3 revealing a sharp rise of pH after the year 1995. For sodium hazard i.e., Na^+ years 2001 and 2002 are identified as significant ($p < 0.001$) trend-increasing breakpoints at three stations also evident from Fig. 3 and 2002, 2003, and 2005 are significantly ($p < 0.001$) trend-increasing breakpoints of SAR in individual time series (Table 5 and Fig. 3).

3.2. Temporal Evolution of SAR as an indicator of water quality across three stations

Currently, SAR is computed as a significant parameter of water quality for agricultural monitoring in Hungary. A large proportion of agricultural water in Hungary is dependent on rainfall, therefore assessment and prediction of SAR regarding other ionic species is quite essential. The temporal progression of SAR at a seven-point scale over the historical time of 37 years (1985–2021) at KP station presented in Fig. 4 shows no significant rise in SAR from 1985–2002 and started increasing after 2003 till 2021. The highest SAR value of 2.5 is identified in the year 2006 followed by 2.05 in the year 2017. Like KP station, the other two stations also exhibited high SAR values from 2004 – 2019 with the highest SAR value of 3.3 in the year 2008 at FAK and 3.2 in the year 2019 at NYR station (Fig. 4). Overall, the temporal evolution of SAR from rainwater at three stations reveals to rise in the past decade which might be attributed to an increase in anthropogenic activities.

3.3. Relationship between rainwater ions and SAR (Pearson Correlation and Principal Component Analysis)

Pearson correlation between ionic species and SAR from rainwater at KP station demonstrates a significant positive correlation ($r = 0.68$, 0.81) of EC with two anions (SO_4^{-2}, NO_3^-) and $r = 0.74$, 0.63, 0.85, 0.87 with cations (NH_4^+, K^+, Mg^{2+}, Ca^{2+}) and significant negative correlation $r = -0.5$ and − 0.7 with Na^+ and SAR. A group of ions SO4–2^{2-}, NO_3^-, NH_4^+, have significant positive correlation range $r = 0.5$ – 0.7 with K^+,

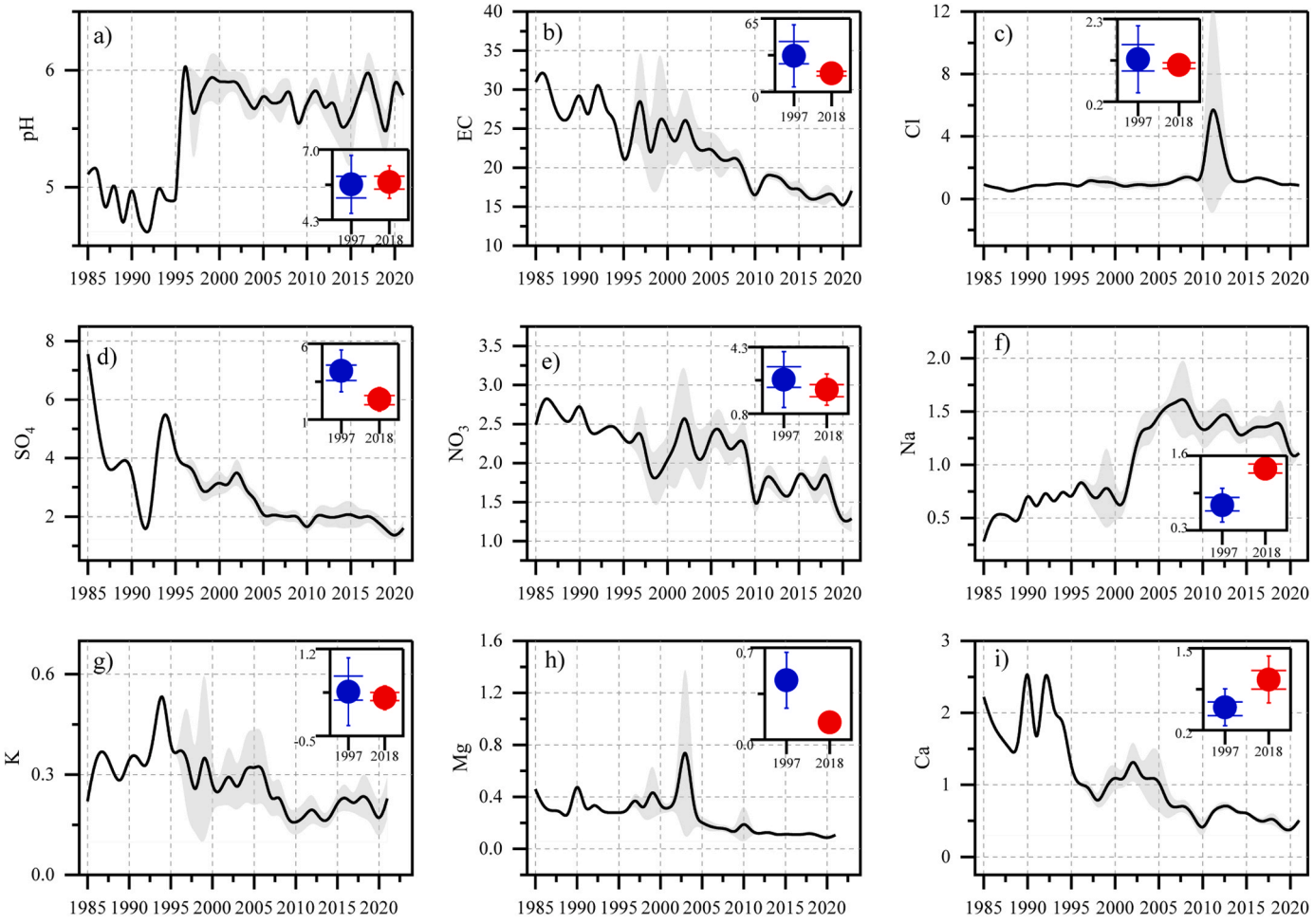


Fig. 3. Temporal evolution of rainwater ionic species across Hungary based on the average data for the three (KP, FAK, and NYR) stations: black lines represent the average of the three stations, gray shadow refers to the standard deviation between stations; I shaped box plot for the year 1997 in blue and 2018 in red.

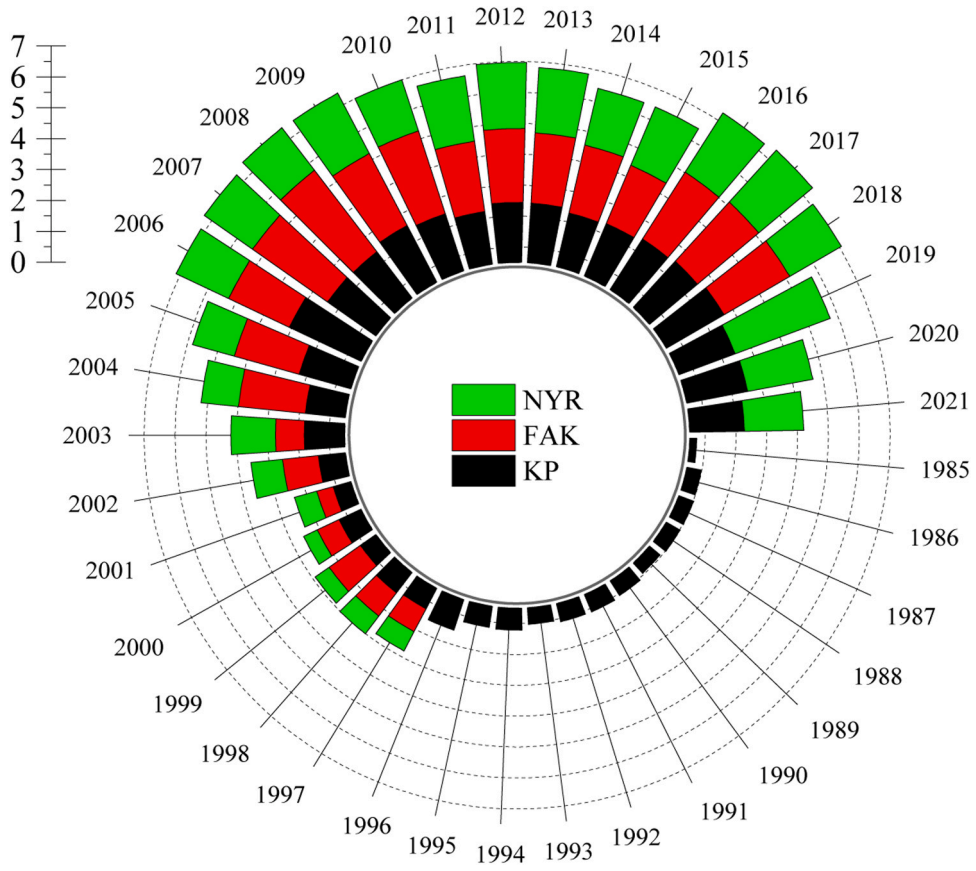


Fig. 4. Temporal evolution of SAR at the three stations (KP, FAK, and NYR) between 1985 and 2021.

