https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences Discussions
Open Access

# Prediction of groundwater quality index to assess suitability for drinking purpose using averaged neural network and geospatial analysis

Seok Hyun Ahn[1,¶], Do Hwan Jeong[2,¶], MoonSu Kim[2], Tae Kwon Lee[1,*], Hyun-Koo Kim[2,*]

[1]Department of Environmental Engineering, Yonsei University, Wonju 26493, South Korea

[2]Soil and Groundwater Division, National Institute of Environmental Research, Incheon 22689, South Korea

*Correspondence:

HK Kim, email: khk228@korea.kr,

TK Lee, email: tklee@yonsei.ac.kr

[¶]These authors contributed equally to this work.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions

**Abstract**

The aims of this study were to determine the groundwater quality index (GQI) using an averaged neural network and evaluate its field applicability with two-dimensional (2D) spatial analysis. The GQI was computed using 29 water quality parameters obtained at 3,552 portable groundwater wells used as drinking water sources. The GQI was divided into the following three grades: 'worrisome', <0.89 (20.1% of the wells); 'good', 0.89–0.94 (62.8%); and 'very good', >0.94 (17.1%). Based on the random forest, the most important water quality parameters were general bacteria, turbidity and nitrate. The 2D spatial analysis confirmed notable differences in the GQI grades among regions. The 10-year long-term groundwater quality monitoring in the 'worrisome' grade showed the nitrate and chloride concentrations have continuously increased. These results indicate that the coupling of the GQI with 2D spatial analysis is a promising approach that can be applied in groundwater management and vulnerability assessment.

Keywords: Groundwater, Water quality index, Data-Driven modelling, 2D spatial analysis, vulnerability assessment

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions

## 1 Introduction

Groundwater is the preferred source of drinking water worldwide (Guppy et al., 2018) . Groundwater accounts for almost half of all drinking water, ~40% of irrigation water, and a third of the water used for industrial purposes (Centre, 2018). To systematically manage groundwater and prevent contamination, the groundwater quality is monitored by using national groundwater monitoring wells and the wells managed by the local government (Lapworth et al., 2019). Because these wells were established in consideration of the land use and hydrogeological characteristics of each region and monitored for a long-time, it is easy to identify correlations between groundwater pollution and environmental factors (Li et al., 2019). To analyse the groundwater quality at a higher spatial resolution, the authorities classify groundwater quality categories into microbial, chemical, acceptability, and radiological aspects and manage the groundwater quality by using various parameters for each category. With the increasing accumulation of groundwater quality data reflecting the land use and hydrogeological characteristics, the number of attempts to predict the water quality or derive water quality management policies has increased (Bhanja et al., 2017).

The water quality index (WQI) is an algebraic approach to evaluate the water quality by converting chemical, physical, and biological parameters with different units into a single value (Misaghi et al., 2017). Many WQIs have been established to characterise surface water such as the WQI developed by the National Science Foundation (Brown et al., 1970), the Oregon Water Quality Index (Cude, 2001), Canadian Council of Ministers of the Environment WQI (Lumb et al., 2006), and Weighted Arithmetic WQI (Brown et al., 1972). Attempts have been made to apply the WQI to other water resources, such as ground- and seawater, but in most cases, the surface water indices were used (Jha et al., 2015). Because groundwater has different properties than surface water due to chemical and physical processes, including the hydrochemistry, mineralogy of aquifer, and precipitation–dissolution processes, it is necessary to develop a WQI suitable for groundwater. Although in a few groundwater quality index (GQI) studies, it has been attempted to evaluate the groundwater quality with water temperature, nitrate, pH as main parameters, it is difficult to accurately determine the water quality with a limited number of parameters and data (Abbasnia et al., 2019; Gao et al., 2020). Therefore, the GQI must be improved by incorporating the massive amount of groundwater data, which was recently collected by the authorities.

Recently, many attempts have been made to predict the groundwater quality by using data-driven models. Data-driven models predict the output using a mathematical model of input variables derived by the supervised learning of a given dataset. Algorithms for learning the datasets include neural network, fuzzy systems, support vector machine (SVM), ensemble trees, and discriminant analysis (Wei et al., 2018). Previously, researchers predicted water quality indicators such as the dissolved oxygen, turbidity, pH, and ammonia using data-driven modelling (artificial neural network, ANN; random forest, RF; SVM) and yielded model prediction errors below 15% (Antanasijević et al., 2014; Meyers et al., 2017; Najah Ahmed et al., 2019). In most previous studies using data-driven models, single water quality parameters were predicted by utilising multiple water quality parameters as input variables. However, a data-driven model that can be used to evaluate or predict the overall water quality, such as the WQI or vulnerability assessment, has not been developed.

The objective of this study was to test the hypothesis that data-driven models can be applied to potable groundwater in South Korea to accurately predict the groundwater vulnerability and

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.



determine the groundwater quality. The prediction of the groundwater vulnerability is necessary to establish policies based on prioritising regions requiring groundwater quality management. We collected water quality datasets including 47 parameters and 8,326 wells through the ‘Safe Groundwater Project in Unsupplied Areas (2017–2020)’ (Fig 1). We calculated the single distance score for each well with potable groundwater by determining the difference between the water quality parameters and water quality standards. We then created a model by using the water quality datasets and distance score utilising data-driven techniques including averaged neural networks.. Regions with a high groundwater pollution vulnerability were selected by linking the binning technique with two-dimensional (2D) spatial analysis, and the accuracy of the groundwater pollution vulnerability was evaluated by analysing the long-term monitoring results obtained at national groundwater monitoring wells in the selected regions. The results show that a simple WQI with data-driven modelling is sufficient to select priority groundwater quality management areas. Although the focus of this study was placed on potable groundwater, our results can be used as guidance for data-driven modelling efforts considering other water resources that are directly related to human health (e.g., irrigation and drinking water).

## 2 Methods

### 2.1 Study sites

Data were acquired from 2017 to 2020 within the framework of the ‘Safe Groundwater Project in Unsupplied Areas (2017–2020)’ managed by the National Institute of Environmental Research, South Korea. The data included 47 drinking water quality parameters for a total of 8,326 Korean groundwater wells (2017: 2,061 wells; 2018: 2,142 wells; 2019: 2,019 wells; 2020: 2,104 wells).

© Author(s) 2022. CC BY 4.0 License.


The locations of the wells are shown in Fig. 1. All wells used in this study have been used as drinking water sources by Koreans. Information about the year of the well development and land use type was obtained by surveys. The groundwater quality parameters used in this study, including harmful inorganics and organics, microorganisms, and substances affecting the aesthetics, and standards are summarised in Table S1.

