

Groundwater productivity potential mapping using frequency ratio and evidential belief function and artificial neural network models: focus on topographic factors

Jeong-Cheol Kim, Hyung-Sup Jung and Saro Lee

ABSTRACT

This study analysed groundwater productivity potential (GPP) using three different models in a geographic information system (GIS) for Okcheon city, Korea. Specifically, we have used variety topography factors in this study. The models were based on relationships between groundwater productivity (for specific capacity (SPC) and transmissivity (T)) and hydrogeological factors. Topography, geology, lineament, land-use and soil data were first collected, processed and entered into the spatial database. T and SPC data were collected from 86 well locations. The resulting GPP map has been validated in under the curve analysis area using well data not used for model training. The GPP maps using artificial neural network (ANN), frequency ratio (FR) and evidential belief function (EBF) models for T had accuracies of 82.19%, 81.15% and 80.40%, respectively. Similarly, the ANN, FR and EBF models for SPC had accuracies of 81.67%, 81.36% and 79.89%, respectively. The results illustrate that ANN models can be useful for the development of groundwater resources.

Key words | artificial neural network, frequency ratio, GIS, groundwater productivity potential, South Korea

Jeong-Cheol Kim

National Institute of Ecology (NIE),
1210 Geumgang-ro, Maseo-myeon, Seocheon-gun,
Chungcheongnam-do 33657,
Korea

Jeong-Cheol Kim

Hyung-Sup Jung
Department of Geoinformatics,
University of Seoul,
163 Seoulsiripdaero, Dongdaemun-gu, Seoul
02504,
Korea

Saro Lee (corresponding author)

Geological Research Division, Korea Institute of
Geoscience and Mineral Resources (KIGAM),
124, Gwahak-ro Yuseong-gu, Daejeon 34132,
Korea
and
Korea University of Science and Technology,
217 Gajeong-ro Yuseong-gu, Daejeon 305-350,
Korea
E-mail: leesaro@kigam.re.kr

INTRODUCTION

Water resources, including surface water and groundwater, are recycled through evaporation, precipitation and runoff on the Earth's surface. The circulation of water depends on the timing and spatial patterns of precipitation and evaporation. These climatological aspects determine the spill pattern, runoff time and water availability (WWDR 2017). Over the past decades, climate observations and climate change projections have shown increased spatial and temporal heterogeneity in the water cycle. As a result, discrepancies in water supply and demand are increasing (IPCC 2013). In 2017, a lack of spring rain triggered historic droughts in Korea. This resulted in a shortage of drinking water and economic damage due to reduced productivity of agricultural land.

doi: 10.2166/hydro.2018.120

The demand for water resources, including groundwater, is expected to increase significantly over the next few decades (Mogaji & Lim 2018). For the past five years, the World Economic Forum (WEF) has stated that the water crisis is one of the most pressing global risks. In 2016, the WEF stated that this crisis will have the greatest impact on societies in the coming ten years (WEF 2016). Currently, around 20% of human consumptive use globally is generated from groundwater sources, and this is projected to increase rapidly over time. Research on and development of available groundwater can help reduce economic losses affiliated with this global crisis.

Groundwater is one of the important natural resources used in agriculture, in the industrial sector, and for public

water supply (Ganapuram *et al.* 2016). Thus, efforts to locate high-quality groundwater sources are increasing worldwide. In Korea, the use of groundwater increased by more than 225% between 1994 and 2014, and the current national supply of groundwater no longer meets the needs of society (MLTM 2016). Therefore, reliable analytical models predicting locations of groundwater from GPP mapping are needed for efficient management of groundwater use.

Groundwater productivity potential (GPP) mapping is defined as a map estimating the probability that groundwater will occur in a study area. Generally speaking, estimating the groundwater potential over an area involves the use of an objective statistical analysis based on various types and sources of field data. However, developing the requisite field data set over a large area involves significant cost and time (Lee & Talib 2005). A geographic information system (GIS) can be used to detail large areas in a more cost-effective manner (Oh *et al.* 2011; Kim *et al.* 2014, 2016, 2017; Lee *et al.* 2016a). Over the past few years there have been significant improvements in GIS technology (Lee *et al.* 2016b; Hwang *et al.* 2017; Kim & Jung 2017; Oh *et al.* 2017), and various spatial modelling techniques have been developed to evaluate groundwater potential. In practice, it is not possible to drill any groundwater wells, so by applying machine learning and statistical methodologies using GIS techniques, it is possible to analyse the relationship between groundwater and other topographic or geographical properties based on the learned data; as a result, the percentage of the estimated potential groundwater well is shown on the map. According to the results, it can reduce the cost and time required for groundwater detection.