Mg^{2+}, Ca^{2+}. Sodium ions Na^{+} is found to have a positive correlation with pH ($r = 0.57$), Cl^{-} ($r = 0.32$), and SAR ($r = 0.92$) and a negative correlation of Na^{+} ($r = 0.4 – 0.6$) with K^{+}, Mg^{2+}, and Ca^{2+} is observed. Positive correlation $r = 0.5$ and 0.6 of K^{+} with Mg^{2+} and Ca^{2+} and $r = 0.84$ between Mg^{2+} and Ca^{2+} is observed (Fig. 5). Moreover, with few differences, similar relationships are observed between rainfall ionic species at the remaining stations (FAK, and NYR). For example, a low and non-significant correlation of pH is identified with EC ($r = -0.22$), and SAR. However, a similar positive correlation cluster of EC with SO_4^{-2}, NO_3^{-} and NH_4^{+} ($r = 0.7$). Similarly, a negative correlation cluster of SO_4^{-2}, NO_3^{-} and NH_4^{+} is reported with Na^{+} ($r = (-0.3) – (-0.6)$ and SAR ($r = -0.6$). Furthermore, NYR station report somehow weak relationships between cations and anions of rainwater chemicals, but Na^{+} and SAR exhibit a significant negative correlation ($r = -0.6$) with SO_4^{-2}, Mg^{2+} ($r = -0.43, -0.6$). Further, SAR is identified to have a significant negative correlation with pH and EC ($r = -0.4$) (Fig. 5).

Clusters of principal components of rainwater ionic species at three stations with confidence ellipse at 95% level of confidence presented in Fig. 6**.** Overlapping ellipses with closed parameter values reveal that there is not a high spatial statistical difference of principal components of rainwater ions, suggesting similar composition and pattern at three stations (Fig. 6**)**. Moreover, the loading biplot of principal components demonstrates that PC1 accounts for 46.9% of variance followed by 14.2% by PC2 scattered in distinct vector groups. In this context, SO_4^{-2}, NH_4^{+}, and Ca^{2+} are more closely related to each other and have a stronger positive influence on PCI while Na^{+} and SAR, pH, and Cl^{-} are closely related to each other and have negative loadings on PC1 and positive loadings on PC2. The cluster demonstrates a significant correlation between SAR with Na^{+} as also identified in Pearson Correlation. Moreover, another group of interrelated ionic species includes K^{+}, NO_3^{-}, and EC with a stronger positive influence on PC2.

Overall, PCA analysis showed a close association between all variables influencing the SAR in two PCs. Pearson correlation also reveals a good deal of significant positive or negative relationship of all rainwater ions with SAR. Hence, based on the theoretical background and empirical evidence of significant correlation, no predictor is extracted or removed in ML input through any feature selection method. All variables are employed for predicting SAR in rainwater at selected stations in two scenarios (Fig. 6). For the SC-1, a broader set of all rainwater ions (pH, EC, Cl^{-}, SO_4^{-2}, NO_3^{-}, NH_4^{+}, Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}) are chosen to explore their potential contributions for predicting SAR followed by specific highly significant Na^{+}, Mg^{2+}, and Ca^{2+} ions in the SC-2.

3.4. Machine learning performance for SAR prediction at KP station

The performance evaluation based on RMSE and MAE of four machine learning algorithms (ANN-MLP, GU, RF, and RSS) for two scenarios SC-1 and SC-2 at the training and testing stage is presented in Fig. 7. The boxplots of performance evaluation present the range of RMSE and MAE for 10 repetitive model runs for each ML algorithm.

At the training stage of SC-1, performance evaluations of 10 model runs for each algorithm revealed ANN-MLP with the lowest RMSE ranges between 0.02 to 0.14 and mean RMSE = 0.05, followed by RF with RMSE range between 0.09 to 0.11 and mean RMSE = 0.10, and GU with a constant RMSE of 0.18 in all models runs. RSS was performed with a wider range of RMSE between 0.17 to 0.26 with mean RMSE = 0.16. Like RMSE, MAE also showed ANN-MLP with the lowest range between 0.01 to 0.11 followed by RF with MAE between 0.06 to 0.07, and RSS MAE range between 0.1 to 0.2. Overall, performance evaluation of ML algorithms at the training stage demonstrates that change in seed value altered the model performance of ANN and RSS. But ANN-MLP showed the lowest RMSE and MAE in all runs (Fig. 7a). Further, SC-2 of the training stage revealed a closer RMSE range of all models runs.

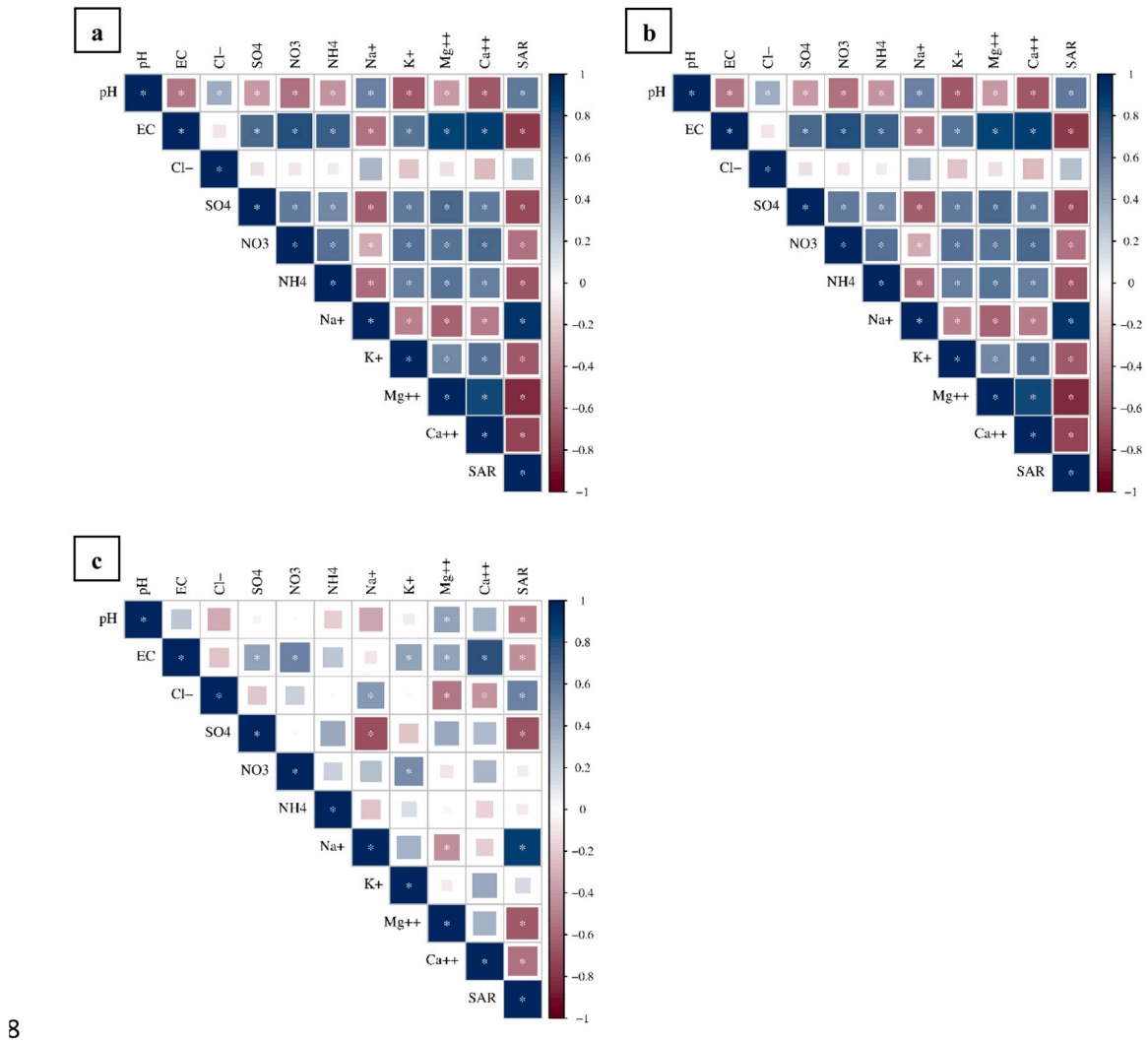


Fig. 5. Pearson correlation matrix of rainwater cations and anions for three stations a) KP b) FAK c) NYR (*P < 0.05).