## 2.2 Data preprocessing

The R software (version 3.6.1) was used for data preprocessing. Because the aim of this study was to evaluate the potential use of groundwater as drinking water based on the prediction of groundwater quality, wells exceeding the groundwater quality standard were excluded from the analysis. In total, 4,774 wells had an inappropriate groundwater quality, representing 57.3% of the total groundwater (8,326 wells). Thus, 3,552 wells with potable groundwater were used for data-driven modelling.

Among the 47 water quality parameters, 18 parameters that were not detectable in the potable groundwater were removed. The 18 parameters included three with the aesthetic effects (detergents, smell, and taste), two harmful inorganic materials (cadmium and cyanide), 10 harmful organic materials (phenol, diazinon, parathion, fenitrothion, carbaryl, 1,1,1-trichloroethane, tetrachloroethylene, 1,2-dibromo-3-chloropropane, carbon tetrachloride, and 1,1-dichloroethylene), and three microorganisms (total coliform, faecal coliform, and *Escherichia coli*). The pH was not considered for further study because it insignificantly affects the analysis if the drinkable level is satisfied. Therefore, the analysis and modelling were carried out with 28 parameters (general bacteria, lead, fluorine, arsenic, caesium, mercury, chromium, boron, copper,

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


zinc, chlorine, iron, manganese, aluminium, ammonium, nitrate, sulphate, potassium permanganate consumption, trichloroethylene, dichloromethane, benzene, toluene, ethylbenzene, xylene, 1,4-dioxane, total hardness, colour, and turbidity).

### 2.3 Calculation of the groundwater quality index (GQI)

To evaluate the water quality indicators used in this study in the form of a single quantitative index, the difference between each water quality indicator and the water quality criterion was calculated. First, for potable groundwater, min–max normalization (Patro and Sahu, 2015) was performed, where the water quality standard was used as max for each parameter (Fig. 2A). The following equation was used:

$$P_n = \frac{P_e - P_{e\,min}}{P_s - P_{e\,min}}, \qquad (1)$$

where $P_n$ denotes the normalised value, $P_e$ represents the value of each parameter, and $P_S$ is the groundwater quality standard for each parameter. $P_{e\,min}$ represents the minimum of $P_e$. Because all data used for the modelling represent potable groundwater, $P_S$ is the maximum value.

Second, the deviation was calculated using the following simple equation:

$$P_d = 1 - P_n, \qquad (2)$$

where $P_d$ is the parameter's deviation value and 1 is the groundwater quality standard for each parameter. In min–max normalization, the minimum value becomes 0 and the maximum value becomes 1 because the maximum value is based on the groundwater quality.

In the third step, the distance score of the well as the single quantitative index was calculated using the following equation:

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


$$W_d = \frac{\sum_{i=1}^{x}(P_d)^2}{x}, \tag{3}$$

where $W_d$ is the distance score and $x$ represents the number of parameters. The calculated distance score is called groundwater quality index (GQI).

Based on the GQI, the data were divided into three grades, where $W_d =$ 0.94 corresponds to the top 20% and $W_d = 0.89$ corresponds to the bottom 20%. A value <0.89, >0.94, and ranging between 0.84–0.94 represents worrisome, very good, and good areas, respectively.

## 2.4 Model setup

Data-driven modelling was carried out using the R package 'caret' (version 6.0-86) (Kuhn, 2008). 'Caret' is an abbreviation for classification and regression training and the package contains useful functions for the creation of predictive models (Fig. 2B). It focuses on simplifying training and tuning processes. It also contains functions for training data preprocessing, parameter importance calculation, and model visualization. It has the advantage of enabling the parallel processing of multiple models. A total of ten classification models were chosen and used in this study. The ten models were used to classify unlearned groundwater vulnerability grades by learning the previously preprocessed 28 water quality parameters. The models used include averaged neural network, RF, SVM, ensemble trees, bagged flexible discriminant analysis, gradient boosting, penalised discriminant analysis, boosted logistic regression, ROC-based classifier, and Naïve Bayes. The averaged neural network, an ANN model, was created by applying an averaging technique to the neural network model. It is generated by modifying the functions of ordinary differential equations applied to the neural network and can be used for both regression and classification analysis. Decision tree and RF are tree-based models that operate with the goal of

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


dividing feature space into multiple areas. The RF is a model in which the performance is improved based on the use of an ensemble technique called bagging. Discriminant analysis is an analysis method that is used to identify criteria that can determine which population these samples were extracted from using sample information from two populations. In this study, bagged flexible discriminant analysis and penalised discriminant analysis were used. The Naïve Bayes classifier is a technique prediction based on simple probabilistic and on the application of the Bayes theorem (or Bayes rule) with a strong independence assumption. The SVM is one of the representative data-driven modelling methods, which is based on an algorithm that identifies boundaries dividing groups of data by the largest margin. The ROC-based classification is a model based on ROC analysis and can only be used for classification analysis. The ROC analysis is a method based on which the below areas are compared by visualising the performance of the classifier with an ROC curve. With respect to ensemble techniques, we used gradient boosting and boosted linear regression models, which improve the accuracy of the model by continuously reducing the residual during the learning process. The ten models described above were trained using caret's default training and all models involved 5-flod cross-validation by splitting the data into five subsets to compare the model performance.

The classification performance of the different models was measured using common error metrics, that is, the accuracy of the confusion matrix and kappa value. Datasets with potable groundwater were divided into training sets (80%) and test sets (20%; Fig 2B). The training sets were then further divided into two parts: 80% training set and 20% validation set. The goal of these procedures was to avoid overfitting issues during the modelling process.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences Discussions

**2.5 Feature selection**

Because the groundwater quality strongly correlates with hydrogeological properties, several water quality parameter may be biased in certain areas and the water quality predictions based on the classification model may proceed with low accuracy. To better understand the effect of the water quality parameters on the data-driven model, the input feature selection was conducted by RF. The RF can be used to extract the feature importance based on how much each feature contributes to decreasing the impurity of the trees ('meandecrease gini'; MDG) (Han et al., 2016). This parameter can be used to rank the different features.

**2.6 Binning and 2D spatial analysis**

Binning is a method that is used to group a number of continuous values into a smaller number of bins. We used binning to group multiple GQIs into one value of a grid of uniform size in the map. Based on this method, a representative GQI is obtained for a region and problems caused by outliers in the closed area can be alleviated. The binning was calculated and visualised using the 'stat_summary_2d' function of the R package 'ggplot2' (version 3.3.3). When analysing with low resolution using binning on national maps, the size of the bin was set to 9 km × 11 km (longitude × latitude). When analysing with high resolution using binning for Chungcheongbuk-do, the size of the bin was set to 4.5 km × 5.5 km (longitude × latitude), representing one fourth of the nationwide bin size. Each bin was coloured based on the GQI.