As the use of GIS-based models has increased, various models have been proposed to assess GPP. Recently, various data mining models have been applied in many studies, including frequency ratio (FR) (Elmahdy & Mohamed 2015; Jothibasu & Anbazhagan 2017), artificial neural network (ANN) (Lee *et al.* 2006, 2017b; Sokeng *et al.* 2016), random forest (RF) (Rahmati *et al.* 2016; Zabihi *et al.* 2016), logistic regression (LR) (Ozdemir 2011b; Zandi *et al.* 2016; Park *et al.* 2017), boosted regression tree (BRT) (Kim *et al.* 2018; Lee *et al.* 2017a; Mousavi *et al.* 2017), support vector machine (SVM) (Lee *et al.* 2018), weights of evidence (WoE) (Tahmassebpour *et al.* 2016; Ghorbani Nejad *et al.*

2017), and evidential belief function (EBF) (Pourghasemi & Beheshtirad 2015; Zeinivand & Ghorbani Nejad 2018), and have been applied in GPP mapping, as well.

Especially for GPP mapping and other disaster susceptibility mapping using the ANN or FR model, Lee *et al.* (2017b) compared the predictive ability of ANN and SVM models, and the results show that the ANN model is more suitable for landslide susceptibility mapping. Lee *et al.* (2018) carried out an analysis using the data mining models of the ANN and a SVM in Boryeong city, Korea. Naghibi *et al.* (2017) compared the predictive ability of RF, ANN, FR, SVM, and boosted tree models, etc. Mousavi *et al.* (2017) compared the predictive ability of boosted tree with that of FR models. Also, a comparison of statistical index (SI), FR, WoE and EBF techniques for finding the groundwater potential area was performed by Zeinivand & Ghorbani Nejad (2018). Lee *et al.* (2012) applied the mapping of regional GPP for the area around Pohang city, Korea. Ozdemir (2011a) explored GIS-based GPP maps in the Sultan Mountains (Konya, Turkey) using FR, WoE and LR methods and compared them.

This study applied ANN (data mining) and FR and EBF (probabilistic) models to GPP mapping to obtain more accurate GPP maps, and identified important factors within each model that affect the productivity potential of groundwater. Several prior studies have generated GPP maps using various data mining techniques, but there have been few attempts to create GPP maps using an ANN model. Thus, this study also confirmed and compared the accuracy and adequacy of the ANN and FR and EBF models for GPP analysis. Therefore, this study can be used as a baseline reference for future work with ANN models. The productivity of groundwater is affected by various factors including hydraulic conductivity. Among the various factors, geomorphic factors that are related to the concentration and flow of water by altitude differences have a great impact on groundwater productivity. In this study, it was used in terms of topographic factors because we aimed to confirm whether the potential productivity of groundwater can be estimated with high accuracy when using various geographical factors which make it relatively easy to acquire data.

The process of GPP map analysis is displayed in Figure 1. T and SPC point data were obtained and randomly