For example, for the high performer ANN-MLP, the RMSE range is between 0.02 to 0.04 followed by RF with RMSE range = 0.07 to 0.09. The lowest mean RMSE = 0.03 in SC-2 by ANN-MLP provides a clear opinion about the better performance of SC-2 variables (Na, Ca, Mg) in predicting SAR. Overall, at the training stage ML model performance can be sequenced as ANN-MLP > RF > GU > RSS (Fig. 7a).

At testing stage SC-1, performance evaluation of all models runs also revealed ANN-MLP with the lowest RMSE range between 0.1 to 0.3, followed by GU with RMSE = 0.18, RF with RMSE range = 0.15 to 0.23, RSS with RMSE = 0.13 to 0.34. At the testing stage, the performance of ANN-MLP is extraordinary in predicting SAR in SC-2 with a narrow RMSE range between 0.02 to 0.05, followed by RF with RMSE = 0.14 to 0.19, RSS (RMSE = 0.14 to 0.27), GU (RMSE = 0.25). The results revealed that at the testing stage, ANN-MLP performance is higher than all other algorithms in SC-2 with the lowest RMSE = 0.02 at seed S = 10 (3rd model run). Overall, at the testing stage ML model performance is sequenced as ANN-MLP > GU > RF > RSS (SC-1) and ANN-MLP > RF > RSS > GU (SC-2). Hence ANN-MLP is proven to be the best algorithm in SC-2 at both training and testing stages and chosen for validation of SAR prediction at two other (FAK, and NYR) stations at a selected seed = 10 (Fig. 7b).

3.5. Validation of best ML algorithm and Scenario at FAK and NYR stations

The selected ANN-MLP and SC-2 are validated for predicting SAR at FAK and NYR stations. Evaluation metrics clearly show a high performance of the algorithm with RMSE = 0.08 and 0.05 and MAE = 0.06 and 0.04 at both stations, respectively. Moreover, the high correlation coefficient $r = 0.9$ reveals highly accurate forecasting of SAR with very little over and underfitting (Fig. 8). The tylor diagram (Fig. 9) further proves a closed correlation between actual and predicted values of SAR validated by ANN-MLP at FAK and NYR stations.

4. Discussion

4.1. Causes and relationship of ionic species in rainwater chemistry

The chemical properties of rainwater are region or site-specific and are mainly related to the geological or physical structure and anthropogenic activities of the region (Nasiruddin Khan and Sarwar, 2014; Singh et al., 2007; Vlastos et al., 2019). Several ionic species are researched to determine the chemical properties of rainwater for agricultural purposes (Wu et al., 2012). The normal pH range of rainwater range between 5 to 5.5 and 5.6 is considered to be a threshold to examine its acidity (Keresztesi et al., 2020a). The findings from our study revealed that the maximum pH for the standard KP station is found

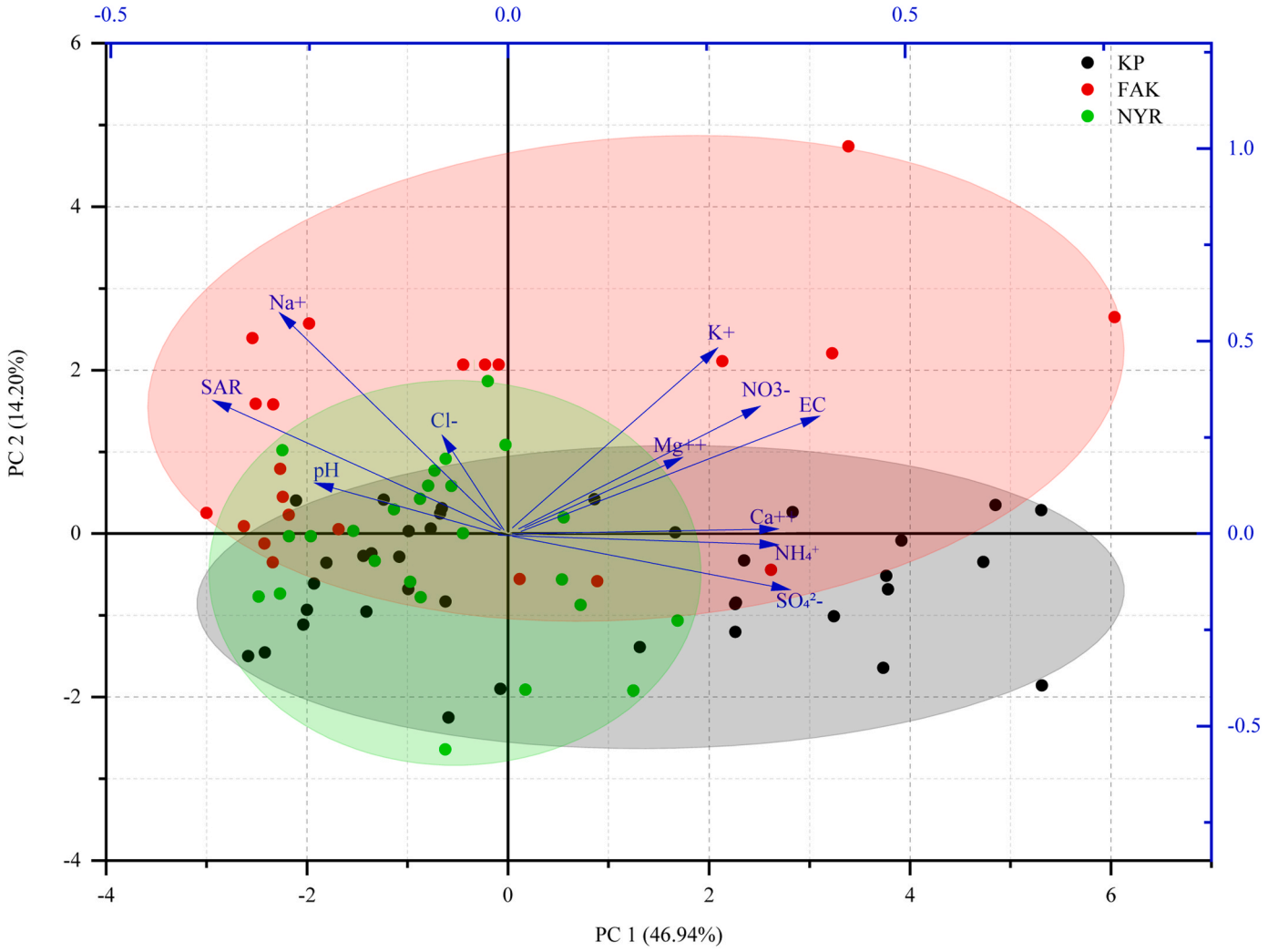


Fig. 6. Principal component analysis of rainwater ionic species in KP (black square), FAK (red square), and NYR (green square) with a 95% level of confidence ellipse.

to be 6.06 in 2017 and the minimum was 4.6 in year 1992 with a mean of 5.5 over 37 years (Table 2). Hence, a good range of pH at KP station provides significant opinions about the presence of high and low concentration of several ionic species in rainwater in different years of timescale (Fig. 2). However, several origination factors for the occurrence and interrelationships of ionic species are reported in the literature (Facchini Cerqueira et al., 2014; Ge et al., 2021; Zhou et al., 2019). For example, Celle-Jeanton et al. (2009) reported that SO_4^{-2}, NO_3^-, and NH_4^+ ions in rainwater can have both marine and anthropogenic origins like fuel combustion, traffic, or emissions from agricultural land. Similarly, rainwater-neutralizing alkaline species like HCO_3^-, Ca^{2+}, Mg^+, and K^+ have a terrestrial origin (Keresztesi et al., 2020b; Xiao et al., 2013). A strong positive correlation ($r > 0.5$) of SO_4^{-2} with NO_3^- and NH_4^+ (Fig. 5) reveals a similar origin might be attributed to oceanic salt and anthropogenic activities (Keresztesi et al., 2019; Keresztesi et al., 2020a) but with a declining trend (Table 5) as also reported by Chang et al. (2022). The findings of PCA in our study also provide an opinion about the high loading of SO_4^{-2}, and NH_4^+ on PC1 (Fig. 6) might be linked to anthropogenic sources supported by the findings of Cao et al. (2009). Similarly, the positive correlation between Na^+ and Cl^- ions (Fig. 5) in our study can be aligned with the findings of Keresztesi et al. (2020b) might be attributed to marine sources and sea salt spray from the Atlantic Ocean. Further, a positive correlation between K^+, Mg^+, and Ca^{2+} (Fig. 5) links the sources from the same geological structure of crustal features in the country as explained by Li et al. (2019). Positive attribution of Mg^+ with Ca^{2+} (Fig. 5) with a declining trend (Table 5) over time in rainwater is also linked to the sedimentary deposition of limestone and dolomite in the region (Keresztesi et al., 2019). Furthermore, the increasing trend of Na^+ ions (Table 5) with a significant i.e., 10 – 14% of concentration (Fig. 2) at selected stations of Hungary can be linked with several factors. For example, sea salt transportation from the Atlantic Ocean but with less chance due to its landlocked position. Further, a negative correlation of Na^+ with Mg^+ and Ca^{2+} (Fig. 5) provides less opinion or evidence about the geological origin of Na^+ in rainwater. Multivariate PCA revealed a significant share of Na^+ and Cl^- ions linked to both marine and anthropogenic origin, explaining 61% of the variance on PC1 and PC2 (Fig. 6) as also reported by the findings of Celle-Jeanton et al. (2009). Several anthropogenic factors like industrial, and urban emissions, intensive agricultural practices (fertilizers applications) might be the possible cause of increasing Na hazard in Hungary which needs to be researched more.