To compare changes in the long-term groundwater quality based on the GQI, an area with a worrisome (site A: Seangkeuk in Eumseong), good (site B: Gageum in Chungju), and very good (site C: Heoin in Boeun) GQI was selected. Water quality (nitrate and chloride concentrations) datasets were collected from the national monitoring wells in the above-mentioned area. Nitrates and chlorides were selected because they overlap with the water quality parameters monitored at the national groundwater monitoring wells and water quality parameters used in this study.

## 2.7 Statistical analysis

All statistical analysis were conducted using R software (version 3.6.1). To examine the differences in the GQIs and water qualities of GQI grades, Kruskal–Wallis analysis was used according to the normality of data. The significant differences between the GQI grades was further confirmed using the 'mctp' function in the R package 'nparcomp' (version 3.0) as a nonparametric post-hoc method.

# 3 Results

## 3.1 Pollution characteristics of groundwater

Based on Korean drinking water quality standards, the water quality of 65.2% (1,344 wells) of 2,061 wells in 2017; 64.3% (1,377 wells) of 2,142 wells in 2018; 46.0% (928 wells) of 2,019 wells in 2019; and 53.5% (1,125 wells) of 2,104 wells in 2020 was inappropriate (Fig. S1A). The major sources of groundwater pollution were microorganisms (42.4%–45.9%), followed by complex pollution (24.6%–35.0%), harmful inorganics (16.1%–27.1%), substances with an aesthetic effect (2.4%–6.5%), and harmful organics (one well in 2018; Fig. S1B). The complex pollution containing microorganisms or nitrate accounted for 84.5% of the total proportion. Considering these results, microorganisms and harmful inorganics were major groundwater pollutants in the study area. However, it was difficult to determine specific external factors (e.g., land use and well

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


development year) causing the groundwater pollution (Fig. S2).

### 3.2 Characterization of the GQI and grades

The distribution of the GQI calculated in each well has been visualised in a bar graph (Fig. 3A). From the figure, a right-skewed distribution can be observed for all potable groundwater. The minimum, maximum, median, and average values of the GQI were 0.7344, 0.9770, 0.9160, and 0.9127, respectively. These results indicate that the water quality of more than half of the wells was on average within 10% of the water quality standard. Note that the GQI correlates with the Weighted Arithmetic WQI, one of the well-known single indices for evaluating the surface water quality (cor = -0.38, Fig. S3). The GQI was divided into three grades: 'worrisome', <0.89 (714 wells); 'good', 0.89–0.94 (2,229 wells); and 'very good', >0.94 (609 wells; solid line in Fig. 3A). The GQI significantly varied depending on the grade (Kruskal test, p-value < 0.05, Fig. 3B).

To determine the factors that control the GQI of the wells of each grade, the water quality parameters were analysed using two approaches. First, the number of water quality parameters higher than half of the water quality standards was calculated for each groundwater well (Fig. 3C). In the 'very good' grade, one or less parameters accounted for more than 95% of the total, whereas one or two parameters accounted for more than 95% of the total in the 'good' grade. In the 'worrisome' grade, more than 50% of the water quality standards of one or more parameters was observed in all wells, and three or more water quality parameters accounted for more than 50% of the wells. As expected, when the grade changed from 'very good' to 'worrisome', many water quality parameters approached the water quality standards. Second, the GQIs of the water quality parameters contributing to the grade division were statistically compared based on the grade (Fig.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


4). Based on the selection of the ten most important water quality parameters using the RF model, the most important parameter was 'general bacteria', followed by the turbidity, nitrate, total hardness, sulphate, chloride, zinc, potassium permanganate consumption, fluoride, and iron (Fig. 3D). The deviation value of all selected parameters significantly decreased from 'very good' to 'worrisome' (Kruskal test, p-value < 0.05, Fig. 4). These results imply that various water quality parameters were close to the water quality standards in the 'worrisome' grade.

### 3.3 Data-driven model selection

The performance of ten classification models for the prediction of the GQI was compared based on the accuracy of the confusion matrix and kappa value (Fig. 5). The ANN model yielded the best classification performance (96.5%–98.6%), followed by the SVM with an average classification accuracy of ≥90%. The average classification accuracy of Naïve Bayes and decision tree models did not exceed 60%. Therefore, the ANN model with a 98.6% classification accuracy was selected as the optimal classification model. We also applied the ANN model to predict the grades using individual annual datasets for additional cross-validation. The grades of all annual datasets were predicted with an accuracy of ~99%.

### 3.4 Spatial analysis using GQI binning

Binning is useful for the conversion of large point-based data to a regular grid representing the aggregation of points in the map and makes it easy to visualise the data at different map scales. We binned the GQI and grades and plotted them on a nationwide map (each grid: 9 km × 11 km) for South Korea (Figs S5A and S5B). Among all 1,496 grid cells, 537 (35.9% of total) grid cells

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


indicate the GQI and grades, and on average seven wells were included based on the coloured grid cells. The Chuncheongbuk-do province was selected for the visualization and analysis of the GQI and grades at a higher resolution by using half of the size of the previous grid (each grid: 4.5 km × 5.5 km; Figs S5C, S5D, and 6A). Compared with other provinces, the wells are evenly spatially distributed in the Chuncheongbuk-do Province. Among all 460 grid cells, 80 (17.4% of total) grid cells indicate the GQI and grades. Among the coloured grid cells, 16 (20.0% of coloured grid), 55 (68.7%), and nine (11.3%) grid cells represent a 'worrisome', 'good', and 'very good' grade, respectively. In general, the grid cells representing 'worrisome' areas are mainly distributed at the edge of the province. We compared the long-term trends of the water quality of national monitoring wells by selecting a representative region for each grade to confirm that our results are reliable (Fig. 6A). Because the main water quality parameters used in this study and those regularly monitored by the national groundwater monitoring wells are nitrate and chloride, the change of the two parameters over ten years was analysed (Fig. 6B). Both the nitrate ($R^2$ = 0.289) and chlorine ($R^2$ = 0.696) concentrations at site A (Saengkeuk in Eumseong), which was determined to be a 'worrisome' area, have rapidly increased over the past decade. Both the nitrate ($R^2$ = -0.413) and chloride ($R^2$ = -0.05) concentrations show a decreasing trend at site C; however, at site B, the chloride content slightly increased ($R^2$ = 0.149), but the nitrate concentration did not change ($R^2$ = -0.02). The differences in the long-term water quality trends observed at the national groundwater monitoring wells were confirmed based on the GQI grades.