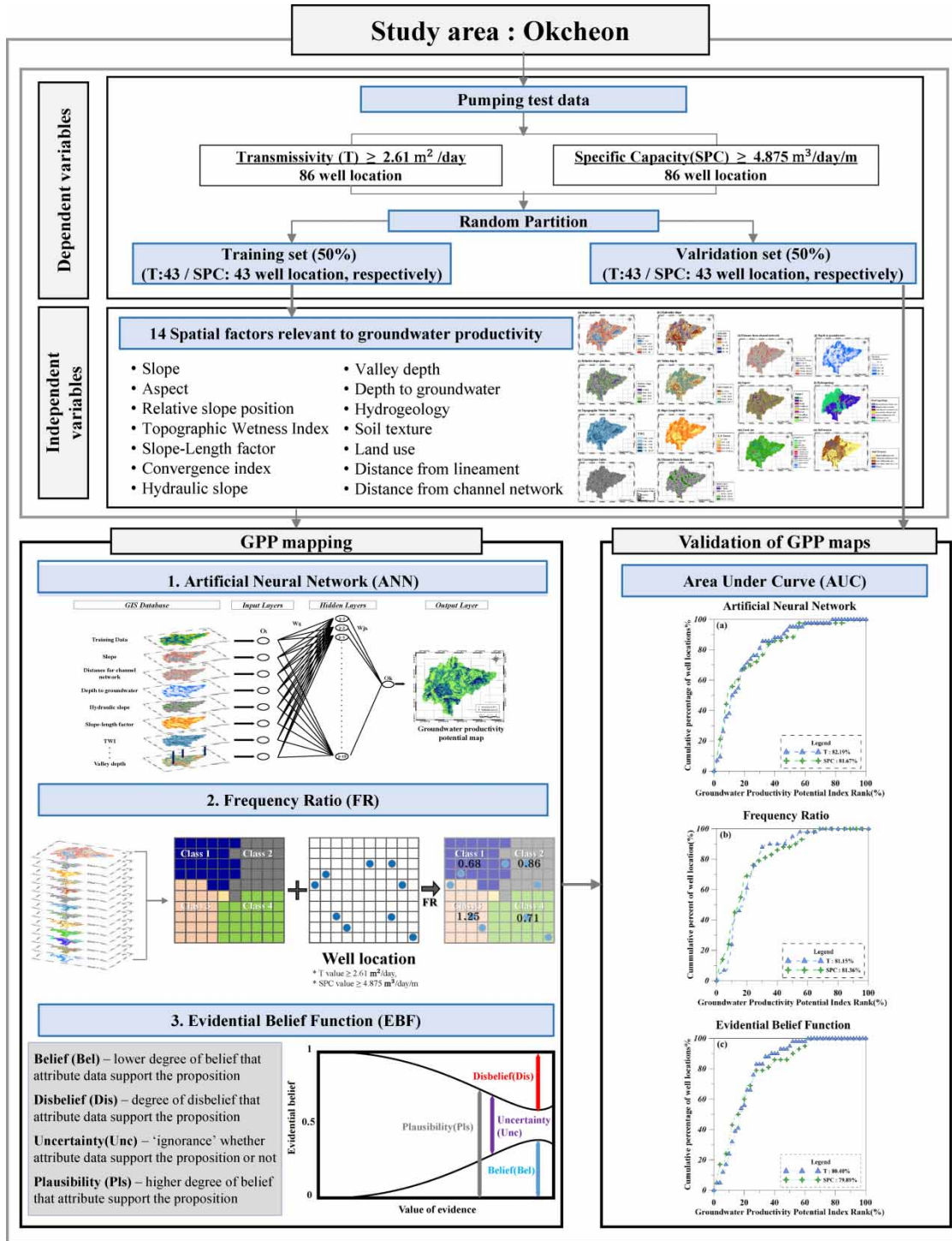


Figure 1 | Flowchart of the study procedures for groundwater productivity potential mapping.

classified as either training data (50%) or validation data (50%). Geology, topography, soil texture and land cover data were combined into a spatial database (Hou et al.

2018; Jenifer & Jha 2017). Hydrogeological factors, including slope, aspect, slope gradient, relative slope position, hydraulic slope, valley depth, topographic wetness index (TWI,

slope length (LS) factor, convergence index, depth from groundwater, distance from lineament, distance from channel network, and so forth, were extracted from spatial databases. Then, T and SPC data were selected (T values $\geq 2.6 \text{ m}^2/\text{day}$, SPC values $\geq 4.875 \text{ m}^3/\text{day}/\text{m}$) as training data for the three models. Finally, the GPP maps were assessed using area under the curve (AUC) techniques.

STUDY AREA

The study focused on Okcheon city of South Korea. This area uses $45,032,000 \text{ m}^3$ of groundwater per year from 23,856 pumping stations. Of that, 67.2% is used for human consumption, 32.1% for agriculture and 0.5% for industry (Ministry of Land, Transport and Maritime Affairs 2016). The average annual precipitation is 1,297.4 mm, similar to Korea's average annual precipitation of 1,277.4 mm (1978–2007) (MLTM 2012). The Okcheon region lies between $36^\circ 10' \text{N}$ and $36^\circ 26' \text{N}$ latitude and $127^\circ 29' \text{E}$ and $127^\circ 53' \text{E}$ longitude (Figure 2) and covers 537.06 km^2 . The altitude of this area ranges from 0 to 530.5 m above sea level. The terrain gradient ranges from 0° to 81.4° , with a mean value of 20.1° . It was computed from a $30 \times 30 \text{ m}$ digital elevation model (DEM) extracted from a 1:5,000 scale

topographic map. Geologically speaking, it belongs to the Okcheon era and includes the unrecorded Okcheon-supergroup, Paleozoic-Choseon-supergroup and Pyeongan-supergroup. Triassic and Jurassic granitic rocks, Cretaceous sedimentary rocks, volcanic and intrusive igneous rocks, and the fourth alluvial fan are present. A total of 65% of the study area is covered by land for agriculture, orchard or forest. Since groundwater is associated with drinking and irrigation water supplied to communities, it is very meaningful to estimate GPP.

SPATIAL DATA

In this study, groundwater productivity factors such as SPC and T as well as various geologic factors were applied as parameters in order to detect the potential productivity of groundwater. In calculations of groundwater productivity, SPC and T were dependent variables and various hydro-geological factors were independent variables.

Specific capacity (SPC) is defined as the amount of water that can be produced per unit drawdown. SPC is calculated by dividing the pumping rate by the drawdown and is often used as a well performance. Lower SPC will result in deeper pumping water levels, resulting in higher