4.2. Sodium hazard (SAR) valuation for agricultural water monitoring

Water from rain contributes a significant footprint of agricultural water in central Europe and is supplemented by irrigation in the absence or in less rainfall months. Despite the direct contribution of rainwater to agriculture production, irrigation also depends upon surface and groundwater which is continuously recharged through rainwater (Bussay et al., 2015; Pinke et al., 2020). Hence, monitoring agricultural

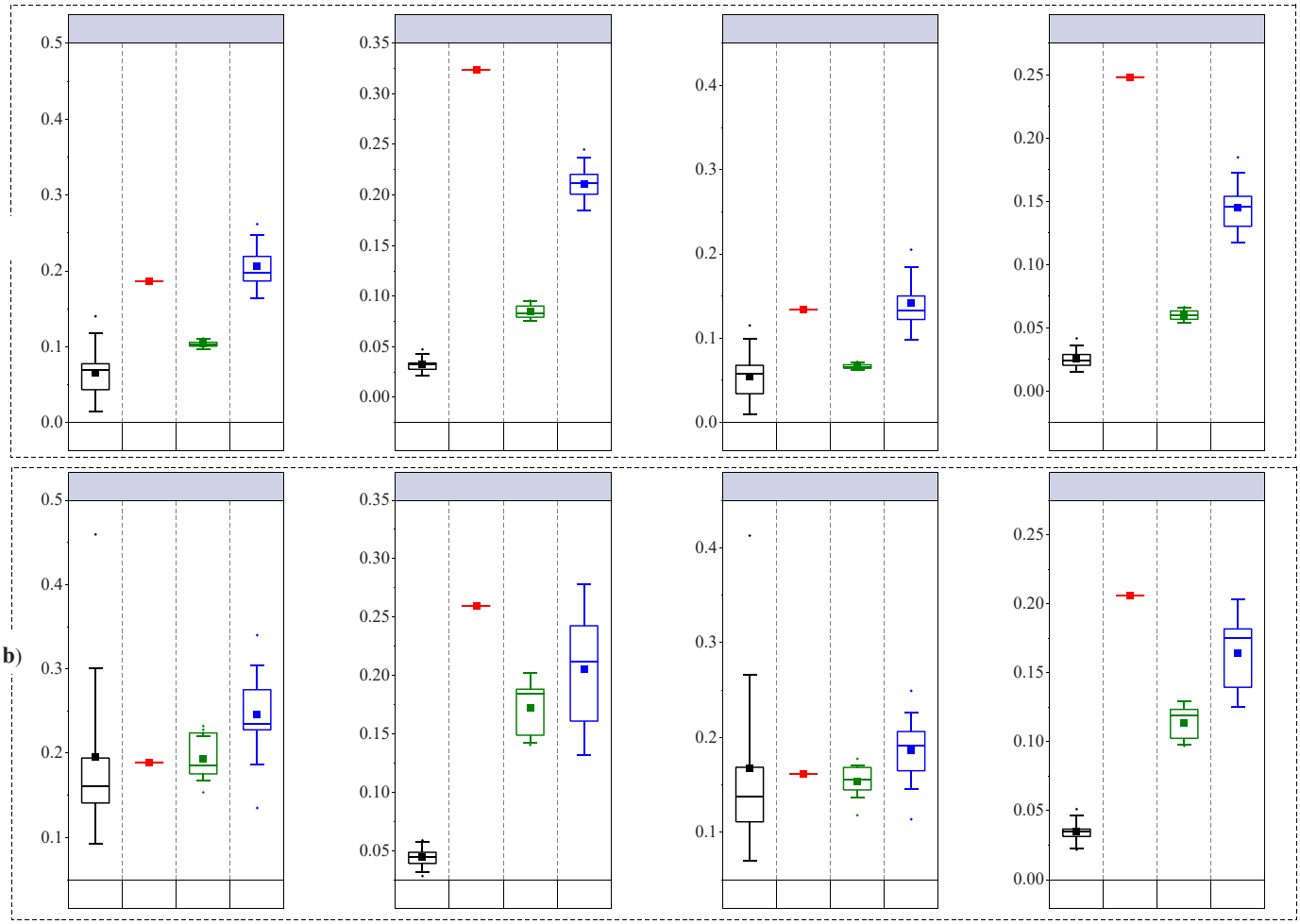


Fig. 7. Performance evaluation of ML algorithms (ANN-MLP, GU, RF, and RSS) in predicting SAR value in KP station (1985–2021) based on SC1 & SC2: a): training 70% and b): testing 30%.

water quality is a key parameter of monitoring crop health and production. Salinity and sodicity hazard recognized as SAR has obvious effects on the physical properties of soil disrupting the rate of infiltration (Suarez et al., 2006). Therefore, monitoring rainwater chemistry is evidently reported in several agricultural regions of the world (Wu et al., 2012; Zeng et al., 2020a; Zeng et al., 2020b). The findings from our study aligned with the review of Mohanavelu et al. (2021) with an obvious increasing trend of sodicity hazard with increasing SAR value above 2 since the early years of the 2000 s at all stations (Table 5 and Fig. 4). The high SAR values in particular years might reduce the soil infiltration and hydraulic conductivity (Suarez et al., 2008) causing below-average crop yield in the early years of the 2000 s (Harsányi et al., 2023). In other regions like the North China Plain, Wang et al. (2023) recently reported the long-term impacts of salinity and sodicity on wheat and maize yield. Similarly, Alsubih et al. (2022) also resulted that high concentrations of SAR and Na% in surface water from dams can be hazardous for irrigation purposes. Hence, predicting SAR from multi-statistical and machine-learning approaches is crucial for sustainable agriculture water management.

4.3. Machine learning application for predicting SAR from different rainwater input scenarios

Selection of the input variables and ML algorithms is a key to derive and explore non-linear relationships in complex environmental settings. Several studies have adopted machine and deep learning algorithms for predicting agricultural water quality (i.e., SAR) from several parameters (Neissi et al., 2020; Rahnama et al., 2020; Sattari et al., 2020; Sepahvand et al., 2021). For instance, we predicted SAR from ten rainwater inorganic ions employing four ML algorithms from the perspective of agriculture water management as also predicted in other studies (El Bilali and Taleb, 2020; El Bilali et al., 2021; Gharaibeh et al., 2021) from ground and surface water. Although, the findings from our study revealed that all ML algorithms including RSS, RF, GU, and ANN-MLP performed well in predicting SAR. However, ANN-MLP followed by RF provided the lowest RMSE range in multiple models runs of each scenario (SC-1, and SC-2) (Fig. 7). Rainwater ionic dataset for our study presented a high dimensionality with complex relationships. Ensemble ML methods of RSS and RF are proved to be appropriate to capture and understand such complexities and dimensionalities in the data for accurate predictions. The procedure of bootstrap sampling (randomly splitting the dataset with replacement) introduces variability in the model with diversified subsets (Arshad et al., 2023). Each subset is further used to train individual decision trees where the model learns the intricate non-linear relationships between the variables for providing accurate predictions in high dimensional hydrological applications (Chen et al., 2020). Another study by Castrillo and García (2020) also proved the efficiency of RF algorithm for water quality prediction attributed to its ensemble properties, mitigating the overfitting problems and demonstrating high generalization in a diversified environment. Another study by Alnahit et al. (2022) also provided a good opinion about the random forest RF model in predicting stream water quality from climatic and catchment-related variables. Moreover, GPR is also proved to be an accurate prediction model inspired through