## 4 Discussion

The 'Safe Groundwater Project in Unsupplied Areas (2017–2020)' was conducted including wells

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


used by citizens for which regular water quality surveys were not carried out. It is a public service to provide realistic policies to citizens based on the analysis of the quality of groundwater used by citizens. In contrast to previous reports of massive water quality monitoring in South Korea, which mainly included a limited number of water quality parameters monitored in national groundwater monitoring wells, the data from the 'Safe Groundwater Project in Unsupplied Areas (2017–2020)' are of great value because they include 47 water quality parameters monitored in wells that citizens use as drinking water sources. More than 50% of the wells used in this study are inappropriate as drinking water source (at least one water quality parameter exceeded the standard value). Considering that the ratio of wells with an inappropriate water quality was low (6.5%–8.0%) based on a previous massive survey of the groundwater quality, the number of wells with an inappropriate water quality obtained in this study is very high (Lee and Kwon, 2016). The analysed water quality parameters are significantly higher and various water quality parameters exceed the thresholds; thus, the proportion of wells with an inappropriate water quality has increased. However, because both studies showed that the main sources of groundwater pollution are nitrate and total coliform, it is impossible to simply explain the cause of the increase in the proportion of inappropriate wells with the increase in the number of analysed water quality parameters (Yun et al., 2014). Because the characteristics of land use or well development insignificantly affect the proportion of inappropriate wells, it is necessary to analyse these data with a new approach to link them to groundwater management policies. In particular, it is necessary to develop an approach by selection of areas that require groundwater management based on potable water quality data.

We calculated the GQI, a single index of the water quality of each groundwater well, and divided it into three grades. This approach is similar to the WQI used for surface water (WHO, 2011). In

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


particular, the GQI is similar to the WQI suggested by Brown in that it is calculated as a single index using the difference between the water quality standard and observed value (Seifi et al., 2020). Because the GQI calculated in this study does not include a parameter selection process nor a weight determination process for the parameter, the process for calculating the GQI is very simple and an index bias can be avoided. The correlation between the Weighted Arithmetic WQI and GQI indicates the similarity between the two indices. However, in wells with a good water quality (low WAWQI value), the GQI can be analysed with higher resolution compared with the Weighted Arithmetic WQI (high variability of the GQI in the same Weighted Arithmetic WQI). This indicates that the GQI is more suitable for evaluating the water quality of potable groundwater. Despite the calculation for potable groundwater, the statistical differences in major water quality parameters based on the GQI grade confirm the usefulness of the GQI. In addition, it was confirmed that the difference in the number of parameters with a distance score from the standard of less than 50% depends on the grade. This means that, because the GQI is calculated in consideration of the effects of complex parameters rather than one parameter, it is suitable for Korea, which is facing significant groundwater pollution caused by complex pollutants. Pollution sources change several groundwater quality parameters at the same time. Therefore, considering multiple parameters at the same time is more important in tracking pollutants and characterising groundwater pollution conditions than focusing on one parameter (Menció et al., 2016). Because the GQI has the same effect as multivariate analysis based on the simultaneous reflection of changes of multiple water quality parameters, it is of advantage in water quality surveys for both complex and single pollutants (Menció et al., 2016; Wu et al., 2019).

To develop a data-driven model that can be used to predict the GQI grade, we compared the

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions

predictive performance of ten classification models. The ANN performed the best, with a prediction accuracy of ~95%, followed by the SVM and bagged flexible discriminant analysis. The classification performance of the ANN is high because this ensemble method uses the average of the predictions from each model by fitting multiple neural network models to the same datasets (Lone et al., 2021). Based on previous studies, the ANN did not have a good prediction performance compared with other models (Ehteshami et al., 2016; Naghibi et al., 2018). However, in most previous studies, ANNs were used for regression analysis to predict specific water quality indicators rather than grades or classes, and the amount of data was considerably small for ANN analysis (Ehteshami et al., 2016). In this study, a high performance was obtained because the water quality parameters were divided into grades and the amount of learning data was suitable for ANN applications.

Because groundwater pollution will likely quickly contaminate the groundwater in surrounding areas depending on the flow rate and hydrogeological characteristics, it is important to investigate the groundwater quality of multiple wells at the regional scale to mitigate groundwater pollution in a timely manner. However, because it is difficult to proceed for realistic reasons (e.g., well selection, sampling cycle, and analysis cost), datasets are often converted into data suitable for grid cell or spatial indices, and geographic areas are partitioned by using an analysis technique such as binning and displaying in maps (Shrestha et al., 2015). We recalculated the GQI by using grid cells and binning and determined GQI grades and visualised them on a map. Although it does not completely match the existing administrative districts, the map indicates the current status of the groundwater quality for Myun-sized (the second smallest administrative unit in South Korea) grid cells (4.5 km × 5.5 km). In the case of Chuncheonbuk-do, GQI grades were determined in

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


only 17.5% of the province because areas with an abundant water supply and unoccupied mountainous areas are not subject to water quality monitoring. These simple visualization results are more intuitive and user-friendly than existing groundwater pollution vulnerability assessments or data-driven groundwater quality models (Knoll et al., 2019; Ouedraogo et al., 2016). In addition, the data can be easily linked to groundwater management policies because water quality management agencies or local governments can utilise existing groundwater quality data without further monitoring processes. Because the water quality of multiple wells is converged to an average value during binning, it is necessary to evaluate the reliability of the binned GQI or GQI grade (Kumar and Krishna, 2018). Reliability assessment should be used to track the groundwater contamination by investigating the water quality of local wells by grade in the long term. However, it is difficult to establish a general reliability assessment because a water quality monitoring had to be conducted only once in one groundwater well in this project. To indirectly evaluate the reliability, we analysed the trend of the water quality changes over the past ten years based on the GQI grades using the results of long-term water quality monitoring at the national groundwater monitoring wells. The water quality results significantly differ depending on the GQI grades. In particular, in the 'worrisome' grade, the contamination of groundwater has rapidly increased over the past decade, which was not observed for other grades. These results indicate that the reliability of the water quality evaluation is high, even if the GQI, which is sensitive to the changes in multiple water quality parameters, is binned and analysed at a regional level. In addition, spatial analysis including the GQI grade can provide important information for establishing policies with respect to the selection of groundwater management priority areas.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions

## 5 Conclusion

A method to evaluate the groundwater quality for drinking purposes based on the GQI was introduced. The regional characteristics of the GQI were assessed using 2D spatial analysis. Because the GQI was computed based on mass water quality data (47 water quality parameters and 8,326 wells) and neural networks, the groundwater quality could be accurately determined. Overall, the results show that the groundwater in a large number of wells that are currently used by citizens as drinking water sources, especially in regions with low GQI (e.g., 20.1% for the 'worrisome' grade), is polluted. This approach can be used to predict potential groundwater pollution based on the comprehensive evaluation of the groundwater quality beyond the dichotomous judgement of the drinking water quality based on the water quality standard. In this study, a GQI was developed based on several water quality parameters, and it only partially reflects the water quality characteristics of the groundwater. More parameters, including hydrogeological, meteorological, and land use parameters, should be added to improve the GQI and effectiveness of groundwater management and risk assessment.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.