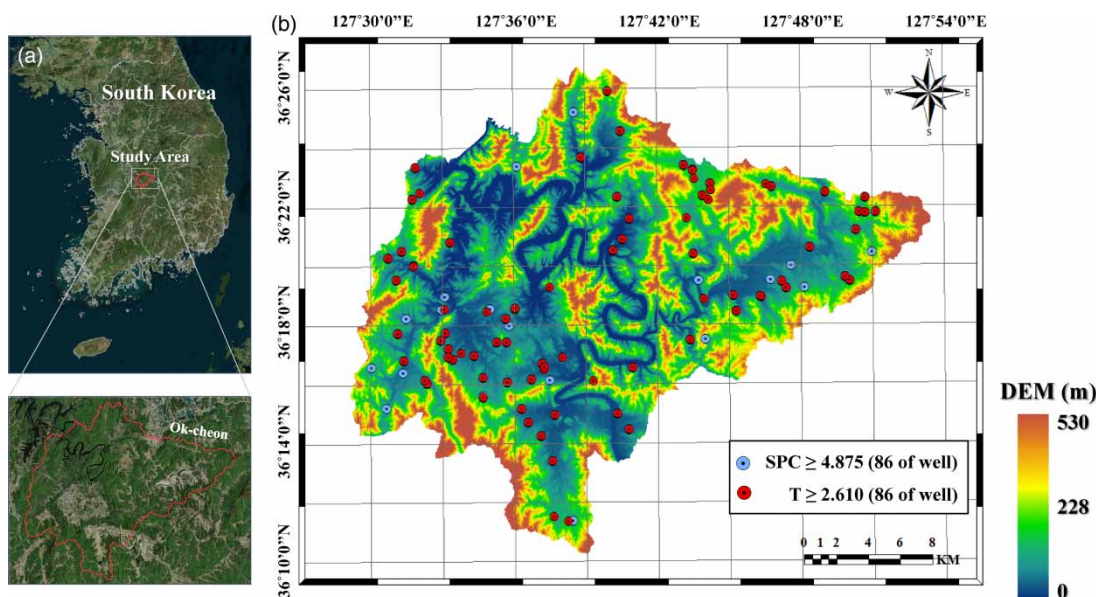


Figure 2 | Study area in South Korea with T/SPC values.

energy costs to pump. For the drawdown test to compute SPC, the pumping rate should be constant, and the period of pumping should be sufficiently long that the rate of change in drawdown is small. The well efficiency is related to the ratio between the theoretical reduction of the actual sewer and the aquifer within the well structure. This information is derived from pumping tests that last at least 24 h. SPC is expressed as:

$$S_c = \frac{Q}{h_0 - h} \quad (1)$$

where S_c is the SPC ((L^2T^{-1}) ; $m^3/day/m$), Q is the pumping rate ((L^3T^{-1}) ; m^3/day) and $h_0 - h$ is the drawdown ((L) ; m). SPC is the value obtained during the pumping test, which is used to identify potential aquifer problems or to establish maintenance plans. In addition, T of the aquifer can be estimated using SPC.

Transmissivity (T) is defined as the flow rate under a unit pressure gradient through a unit width of a given saturated thickness aquifer; in other words, the transmission capability of an entire thickness of aquifer. Also, hydraulic conductivity is a measure of a material's capacity to transmit water. It is defined as a constant of proportionality relating to the specific discharge of a porous medium under a unit hydraulic gradient in Darcy's law. In other words, hydraulic conductivity means the volume of water flowing through the $1ft \times 1ft$ cross-sectional area of an aquifer under a hydraulic gradient of $1ft/1ft$ in a given amount of time (usually 24 h). The T of an aquifer is related to its hydraulic conductivity as follows:

$$K'(x, y) = \frac{1}{b} \int_0^b K(x, y, z) dz \quad (2)$$

$$T = Kb \quad (3)$$

where T is transmissivity (L^2T^{-1}), b is aquifer thickness (L) and K is hydraulic conductivity. A higher T denotes a thicker aquifer and less well drawdown. In this study, both T and SPC were appropriately used and we applied two values to three models.

The DEM generated by the National Geographic Information Institute (NGII) had a resolution of 10 m.

We prepared an improved DEM by digitizing contour lines at 5 m intervals from a topographic map. Using this DEM, we calculated the slope gradient, slope aspect and TWI. Generally speaking, topographic factors such as slope gradient, slope aspect and curvature reflect geographical characteristics. Soil characteristics can influence groundwater potential. For example, soil texture determines the rate of surface water penetration into aquifers. The soil texture of the study area was extracted from a 1:25,000 scale soil map published by the National Institute of Agricultural Science (Lee & Lee 2015).

Various topographic and hydrogeological factors (total of 14 factors) used in the study have been converted to grid format through ArcGIS software. The final grid size was 3,047 rows by 3,642 columns. The study area was composed of 11,097,174 grid cells. T utilized 86 cells: 43 cells (including $T \geq 2.61 m^2/d$) for training and 43 cells for validation. SPC corresponded to a total of 86 cells: 43 cells (including $SPC \geq 4.875 m^3/d/m$) for training and 43 cells for validation. The groundwater productivity data were transformed into binary form by the pumping test where a value above the reference value was 1 and a value below the reference value was 0. These data were binarized from a T value threshold of $2.61 m^2/d$ and a corresponding SPC median value of $4.875 m^3/d/m$. Groundwater productivity data (Table 1), including T and SPC, were obtained from the national groundwater survey report of the Korea Institute of Geoscience and Mineral Resources (KIGAM) and the rural groundwater survey report of the Ministry for Food, Agriculture, Forestry, and Fisheries (MFAFF).