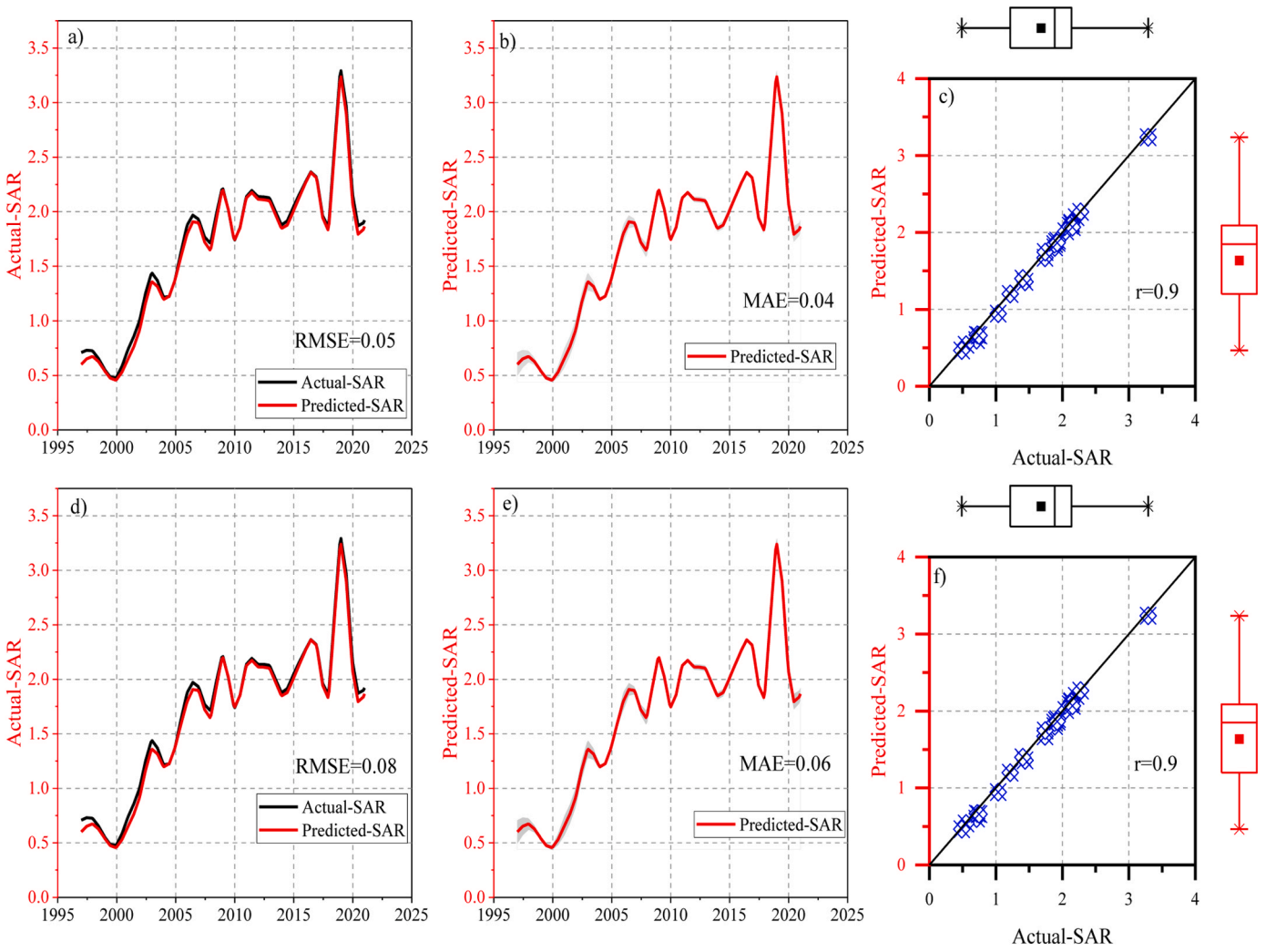


Fig. 8. ANN-MLP validation for forecasting SAR at a,b, c) FAK station, d, e, f,) NYR stations: a) temporal evaluation of observed SAR vs. predicted valued at FAK station, b) error in prediction of SAR value presented in gray shadow, c) scatter plot with box plot between observed SAR vs. predicted one; d) temporal evaluation of observed SAR vs. predicted valued at NYR station, e) error in prediction of SAR value presented in gray shadow, f) scatter plot with box plot between observed SAR vs. predicted one.

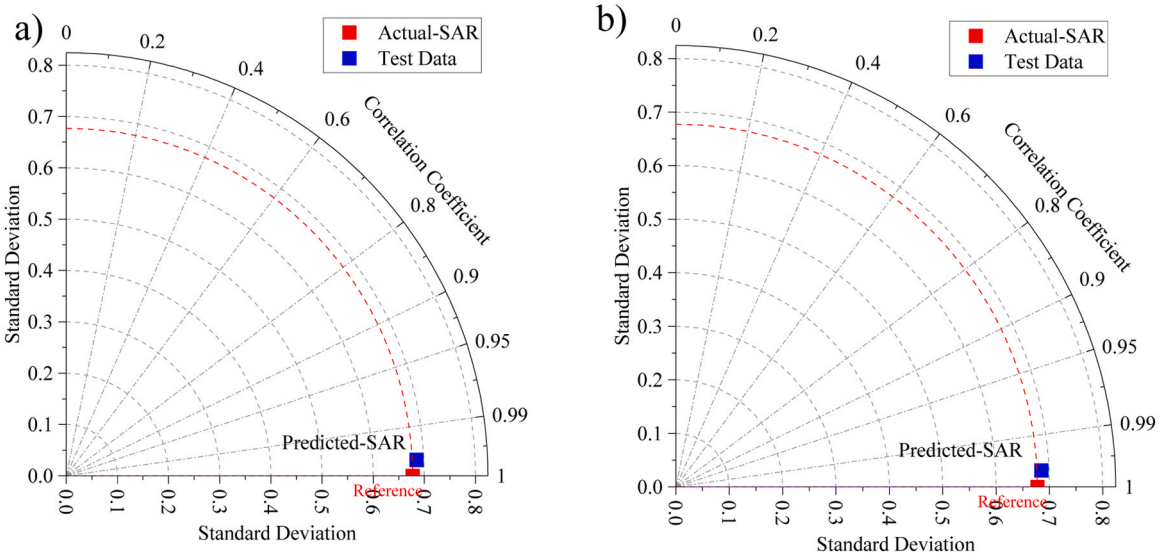


Fig. 9. Tylor diagram showing the relation between observed SAR values and predicted on at: a) FAK station, and b) NYR stations.

Bayesian framework to understand and model the data uncertainties and provide robust predictions from a probabilistic approach in water quality research (Uddin et al., 2023; Wang et al., 2023). In comparison with RF and other ML algorithms, ANN-MLP has been proven to perform superior in water quality studies (Abdel-Fattah et al., 2021; Chen et al., 2020; Ubah et al., 2021). For instance, Najah Ahmed et al. (2019) also provided an opinion about the better performance of ANN-MLP at the training stage for predicting water quality. A recent study by Chauhan and Trivedi (2023) also resulted in the efficiency of ANN model for predicting water quality index from several parameters. The architecture of ANN-MLP comprises several interconnected nodes or neurons which perform weighted computation using activation functions. The weighted connection between the neurons and sigmoid activation helps the model to learn and solve complex and non-linear problems providing robust predictions. Sigmoid activation not only introduces the nonlinearity in the model to understand and capture complex relationships but also facilitates the training process through backpropagation (BP) (a gradient-based optimization algorithm) for efficient model training. Hence, MLP architecture is proven to provide more robust computation for prediction SAR from irrigation water ions (Wagh et al., 2016). Other than ML algorithms, the selection of input variables also costs the prediction performance. For instance, Kouadri et al. (2022) also provided the opinion that the selection of the most significant water quality variables in particular scenarios is necessary to improve the efficiency of ML models. Likewise, our study tested the input variables in two different scenarios with a high performance of SC-2 in ANN-MLP with only three input variables (Ca^{2+}, Mg^{+}, and Na^{+}) providing accurate SAR predictions (Figs. 7–9) which are also supported by the findings of Gautam et al. (2023) and Kushwaha et al. (2023).