**ACKNOWLEDGEMENT**

We gratefully acknowledge the funding of the projects of the National Institute of Environmental Research (NIER) in South Korea (NIER-2020-04-02-049; NIER-2021-04-02-092) by the Korean Ministry of Environment (MOE).

**AUTHOR CONTRIBUTIONS**

**Seok Hyun Ahn:** Data curation, Data analysis, Visualization, Writing-original draft. **Tae Kwon Lee:** Conceptualization, Data analysis, Writing-review & editing. **Do Hwan Jeong:** Validation, Writing-review & editing, **Moonsu Kim:** Conceptualization, Writing-review & editing. **Hyun-Koo Kim:** Conceptualization, Supervision, Writing-review & editing

**COMPETING INTERESTS**

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.

**Code/Data availability**

The data for this project are confidential, but may be obtained with Data Use Agreements with the National Institute of Environmental Research (NIER) in South Korea. Researchers interested in access to the data may contact Do Hwan Jeong at jungdh93@korea.kr.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


## Reference

Abbasnia, A., Yousefi, N., Mahvi, A. H., Nabizadeh, R., Radfard, M., Yousefi, M., and Alimohammadi, M.: Evaluation of groundwater quality using water quality index and its suitability for assessing water for drinking and irrigation purposes: Case study of Sistan and Baluchistan province (Iran), Human and Ecological Risk Assessment: An International Journal, 25, 988-1005, 2019.

Antanasijević, D., Pocajt, V., Perić-Grujić, A., and Ristić, M.: Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis, Journal of Hydrology, 519, 1895-1907, 2014.

Bhanja, S. N., Mukherjee, A., Rodell, M., Wada, Y., Chattopadhyay, S., Velicogna, I., Pangaluru, K., and Famiglietti, J. S.: Groundwater rejuvenation in parts of India influenced by water-policy change implementation, Scientific Reports, 7, 7453, 2017.

Brown, R., Mccleiland, N., Deiniger, R., and O'Connor, M.: Water quality index-crossing the physical barrier,(Jenkis, SH, ed.) Proc, 1972, 787-797.

Brown, R. M., McClelland, N. I., Deininger, R. A., and Tozer, R. G.: A water quality index-do we dare, Water and sewage works, 117, 1970.

Centre, I. G. R. A.: Groundwater Overview: Making the invisible Visible, UN WATER, 2018. 2018.

Cude, C. G.: Oregon water quality index a tool for evaluating water quality management effectiveness 1, JAWRA Journal of the American Water Resources Association, 37, 125-137, 2001.

Ehteshami, M., Farahani, N. D., and Tavassoli, S.: Simulation of nitrate contamination in groundwater using artificial neural networks, Modeling Earth Systems and Environment, 2, 28, 2016.

Gao, Y., Qian, H., Ren, W., Wang, H., Liu, F., and Yang, F.: Hydrogeochemical characterization and quality assessment of groundwater based on integrated-weight water quality index in a concentrated urban area, Journal of Cleaner Production, 260, 121006, 2020.

Guppy, L., Uyttendaele, P., Villholth, K. G., and Smakhtin, V.: Groundwater and sustainable development goals: Analysis of interlinkages, 2018.

Han, H., Guo, X., and Yu, H.: Variable selection using mean decrease accuracy and mean decrease gini based on random forest, 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 219-224, 2016.

Jha, D. K., Devi, M. P., Vidyalakshmi, R., Brindha, B., Vinithkumar, N. V., and Kirubagaran, R.: Water quality assessment using water quality index and geographical information system methods in the coastal waters of Andaman Sea, India, Marine pollution bulletin, 100, 555-561, 2015.

Knoll, L., Breuer, L., and Bach, M.: Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning, Science of The Total Environment, 668, 1317-1327, 2019.

Kuhn, M.: Building predictive models in R using the caret package, Journal of statistical software, 28, 1-26, 2008.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.


Kumar, A. and Krishna, A. P.: Assessment of groundwater potential zones in coal mining impacted hard-rock terrain of India by integrating geospatial and analytic hierarchy process (AHP) approach, Geocarto International, 33, 105-129, 2018.

Lapworth, D. J., Lopez, B., Laabs, V., Kozel, R., Wolter, R., Ward, R., Vargas Amelin, E., Besien, T., Claessens, J., Delloye, F., Ferretti, E., and Grath, J.: Developing a groundwater watch list for substances of emerging concern: a European perspective, Environmental Research Letters, 14, 035004, 2019.

Lee, J.-Y. and Kwon, K. D.: Current status of groundwater monitoring networks in Korea, Water, 8, 168, 2016.

Li, H., Gu, J., Hanif, A., Dhanasekar, A., and Carlson, K.: Quantitative decision making for a groundwater monitoring and subsurface contamination early warning network, Science of The Total Environment, 683, 498-507, 2019.

Lone, K. J., Hussain, L., Saeed, S., Aslam, A., Maqbool, A., and Butt, F. M.: Detecting basic human activities and postural transition using robust machine learning techniques by applying dimensionality reduction methods, Waves in Random and Complex Media, 2021. 1-26, 2021.

Lumb, A., Halliwell, D., and Sharma, T.: Application of CCME Water Quality Index to monitor water quality: A case study of the Mackenzie River basin, Canada, Environmental Monitoring and assessment, 113, 411-429, 2006.

Menció, A., Mas-Pla, J., Otero, N., Regàs, O., Boy-Roura, M., Puig, R., Bach, J., Domènech, C., Zamorano, M., and Brusi, D.: Nitrate pollution of groundwater; all right…, but nothing else?, Science of the total environment, 539, 241-251, 2016.

Meyers, G., Kapelan, Z., and Keedwell, E.: Short-term forecasting of turbidity in trunk main networks, Water Research, 124, 67-76, 2017.

Misaghi, F., Delgosha, F., Razzaghmanesh, M., and Myers, B.: Introducing a water quality index for assessing water for irrigation purposes: A case study of the Ghezel Ozan River, Science of the Total Environment, 589, 107-116, 2017.