Groundwater is profoundly related to the presence of nearby water sources. The properties of the river related to the groundwater are the width and depth of the river and the velocity (Taormina & Chau 2015). Generally, the larger the width of the river, the deeper the river and the faster the flow rate, the greater the amount of water penetrating towards the aquifer (Langbein & Leopold 1964). Also, distance from the channel network may be related to groundwater potential (Chen et al. 2015). In this study, distance from the channel network had a negative correlation, which indicates that the closer the rivers, the greater the GPP.

Table 1 | Data layers of the study area

Original data	Factors	Data type	Scale
Yield	T [m ² /d/] SPC [m ³ /d/m]	Point	
Topographical map ^a	Slope gradient [°] Slope aspect [°] Relative slope position TWI Slope length factor (LS-factor) Convergence index Hydraulic slope [m] Valley depth [m] Depth to groundwater [m]	Grid	1:5,000
Geological map ^b	Hydrogeology	Polygon	1:50,000
Soil map ^c	Soil texture	Polygon	1:25,000
Land cover map ^d	Land use Distance from lineament [m] Distance from channel network [m]	Polygon	1:5,000

^aTopographical factors were extracted from digital topographic map by National Geographic Information Institute (NGII).

^bThe geology map supplied by Ministry of Land, Transport and Maritime Affairs (MLTM).

^cThe soil map supplied by the National Institute of Agricultural Science and Technology.

^dThe land cover map supplied by the Korea Ministry of Environment.

METHODOLOGY

Frequency ratio

In this study, we analysed the correlation between the groundwater well locations and various factors related to groundwater productivity using FR of each factor in groundwater occurrence.

Before applying the FR model, a correlation should be derived between the location of wells in the groundwater and the topography and other factors (Dadgar *et al.* 2017). Then, the FR is calculated by analysing the relationship of each factor. The relationship analysis is the ratio of the area where groundwater productivity (T or SPC) occurred to the total area. Therefore, a value of 1 indicates an average value. A greater value than 1 indicates a higher correlation (Lee *et al.* 2014). The steps for the model for the groundwater occurrence-contributing factor are below:

$$\text{Frequency ratio} = \frac{W/T_i}{P/T_p} \quad (4)$$

where W shows the number of wells in each class, T_i shows the total number of T (or SPC), P shows the number of pixels

in each class, and T_p indicates the total number of pixels. Last, FR values for all of the factors will be summed and the final GPP maps will be produced by Equation (5):

$$\text{Groundwater productivity potential maps} = \sum (FR)_i \quad (5)$$

A value of FR greater than 1 indicates a higher correlation to the potential of groundwater.

Evidential belief function

In this paper, the EBF model used for GPP evolved based on the Dempster–Shafer theory (Dempster 1967, 1968; Shafer 1976). Generally, estimation of EBFs is composed of the degree of belief (Bel), plausibility (Pls), uncertainty (Unc) and disbelief (Dis). Also, each value ranges from 0 to 1 (Lee *et al.* 2013). According to the Dempster–Shafer theory, generalized Bayesian upper and lower probabilities represent Bel and Pls. Bel indicates the degree of belief for the evidence. Pls shows the degree to which the evidence remains plausible (Al-Abadi *et al.* 2015). In general, Pls is equal to or larger than Bel and the difference for Pls and

Bel is Unc. Unc represents doubt about evidence supporting a proposition. Dis represents a belief that the proposition is false; it is equal to $1 - \text{Pls}$ (or $1 - \text{Unc-Bel}$). Thus, the sum of Bel, Dis and Unc is 1. The equations used for GIS-based data-driven EBFs for GPP mapping are not presented here, but can be found in Carranza & Hale (2003), Park *et al.* (2014a) and Tahmassebi *et al.* (2016).

Artificial neural network

ANN is the modelling of human brain action as a connection between neurons. An ANN is a modelling technique that finds hidden patterns in data through repetitive training processes from the data it has. In other words, the ANN model is one of the mechanisms that enables information to be sent from one multivariable space to another, such as a neuron (Garrett 1994). Since ANN has diverse advantages compared with statistical methods, it is widely used for analysis such as classification and diverse recognition. First, ANN is performed independently of the statistical distribution of the data, thus the estimation results can be obtained without using specific statistical parameters. Second, there is the advantage that an accurate analysis of the training set is available because the calculations are pixel-based. In addition, ANN has the advantage that

better results can be obtained using less training data than statistical techniques (Paola & Schowengerdt 1995).

There are many approaches to the ANN algorithm and the most frequently used is the back-propagation algorithm (Lee *et al.* 2016a). The algorithm trains the network until the desired minimum error between the network output and the target output is found. This algorithm is implemented to facilitate the modification of the number of hidden layers and the learning rate to measure the weights between the input, hidden and output layers. The weight between layers can be obtained by learning an ANN, and the importance of each element or contribution can be calculated using the obtained weight. This part can be successfully used to classify new input data that were not previously used (Atkinson & Tatnall 1997).