5. Limitations of research

Our study successfully predicted the SAR from several rainwater cations and anions in two scenario combinations employing four ML algorithms. The performance of ANN-MLP is found to be superior for predicting and forecasting SAR which is essential for agricultural water monitoring and management. Despite the highly accurate SAR forecasting by MLP and other ML algorithms, utilization of several deep learning algorithms like long short term memory (LSTM) is also reported to solve intricate variable relationships for monitoring agricultural and irrigation water quality (Docheshmeh Gorgij et al., 2023). Collection of extended sequential data in the hidden memory of LSTM architecture enables it to capture long term dependencies. Hence, LSTM can perform better in capturing long term temporal dynamics with shorter time intervals (Barzegar et al., 2020). Moreover, inclusion feature selection methods could also be tested to evaluate the performance of varied rainwater ionic performance for accurate predictions (El Bilali et al., 2021; Kushwaha et al., 2023). The study can also be extended more effective by predicting other agricultural water quality parameters like TDS, PS, MAR, for a comparative assessment of the impacts of different ionic species on agriculture output (El Bilali et al., 2021).

6. Conclusion

Agricultural water monitoring and management is highly dependent upon accurate predictions of several water quality indices. Following the exploratory analysis of rainwater chemistry at three selected stations across Hungary, our study employed four ML algorithms namely Random Forest, Gaussian Process Regression, Random subspace, and Artificial Neural Network-Multi layer perceptron to predict sodium adsorption ratio (SAR) in two scenarios combinations (SC-1; with all ten ionic predictors) and (SC-2; with 3 significant ionic predictors). First, both scenarios and all ML algorithms are trained, tested, and cross-validated with 10 repetitive model runs in each scenario and algorithm at a standard KP station. The best-selected ML algorithm and scenario is implemented to further validate its performance at two other (FAK, and NYR) stations with similar rainfall characteristics. The exploratory findings of rainwater chemistry from our study revealed a high concentration of SO_4^{-2}, NO_3^{-}, and Na^{+} ions over the whole timeseries. Man-Kendall trend test along with Sens slope analysis further revealed a decreasing trend of SO_4^{-2} and NO_3^{-} ions and an increasing trend of Na^{+} ions indicating increasing Na hazard in central Europe since the early years of 2000 s. PCA analysis also revealed a high loading of Na^{+} ions explaining more than 60% of variability along with other ionic species. Further, ML-based SAR predictions at KP stations clearly revealed the high performance of ANN-MLP in SC-2 with the lowest RMSE and MAE range in 10 repetitive model runs. Overall, the following sequence of ML models is found for predicting SAR in SC-2: ANN-MLP > RF > RSS > GU. Finally, the validation of best-selected ML algorithm (ANN-MLP) and scenario (SC-2) also provided highly accurate forecasting and predictions with a correlation coefficient of 0.9 between observed and predicted values. Hence, accurate SAR predictions propose proactive measures to mitigate and reduce the potential risk of water contamination hazards in agricultural applications. It can provide a means for sustainable agricultural water management to optimize agricultural production.

Funding

This research was supported by the Researchers Supporting Project, Grant number (RSP2024R296), King Saud University, Riyadh, Saudi Arabia. Also, Project no. TKP2021-NKTA-32 has been implemented with support from Hungary's National Research, Development, and Innovation Fund, financed under the TKP2021-NKTA funding scheme.

CRediT authorship contribution statement

Bashir Bashar: Writing – review & editing. **Vad Attila:** Writing – review & editing. **Alsalman Abdullah:** Writing – review & editing. **Harsányi Endre:** Funding acquisition, Writing – review & editing. **Mohammed Safwan:** Conceptualization, Methodology, Visualization, Writing – original draft. **Arshad Sana:** Methodology, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

References

Abdel-Fattah, M.K., Mokhtar, A., Abdo, A.I., 2021. Application of neural network and time series modeling to study the suitability of drain water quality for irrigation: a case study from Egypt. Environ. Sci. Pollut. Res. 28, 898–914.

Ahaninjan, K., Egdernezhad, A., 2020. Modeling qualitative parameters of SAR, EC, and TDS in groundwater using optimized artificial neural network model (Case Study: Behbahan Plain). Environ. Water Eng. 6, 161–172.

Alastuey, A., Querol, X., Chaves, A., Ruiz, C.R., Carratala, A., Lopez-Soler, A., 1999. Bulk deposition in a rural area located around a large coal-fired power station, northeast Spain. Environ. Pollut. 106, 359–367.

Al-Momani, I.F., Ataman, O.Y., Anwari, M.A., Tuncel, S., Köse, C., Tuncel, G., 1995. Chemical composition of precipitation near an industrial area at Izmir, Turkey. Atmos. Environ. 29, 1131–1143.

Alnahit, A.O., Mishra, A.K., Khan, A.A., 2022. Stream water quality prediction using boosted regression tree and random forest models. Stoch. Environ. Res. Risk Assess. 36, 2661–2680.

Alsubih, M., Mallick, J., Islam, A.R.M.T., Almesfer, M.K., Kahla, N.B., Talukdar, S., Ahmed, M., 2022. Assessing surface water quality for irrigation purposes in some dams of asir region, saudi arabia using multi-statistical modeling approaches. Water 14, 1439.