Naghibi, S. A., Pourghasemi, H. R., and Abbaspour, K.: A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS, Theoretical and applied climatology, 131, 967-984, 2018.

Najah Ahmed, A., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., Ehteram, M., and Elshafie, A.: Machine learning methods for better water quality prediction, Journal of Hydrology, 578, 124084, 2019.

Ouedraogo, I., Defourny, P., and Vanclooster, M.: Mapping the groundwater vulnerability for pollution at the pan African scale, Science of The Total Environment, 544, 939-953, 2016.

Patro, S. and Sahu, K. K.: Normalization: A preprocessing stage, arXiv preprint arXiv:1503.06462, 2015. 2015.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions

Seifi, A., Dehghani, M., and Singh, V. P.: Uncertainty analysis of water quality index (WQI) for groundwater quality evaluation: Application of Monte-Carlo method for weight allocation, Ecological Indicators, 117, 106653, 2020.

Shrestha, P., Sulis, M., Simmer, C., and Kollet, S.: Impacts of grid resolution on surface energy fluxes simulated with an integrated surface-groundwater flow model, Hydrol. Earth Syst. Sci., 19, 4317-4326, 2015.

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., and Zhao, X.: A review of data-driven approaches for prediction and classification of building energy consumption, Renewable and Sustainable Energy Reviews, 82, 1027-1047, 2018.

WHO, G.: Guidelines for drinking-water quality, World Health Organization, 216, 303-304, 2011.

Wu, J., Li, P., Wang, D., Ren, X., and Wei, M.: Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau, Human and Ecological Risk Assessment: An International Journal, 2019. 2019.

Yun, S. W., Choi, H.-M., and Lee, J.-Y.: Comparison of groundwater levels and groundwater qualities in six megacities of Korea, Journal of the Geological Society of Korea, 50, 517-528, 2014.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Open Access
Discussions

Figure list

Figure 1. Geographical information about the sampling sites in Korea.

Figure 2. (A) Preprocessing for the calculation of the groundwater quality index including 28 water quality parameters and establishment of grades for data-driven modelling. (B) Workflow of the data-driven model.

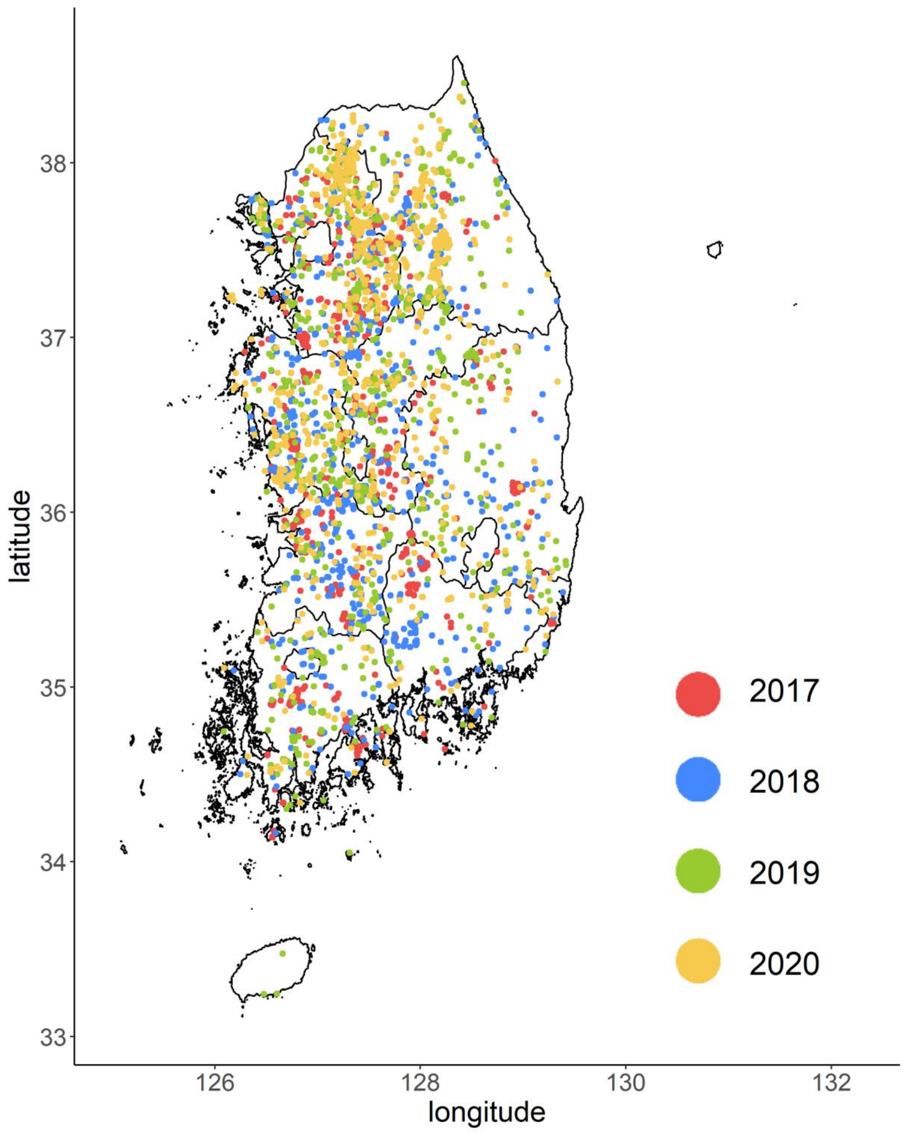
Figure 3. (A) Visualization of the distribution based on the calculated groundwater quality index and grades (grey vertical line). (B) Comparison of the groundwater quality indices for the grades. Significant differences between grades are marked by lowercase letters. (C) Proportion of each grade to the number of water quality parameters with a deviation value of 0.5 or less for each well. (D) Selection of the top ten features contributing to the model performance using random forest.

Figure 4. Boxplot of the deviation value for the top ten features contributing to the model performance. (A) General Bacteria, (B) Nitrate, (C) Turbidity, (D) Total hardness, (E) Chloride, (F) Sulphate, (G) Potassium permanganate consumption, (H) Zinc, (I) Fluorine, and (J) Iron. The significance of the grades was calculated with the Kruskal–Wallis test. Significant differences ($P < 0.05$) between grades are marked by lowercase letters.

Figure 5. Comparison of the data-driven model performance using the accuracy of the confusion matrix and kappa value. Each classification model includes a five-fold cross validation, with ten repeated values.