In this study, we used the ANN model to determine the weights of each layer acquired by neural network training, and in order to get the interpretation of the weight, we used MATLAB software. In the training set, areas with SPC values $< 4.875 \text{ m}^3/\text{d}/\text{m}$ (T values $< 2.61 \text{ m}^2/\text{d}$) were classified as 'not occurrence', and areas with SPC values of $\geq 4.875 \text{ m}^3/\text{d}/\text{m}$ (T values of $\geq 2.61 \text{ m}^2/\text{d}$) were classified as 'occurrence'. Finally, the structure 14 (input layer) \times 30 (hidden layer) \times 1 (output layer) was selected for the network, with input data normalized in the range 0.1–0.9 (Figure 3).

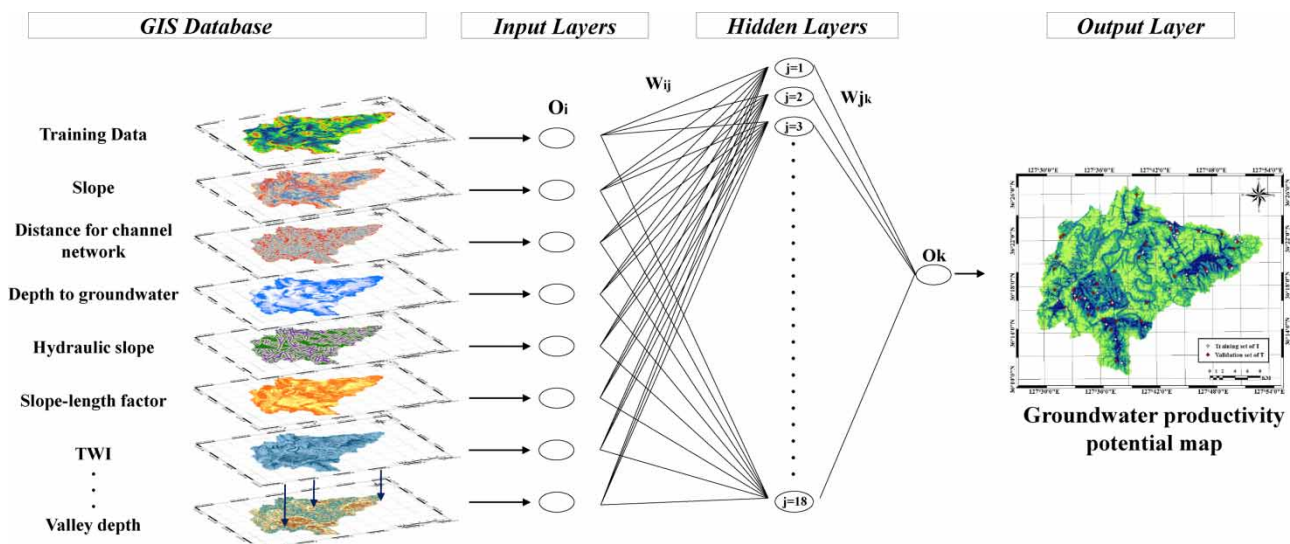


Figure 3 | Architecture of ANN model.

The conditions set for the ANN in the MATLAB program are as follows. First, the training learning rate was set to 0.02, and it was set to perform 1,000 epochs, and the root mean square error (RMSE) value—which is the reference value at which the learning was stopped—was set to 0.01. All iterations met the 0.01 RMSE goal in less than 1,000 epochs. During the training of ANN, weights were calculated for each of the 14 factors on groundwater productivity, and GPP mapping was done using this weight (Figure 4).

GPP mapping and validation

To analyse GPP, we created an ANN model in MATLAB and an FR and EBF model in SPSS. In addition, groundwater productivity data (T or SPC) were obtained from a national groundwater survey (2014) after the study area was determined. These productivity data were used as training and validation data. Three different data mining and probabilistic models were applied to the GPP analysis.

An overall analysis of the maps was conducted using ArcGIS 10.1. Each factor extracted from the map was resampled in a 10×10 m grid format (Park *et al.* 2014b). After converting the ASCII file, we applied the ANN model to the data in MATLAB. The T and SPC points were used as dependent variables (training data), and the factors were set as independent variables. All factors were categorized as either continuous or categorical data. Continuous variables included slope gradient, relative slope position, hydraulic slope, and distance from lineament, depth to groundwater, valley depth, TWI, LS factor, convergence index and distance from river. Categorical variables included slope aspect, hydrogeology, land cover and soil

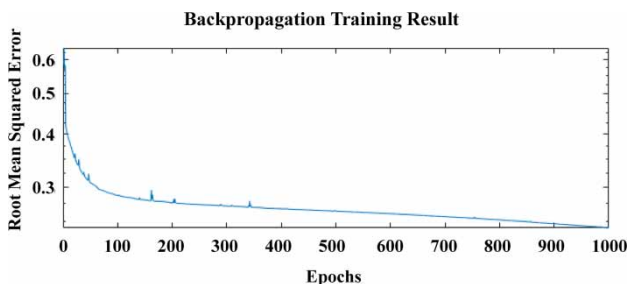


Figure 4 | A backpropagation training result of an ANN model. The graph represents the RMSE value for 1,000 epochs.

map. The predicted GPP maps effectively represented the future potential productivity of groundwater.