Arshad, S., Kazmi, J.H., Javed, M.G., Mohammed, S., 2023. Applicability of machine learning techniques in predicting wheat yield based on remote sensing and climate data in Pakistan, South Asia. Eur. J. Agron. 147, 126837.
Avila, R., Horn, B., Moriarty, E., Hodson, R., Moltchanova, E., 2018. Evaluating statistical model performance in water quality prediction. J. Environ. Manag. 206, 910–919.
Barzegar, R., Aalami, M.T., Adamowski, J., 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. Stoch. Environ. Res. Risk Assess. 34, 415–433.
Birkás, M., Dekemati, I., 2023. Agricultural soil degradation in Hungary. In: Pereira, P., Muñoz-Rojas, M., Bogunovic, I., Zhao, W. (Eds.), Impact of Agriculture on Soil Degradation II: A European Perspective. Springer International Publishing, Cham, pp. 139–157.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.
Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.
Buishand, T.A., 1982. Some methods for testing the homogeneity of rainfall records. J. Hydrol. 58, 11–27.
Bussay, A., van der Velde, M., Fumagalli, D., Seguini, L., 2015. Improving operational maize yield forecasting in Hungary. Agric. Syst. 141, 94–106.
Ćalasan, M., Abdel Aleem, S.H.E., Zobaa, A.F., 2020. On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: a novel exact analytical solution based on Lambert W function. Energy Convers. Manag. 210, 112716.
Çankaya, Ş., Varol, M., Bekleyen, A., 2023. Hydrochemistry, water quality and health risk assessment of streams in Bismil plain, an important agricultural area in southeast Türkiye. Environ. Pollut. 331, 121874.
Cao, Y.-Z., Wang, S., Zhang, G., Luo, J., Lu, S., 2009. Chemical characteristics of wet precipitation at an urban site of Guangzhou, South China. Atmos. Res. 94, 462–469.
Castrillo, M., García, Á.L., 2020. Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. Water Res. 172, 115490.
Celle-Jeanton, H., Travi, Y., Loÿe-Pilot, M.-D., Huneau, F., Bertrand, G., 2009. Rainwater chemistry at a Mediterranean inland station (Avignon, France): local contribution versus long-range supply. Atmos. Res. 91, 118–126.
Chang, C.-T., Yang, C.-J., Huang, K.-H., Huang, J.-C., Lin, T.-C., 2022. Changes of precipitation acidity related to sulfur and nitrogen deposition in forests across three continents in north hemisphere over last two decades. Sci. Total Environ. 806, 150552.
Chauhan, S.S., Trivedi, M.K., 2023. Artificial neural network-based assessment of water quality index (WQI) of surface water in Gwalior-Chambal region. Int. J. Energy Environ. Eng. 14, 47–61.
Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., Ren, H., 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Res. 171, 115454.
Costa, A.C., Soares, A., 2009. Homogenization of climate data: review and new perspectives using geostatistics. Math. Geosci. 41, 291–305.
Daliakopoulos, I.N., Tsanis, I.K., Koutroulis, A., Kourgialas, N.N., Varouchakis, A.E., Karatzas, G.P., Ritsema, C.J., 2016. The threat of soil salinity: a European scale review. Sci. Total Environ. 573, 727–739.
Das, R., Das, S.N., Misra, V.N., 2005. Chemical composition of rainwater and dustfall at Bhubaneswar in the east coast of India. Atmos. Environ. 39, 5908–5916.
Docheshmeh Gorgij, A., Askari, G., Taghipour, A.A., Jami, M., Mirfardi, M., 2023. Spatiotemporal forecasting of the groundwater quality for irrigation purposes, using deep learning method: long short-term memory (LSTM). Agric. Water Manag. 277, 108088.
El Bilali, A., Taleb, A., 2020. Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. J. Saudi Soc. Agric. Sci. 19, 439–451.
El Bilali, A., Taleb, A., Brouziyne, Y., 2021. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agric. Water Manag. 245, 106625.
Elbeltagi, A., Srivastava, A., Deng, J., Li, Z., Raza, A., Khadke, L., Yu, Z., El-Rawy, M., 2023. Forecasting vapor pressure deficit for agricultural water management using machine learning in semi-arid environments. Agric. Water Manag. 283, 108302.
Facchini Cerqueira, M.R., Pinto, M.F., Derossi, I.N., Esteves, W.T., Rachid Santos, M.D., Costa Matos, M.A., Lowinsohn, D., Matos, R.C., 2014. Chemical characteristics of rainwater at a southeastern site of Brazil. Atmos. Pollut. Res. 5, 253–261.
Gautam, V.K., Pande, C.B., Moharir, K.N., Varade, A.M., Rane, N.L., Egbueri, J.C., Alshehri, F., 2023. Prediction of sodium hazard of irrigation purpose using artificial neural network modelling. Sustainability 15, 7593.
Ge, B., Xu, D., Wild, O., Yao, X., Wang, J., Chen, X., Tan, Q., Pan, X., Wang, Z., 2021. Inter-annual variations of wet deposition in Beijing from 2014–2017: implications of below-cloud scavenging of inorganic aerosols. Atmos. Chem. Phys. 21, 9441–9454.
Gharaibeh, M.A., Albalasmeh, A.A., Pratt, C., El Hanandeh, A., 2021. Estimation of exchangeable sodium percentage from sodium adsorption ratio of salt-affected soils using traditional and dilution extracts, saturation percentage, electrical conductivity, and generalized regression neural networks. Catena 205, 105466.
Güçlü, Y.S., 2020. Improved visualization for trend analysis by comparing with classical Mann-Kendall test and ITA. J. Hydrol. 584, 124674.
Harsányi, E., Bashir, B., Arshad, S., Ocwa, A., Vad, A., Alsalman, A., Bácskai, I., Rátonyi, T., Hijazi, O., Széles, A., Mohammed, S., 2023. Data mining and machine learning algorithms for optimizing maize yield forecasting in central Europe. Agronomy 13, 1297.
Hontoria, C., Saa, A., Almorox, J., Cuadra, L., Sánchez, A., Gascó, J.M., 2003. The chemical composition of precipitation in Madrid. Water Air Soil Pollut. 146, 35–54.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366.
Kendall, M.G., 1948. Rank Correlation Methods. Griffin, Oxford, England.
Keresztesi, Á., Birsan, M.-V., Nita, I.-A., Bodor, Z., Szép, R., 2019. Assessing the neutralisation, wet deposition and source contributions of the precipitation chemistry over Europe during 2000–2017. Environ. Sci. Eur. 31, 50.
Keresztesi, Á., Nita, I.-A., Birsan, M.-V., Bodor, Z., Pernyeszi, T., Micheu, M.M., Szép, R., 2020a. Assessing the variations in the chemical composition of rainwater and air masses using the zonal and meridional index. Atmos. Res. 237, 104846.
Keresztesi, Á., Nita, I.-A., Boga, R., Birsan, M.-V., Bodor, Z., Szép, R., 2020b. Spatial and long-term analysis of rainwater chemistry over the conterminous United States. Environ. Res. 188, 109872.
Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. Agric. For. Meteorol. 260-261, 300–320.
Khan, S., Ullah, Q., Khan, A.A., Hassan, S.S., Shakoor, A., Ijaz, M., 2022. Geostatistical investigation of groundwater quality zones for its applications in irrigated agriculture areas of Punjab (Pakistan). Environ. Earth Sci. 81, 91.
Klopp, H.W., Daigh, A.L.M., 2020. Measured saline and sodic solutions effects on soil saturated hydraulic conductivity, electrical conductivity and sodium adsorption ratio. Arid Land Res. Manag. 34, 264–286.
Koseoglu-Imer, D.Y., Oral, H.V., Coutinho Calheiros, C.S., Krzeminski, P., Güçlü, S., Pereira, S.A., Surmacz-Górska, J., Plaza, E., Samaras, P., Binder, P.M., van Hullebusch, E.D., Devolli, A., 2023. Current challenges and future perspectives for the full circular economy of water in European countries. J. Environ. Manag. 345, 118627.
Kouadri, S., Pande, C.B., Panneerselvam, B., Moharir, K.N., Elbeltagi, A., 2022. Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. Environ. Sci. Pollut. Res. 29, 21067–21091.
Kushwaha, N.L., Rajput, J., Suna, T., Sena, D.R., Singh, D.K., Mishra, A.K., Sharma, P.K., Mani, I., 2023. Metaheuristic approaches for prediction of water quality indices with relief algorithm-based feature selection. Ecol. Inform. 75, 102122.
Lai, C., Reinders, M.J.T., Wessels, L., 2006. Random subspace method for multivariate feature selection. Pattern Recognit. Lett. 27, 1067–1076.
Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.
Li, J., Li, R., Cui, L., Meng, Y., Fu, H., 2019. Spatial and temporal variation of inorganic ions in rainwater in Sichuan province from 2011 to 2016. Environ. Pollut. 254, 112941.
Li, Y., Mi, W., Ji, L., He, Q., Yang, P., Xie, S., Bi, Y., 2023. Urbanization and agriculture intensification jointly enlarge the spatial inequality of river water quality. Sci. Total Environ. 878, 162559.
Liu, X., Al-Shaibah, B., Zhao, C., Tong, Z., Bian, H., Zhang, F., Zhang, J., Pei, X., 2022. Estimation of the key water quality parameters in the surface water, middle of Northeast China, based on Gaussian process regression. Remote Sens. 14, 6323.
Liu, Y., Huang, D., Liu, B., Feng, Q., Cai, B., 2021. Adaptive ranking based ensemble learning of Gaussian process regression models for quality-related variable prediction in process industries. Appl. Soft Comput. 101, 107060.
Lu, H., Ma, X., 2020. Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere 249, 126169.
Lü, P., Han, G., Wu, Q., 2017. Chemical characteristics of rainwater in karst rural areas, Guizhou Province, Southwest China. Acta Geochim. 36, 572–576.
MacFarland, T.W., Yates, J.M., 2016. Mann–Whitney U test. In: MacFarland, T.W., Yates, J.M. (Eds.), Introduction to Nonparametric Statistics for the Biological Sciences Using R. Springer International Publishing, Cham, pp. 103–132.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13, 245–259.
Meng, L., Liu, H., L. Ustin, S., Zhang, X., 2021. Predicting maize yield at the plot scale of different fertilizer systems by multi-source data and machine learning methods. Remote Sens.
Minhas, P.S., Qadir, M., Yadav, R.K., 2019. Groundwater irrigation induced soil sodification and response options. Agric. Water Manag. 215, 74–85.
Mohanavelu, A., Naganna, S.R., Al-Ansari, N., 2021. Irrigation induced salinity and sodicity hazards on soil and groundwater: an overview of its causes, impacts and mitigation strategies. Agriculture 11, 983.
Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N., Abdel-Fattah, M.K., 2022. Prediction of irrigation water quality indices based on machine learning and regression models. Appl. Water Sci. 12, 76.
Mustafa, H.M., Mustapha, A., Hayder, G., Salisu, A., 2021. Applications of IoT and artificial intelligence in water quality monitoring and prediction: a review, 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 968–975.
Najah, A., El-Shafie, A., Karim, O.A., El-Shafie, A.H., 2013. Application of artificial neural networks for water quality prediction. Neural Comput. Appl. 22, 187–201.
Najah Ahmed, A., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., Ehteram, M., Elshafie, A., 2019. Machine learning methods for better water quality prediction. J. Hydrol. 578, 124084.
Nasiruddin Khan, M., Sarwar, A., 2014. Chemical composition of wet precipitation of air pollutants: a case study in Karachi, Pakistan. Atmósfera 27, 35–46.
Neissi, L., Golabi, M., Gorman, J.M., 2020. Spatial interpolation of sodium absorption ratio: a study combining a decision tree model and GIS. Ecol. Indic. 117, 106611.
Nong, X., Lai, C., Chen, L., Shao, D., Zhang, C., Liang, J., 2023. Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using