Figure 6. (A) The grades of the binned area for Chungcheongbuk-do are visualised in the map with three sites representative for the grades. Site A: Saengkeuk in Eumseong, Site B: Gageum in Chungju, Site C: Heoin in Boeun. (B) Changes in the nitrate and chloride concentrations in the last ten years measured at national ground monitoring wells in three representative regions.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences
Open Access
Discussions




528
529 Figure 1. Geographical information about the sampling sites in Korea.
530

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

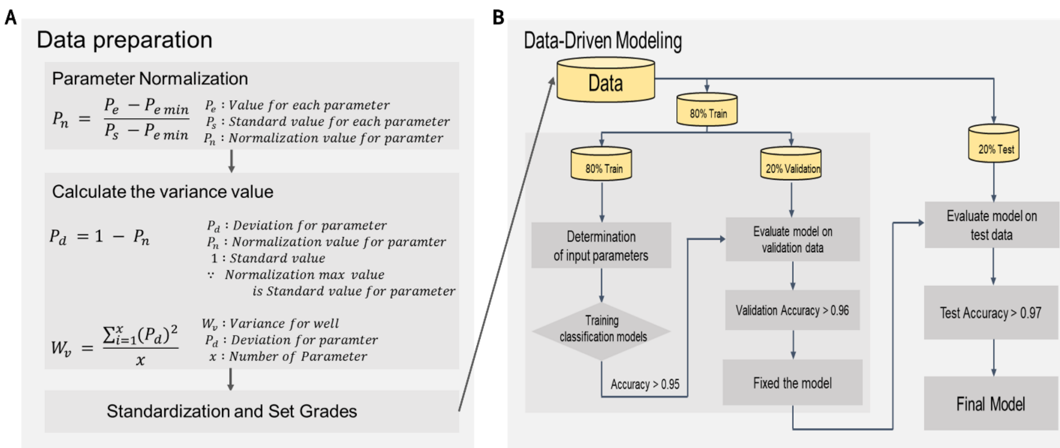



Figure 2. (A) Preprocessing for the calculation of the groundwater quality index including 28 water quality parameters and establishment of grades for data-driven modelling. (B) Workflow of the data-driven model.

© Author(s) 2022. CC BY 4.0 License.

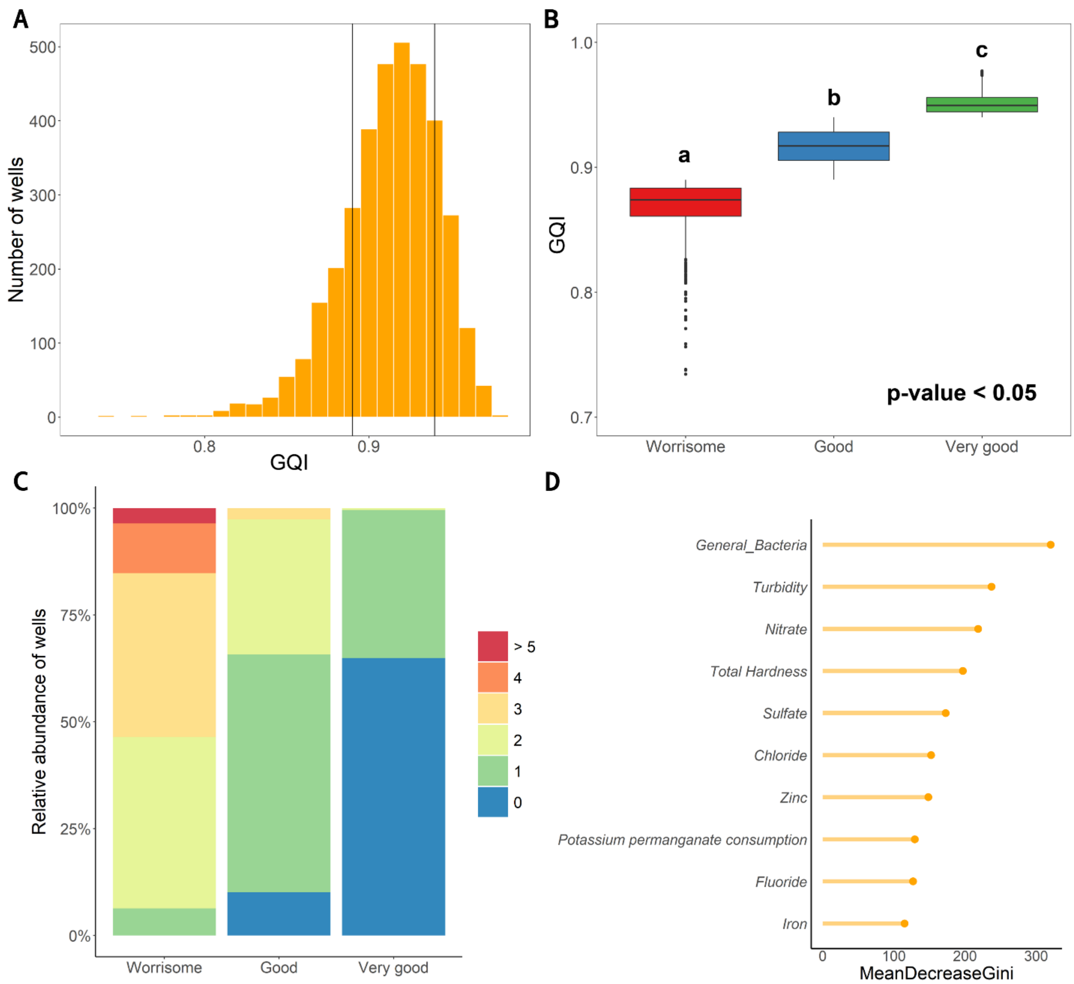




Figure 3. (A) Visualization of the distribution based on the calculated groundwater quality index and grades (grey vertical line). (B) Comparison of the groundwater quality indices for the grades. Significant differences between grades are marked by lowercase letters. (C) Proportion of each grade to the number of water quality parameters with a deviation value of 0.5 or less for each well. (D) Selection of the top ten features contributing to the model performance using random forest.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.