To validate the training data, we randomly and equally divided 86 T data points into training and validation data sets. The same operation was performed using 84 SPC data points. After the models were run with the training data, validation was performed. We constructed receiver operating curves by calculating the AUC. Then, we compiled the relative rankings by sorting in descending order the values calculated in the study area prediction. For a quantitative comparison, the AUCs were recalculated in the total area and used in indicating correct prediction accuracy. In this study, the AUC was used to evaluate and express the predictive accuracy capabilities of the three models.

In summary, GPP mapping was performed as follows: (1) geospatial data were collected and related factors were extracted and calculated; (2) a gridded geospatial database was created; (3) GPP assessment was conducted using FR and ANN models; and (4) potential map validation was performed with validation productivity data (T or SPC) for prediction rates.

RESULTS

This study used several models to estimate potential groundwater production in the Okcheon area. GPP is influenced by factors such as topography, geology, forests and soils (Kang *et al.* 2017). To estimate the correlation between groundwater capacity and these factors, we performed GPP mapping using FR and EBF models (probabilistic model) and an ANN model (a data mining model) (Figure 5).

First, we used an FR model to calculate the relationship between groundwater locations and hydrogeological factors. Then, we validated the model results. Table 2 displays the FR of the factors in each class. These were calculated with respect to groundwater well locations with T values ≥ 2.61 m²/d and SPC values ≥ 4.875 m³/d/m. The relationship between GPP and slope gradient revealed that either flat or gradual slopes have greater groundwater probabilities. For slopes between 0° and 5°, the ratio was >2 , which indicates a high probability of groundwater potential. Similarly, the lower the values for hydraulic slope, relative slope position or slope length, the higher the probability of