support vector regression coupled with multiple feature engineering and optimization methods: a case study in China. Ecol. Indic. 146, 109845.
Nouraki, A., Alavi, M., Golabi, M., Albaji, M., 2021. Prediction of water quality parameters using machine learning models: a case study of the Karun River, Iran. Environ. Sci. Pollut. Res. 28, 57060–57072.
Omeka, M.E., 2023. Evaluation and prediction of irrigation water quality of an agricultural district, SE Nigeria: an integrated heuristic GIS-based and machine learning approach. Environ. Sci. Pollut. Res.
Pinke, Z., Decsi, B., Kozma, Z., Vári, Á., Lövei, G.L., 2020. A spatially explicit analysis of wheat and maize yield sensitivity to changing groundwater levels in Hungary, 1961–2010. Sci. Total Environ. 715, 136555.
Rahnama, E., Bazrafshan, O., Asadollahfardi, G., 2020. Application of data-driven methods to predict the sodium adsorption rate (SAR) in different climates in Iran. Arab. J. Geosci. 13, 1160.
Redington, A.L., Derwent, R.G., Witham, C.S., Manning, A.J., 2009. Sensitivity of modelled sulphate and nitrate aerosol to cloud, pH and ammonia emissions. Atmos. Environ. 43, 3227–3234.
Rodhe, H., Dentener, F., Schulz, M., 2002. The global distribution of acidifying wet deposition. Environ. Sci. Technol. 36, 4382–4388.
Saha, S., Kundu, B., Paul, G.C., Pradhan, B., 2023. Proposing an ensemble machine learning based drought vulnerability index using M5P, dagging, random sub-space and rotation forest models. Stoch. Environ. Res. Risk Assess. 37, 2513–2540.
Sattari, M.T., Feizi, H., Colak, M.S., Ozturk, A., Apaydin, H., Ozturk, F., 2020. Estimation of sodium adsorption ratio in a river with kernel-based and decision-tree models. Environ. Monit. Assess. 192, 575.
Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's Tau. J. Am. Stat. Assoc. 63, 1379–1389.
Sepahvand, A., Singh, B., Sihag, P., Nazari Samani, A., Ahmadi, H., Fiz Nia, S., 2021. Assessment of the various soft computing techniques to predict sodium absorption ratio (SAR). ISH J. Hydraul. Eng. 27, 124–135.
Shadrin, D., Nikitin, A., Tregubova, P., Terekhova, V., Jana, R., Matveev, S., Pukalchik, M., 2021. An automated approach to groundwater quality monitoring—geospatial mapping based on combined application of gaussian process regression and bayesian information criterion. Water 13, 400.
Singh, K.P., Singh, V.K., Malik, A., Sharma, N., Murthy, R.C., Kumar, R., 2007. Hydrochemistry of wet atmospheric precipitation over an urban area in Northern Indo-gangetic plains. Environ. Monit. Assess. 131, 237–254.
Skurichina, M., Duin, R.P.W., 2002. Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal. Appl. 5, 121–135.
Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinforma. 9, 307.
Suarez, D.L., Wood, J.D., Lesch, S.M., 2006. Effect of SAR on water infiltration under a sequential rain–irrigation management system. Agric. Water Manag. 86, 150–164.
Suarez, D.L., Wood, J.D., Lesch, S.M., 2008. Infiltration into cropped soils: effect of rain and sodium adsorption ratio–impacted irrigation water. J. Environ. Qual. 37. S-169-S-179.
Tin Kam, H., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844.
Trabelsi, F., Bel Hadj Ali, S., 2022. Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in Medjerda River Basin, Tunisia. Sustainability 14, 2341.
Ubah, J.I., Orakwe, L.C., Ogbu, K.N., Awu, J.I., Ahaneku, I.E., Chukwuma, E.C., 2021. Forecasting water quality parameters using artificial neural network for irrigation purposes. Sci. Rep. 11, 24438.
Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I., 2023. A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. Water Res. 229, 119422.
Vet, R., Artz, R.S., Carou, S., Shaw, M., Ro, C.-U., Aas, W., Baker, A., Bowersox, V.C., Dentener, F., Galy-Lacaux, C., Hou, A., Pienaar, J.J., Gillett, R., Forti, M.C., Gromov, S., Hara, H., Khodzher, T., Mahowald, N.M., Nickovic, S., Rao, P.S.P., Reid, N.W., 2014. A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus. Atmos. Environ. 93, 3–100.
Vlastos, D., Antonopoulou, M., Lavranou, A., Efthimiou, I., Dailianis, S., Hela, D., Lambropoulou, D., Paschalidou, A.K., Kassomenos, P., 2019. Assessment of the toxic potential of rainwater precipitation: First evidence from a case study in three Greek cities. Sci. Total Environ. 648, 1323–1332.
Wagh, V.M., Panaskar, D.B., Muley, A.A., Mukate, S.V., Lolage, Y.P., Aamalawar, M.L., 2016. Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. Model. Earth Syst. Environ. 2, 1–10.
Wan, X., Li, X., Wang, X., Yi, X., Zhao, Y., He, X., Wu, R., Huang, M., 2022. Water quality prediction model using Gaussian process regression based on deep learning for carbon neutrality in papermaking wastewater treatment system. Environ. Res. 211, 112942.
Wang, H., Zheng, C., Ning, S., Cao, C., Li, K., Dang, H., Wu, Y., Zhang, J., 2023. Impacts of long-term saline water irrigation on soil properties and crop yields under maize-wheat crop rotation. Agric. Water Manag. 286, 108383.
Wang, Y., Feng, B., Hua, Q.-S., Sun, L., 2021. Short-term solar power forecasting: a combined long short-term memory and gaussian process regression method. Sustainability 13, 3665.
Whelpdale, D.M., Summers, P.W., Sanhueza, E., 1997. A global overview of atmospheric acid deposition fluxes. Environ. Monit. Assess. 48, 217–247.
Wriedt, G., Van der Velde, M., Aloe, A., Bouraoui, F., 2009. Estimating irrigation water requirements in Europe. J. Hydrol. 373, 527–544.
Wu, Q., Han, G., Tao, F., Tang, Y., 2012. Chemical composition of rainwater in a karstic agricultural area, Southwest China: the impact of urbanization. Atmos. Res. 111, 71–78.
Xiao, H.-W., Xiao, H.-Y., Long, A.-M., Wang, Y.-L., Liu, C.-Q., 2013. Chemical composition and source apportionment of rainwater at Guiyang, SW China. J. Atmos. Chem. 70, 269–281.
Xu, Z., Wu, Y., Liu, W.-J., Liang, C.-S., Ji, J., Zhao, T., Zhang, X., 2015. Chemical composition of rainwater and the acid neutralizing effect at Beijing and Chizhou city, China. Atmos. Res. 164-165, 278–285.
Yadav, A.K., Malik, H., Chandel, S.S., 2014. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew. Sustain. Energy Rev. 31, 509–519.
Yuan, J., Li, Y., Shan, Y., Tong, H., Zhao, J., 2023. Effect of magnesium ions on the mechanical properties of soil reinforced by microbially induced carbonate precipitation. J. Mater. Civ. Eng. 35, 04023413.
Zaman, M., Shahid, S.A., Heng, L., 2018. Irrigation water quality. In: Zaman, M., Shahid, S.A., Heng, L. (Eds.), Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques. Springer International Publishing, Cham, pp. 113–131.
Zare Farjoudi, S., Alizadeh, Z., 2021. A comparative study of total dissolved solids in water estimation models using Gaussian process regression with different kernel functions. Environ. Earth Sci. 80, 557.
Zeng, J., Han, G., Wu, Q., Tang, Y., 2020a. Effects of agricultural alkaline substances on reducing the rainwater acidification: Insight from chemical compositions and calcium isotopes in a karst forests area. Agric., Ecosyst. Environ. 290, 106782.
Zeng, J., Yue, F.-J., Li, S.-L., Wang, Z.-J., Wu, Q., Qin, C.-Q., Yan, Z.-L., 2020b. Determining rainwater chemistry to reveal alkaline rain trend in Southwest China: evidence from a frequent-rainy karst area with extensive agricultural production. Environ. Pollut. 266, 115166.
Zeng, J., Han, G., Zhang, S., Xiao, X., Li, Y., Gao, X., Wang, D., Qu, R., 2022. Rainwater chemical evolution driven by extreme rainfall in megacity: Implication for the urban air pollution source identification. J. Clean. Prod. 372, 133732.
Zhou, X., Xu, Z., Liu, W., Wu, Y., Zhao, T., Jiang, H., Zhang, X., Zhang, J., Zhou, L., Wang, Y., 2019. Chemical composition of precipitation in Shenzhen, a coastal mega-city in South China: Influence of urbanization and anthropogenic activities on acidity and ionic composition. Sci. Total Environ. 662, 218–226.