Hydrology and Earth System Sciences
Discussions

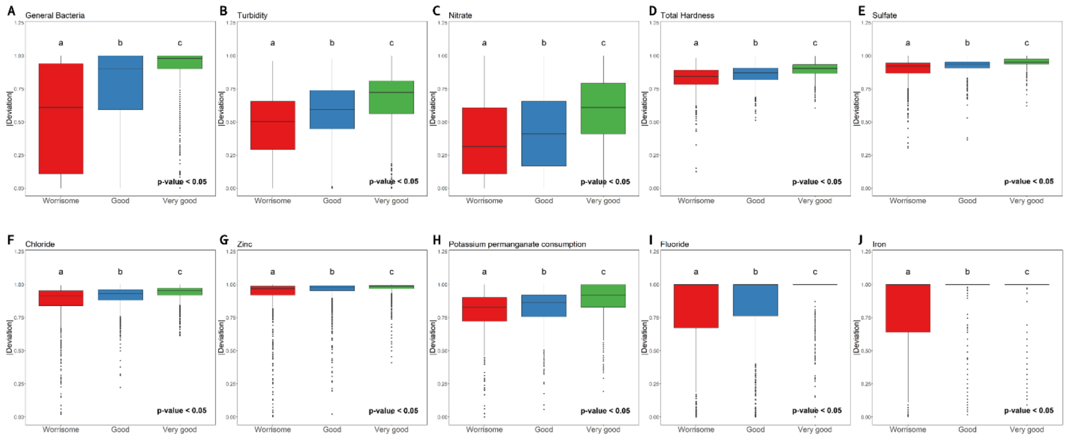



Figure 4. Boxplot of the deviation value for the top ten features contributing to the model performance. (A) General Bacteria, (B) Nitrate, (C) Turbidity, (D) Total hardness, (E) Chloride, (F) Sulphate, (G) Potassium permanganate consumption, (H) Zinc, (I) Fluorine, and (J) Iron. The significance of the grades was calculated with the Kruskal–Wallis test. Significant differences ($P < 0.05$) between grades are marked by lowercase letters.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.



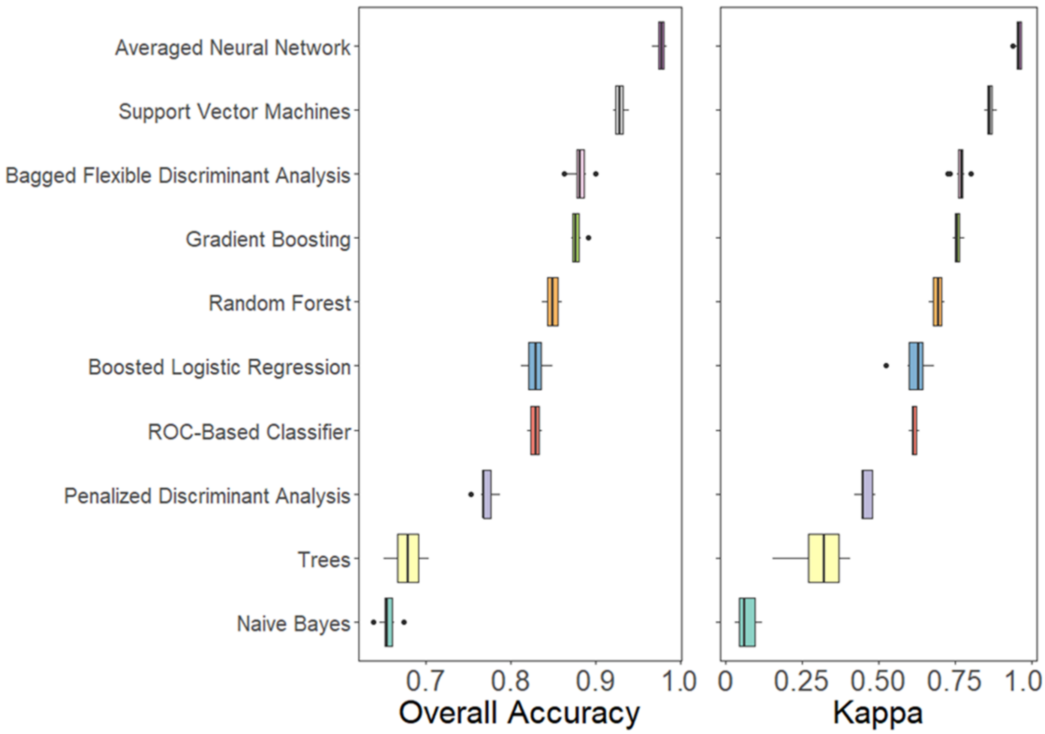



Figure 5. Comparison of the data-driven model performance using the accuracy of the confusion matrix and kappa value. Each classification model includes a five-fold cross validation, with ten repeated values.

https://doi.org/10.5194/hess-2022-86
Preprint. Discussion started: 1 June 2022
© Author(s) 2022. CC BY 4.0 License.
Hydrology and Earth System Sciences
Discussions

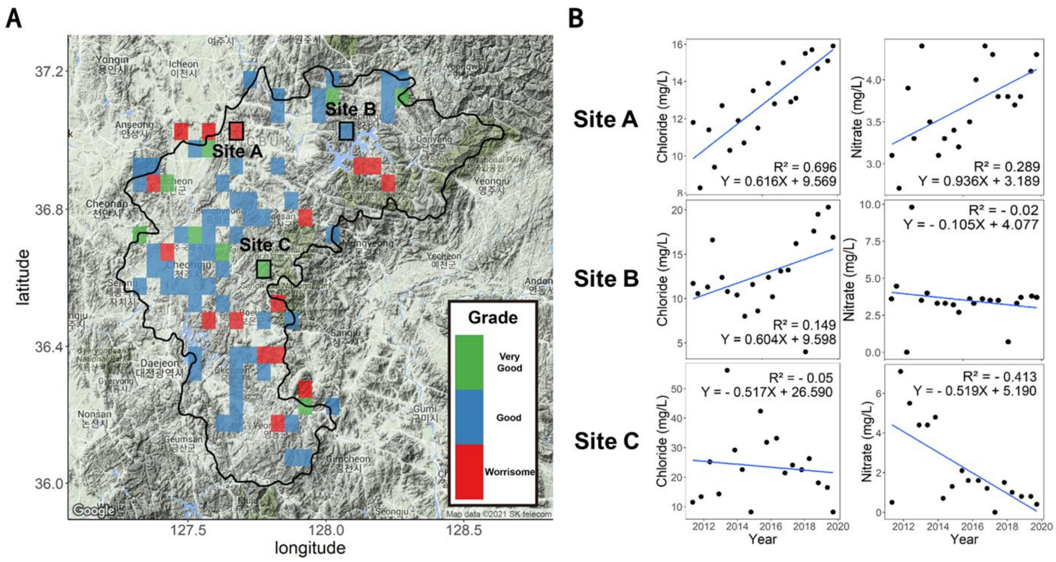



Figure 6. (A) The grades of the binned area for Chungcheongbuk-do are visualised in the map with three sites representative for the grades. Site A: Saengkeuk in Eumseong, Site B: Gageum in Chungju, Site C: Heoin in Boeun. (B) Changes in the nitrate and chloride concentrations in the last ten years measured at national ground monitoring wells in three representative regions.