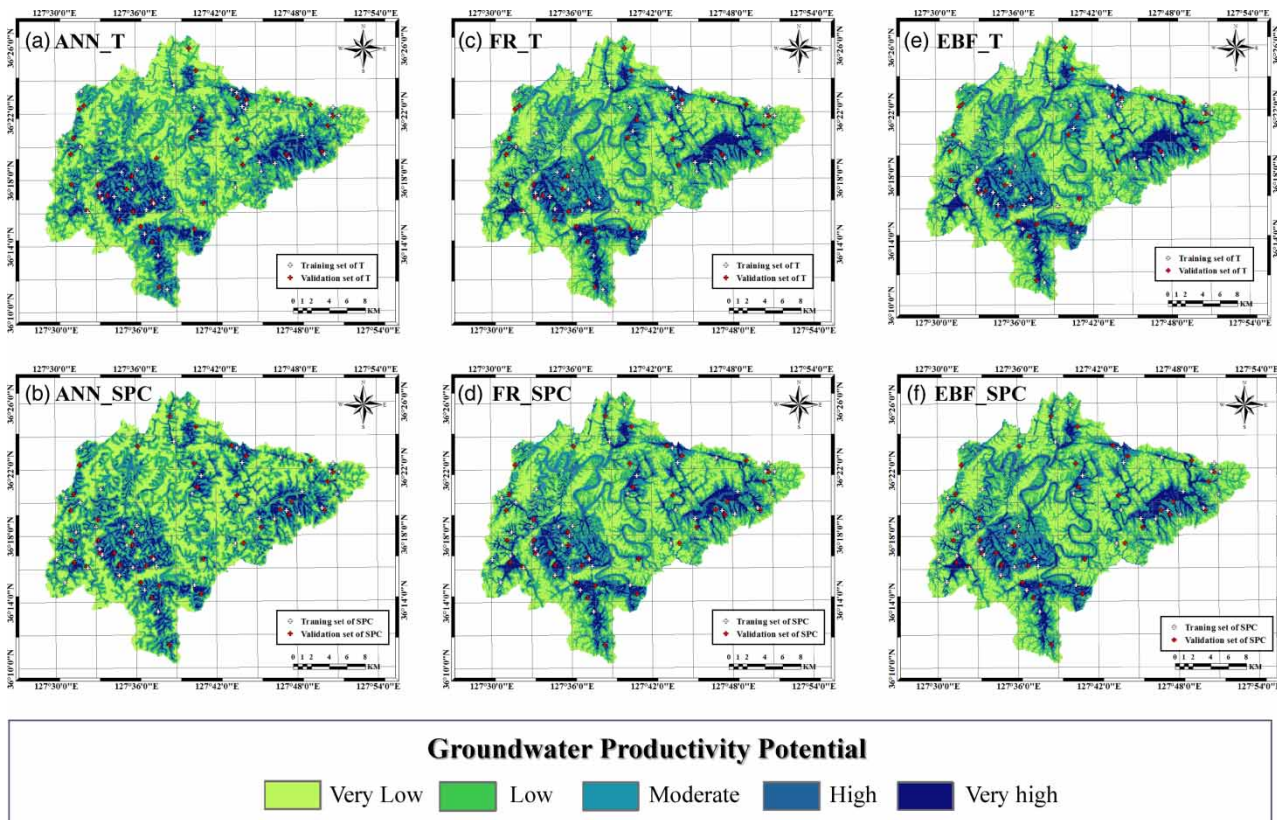


Figure 5 | Groundwater productivity potential map generated using ANN and frequency-ratio models. GPP index was classified into five classes: very high (10%), high (10%), moderate (20%), low (30%) and very low (30%) index ranges of the study area. GPP map with T/SPC values of (a), (b) ANN model, (c), (d) FR model and (e), (f) EBF model.

groundwater potential. The convergence index is a topographic parameter that represents the aspect ratio between the real face and the theoretical maximum divergence direction matrix. When there is maximum divergence, the index is -100 ; when there is maximum convergence, the index is 100 . This means that the nearer the slope is to maximum convergence, the lower the potential occurrence of groundwater.

The FR was higher in areas with unconsolidated clastic sediments compared to areas dominated by carbonate and nonporous volcanic rocks. These rock types are assumed to have poor groundwater potential because of low groundwater movement. Groundwater potential values were higher in urban areas and paddy fields but lower in grassland and mixed forest areas.

Frequency ratios also varied for soils. The ratio was higher (>3) for D soils (very low infiltration rate) and lower ($0.0\text{--}0.6$) for B soils (moderate infiltration rate).

Sandy soils can positively impact groundwater generation because of high permeability. Conversely, clay soils have poor drainage and low permeability. The positive curvature represents the convex surface, the negative curvature represents the concave surface, and the curvature value 0 represents the plane. Concave areas had a ratio of 1.14 compared to 0.80 for convex. Concave surfaces can hold more water, particularly during periods of heavy rainfall. Therefore, groundwater capacity is typically higher in areas with concave surfaces.

Groundwater potential is reduced when distance from a lineament, or channel network, increases. That is, various linear structures and distant areas had many spills, but nearby areas had more recharge and higher infiltration. TWI is a steady-state wetness index that is a function of the slope per unit to the flow direction. Results demonstrated that as TWI increased, the potential for groundwater generation increased because the ground could retain more

Table 2 | FR between groundwater productivity (T, SPC) and related factor

Factor	Class	No. of pixels in domain ^a	% of domain	T ≥ 2.61 ^b			SPC ≥ 4.875 ^b		
				No. of data 1	% of data 1	FR of data 1	No. of data 1	% of data 1	FR of data 1
Slope gradient (degree)	0–4.15	1,084,566	20.04	23	53.49	2.67	21	48.84	2.44
	4.15–15.65	1,116,798	20.64	16	37.21	1.80	17	39.53	1.92
	15.65–24.59	1,062,944	19.64	1	2.33	0.12	2	4.65	0.24
	24.59–32.57	1,100,385	20.33	2	4.65	0.23	2	4.65	0.23
	32.57–90	1,047,035	19.35	1	2.33	0.12	1	2.33	0.12
Hydraulic slope (m)	0–5	1,386,038	25.61	34	79.07	3.09	37	86.05	3.36
	5–10	1,005,238	18.58	7	16.28	0.88	5	11.63	0.63
	10–20	1,629,140	30.10	2	4.65	0.15	1	2.33	0.08
	20–30	903,861	16.70	0	0	0	0	0	0
	30–90	487,451	9.01	0	0	0	0	0	0
Relative slope position	0–0.04	1,204,753	22.26	20	46.51	2.09	23	53.49	2.40
	0.04–0.25	1,066,299	19.70	16	37.21	1.89	15	34.88	1.77
	0.25–0.49	1,051,292	19.43	2	4.65	0.24	0	0	0
	0.49–0.76	1,046,943	19.35	3	6.98	0.36	4	9.30	0.48
	0.76–1	1,042,441	19.26	2	4.65	0.24	1	2.33	0.12
Valley depth (m)	0–19.12	1,068,280	19.74	8	18.60	0.94	6	13.95	0.71
	19.12–37.05	1,134,687	20.97	17	39.53	1.89	13	30.23	1.44
	37.05–58.56	1,105,662	20.43	5	11.63	0.57	9	20.93	1.02
	58.56–88.44	1,079,756	19.95	10	23.26	1.17	9	20.93	1.05
	88.44–304.77	1,023,343	18.91	3	6.98	0.37	6	13.95	0.74
TWI	–0.27–3.56	1,147,597	21.21	3	6.98	0.33	2	4.65	0.22
	3.56–4.26	1,237,405	22.87	2	4.65	0.20	1	2.33	0.10
	4.26–5.36	1,059,243	19.57	3	6.98	0.36	3	6.98	0.36
	5.36–7.78	1,005,413	18.58	16	37.21	2	19	44.19	2.38
	7.78–25.37	962,070	17.78	19	44.19	2.49	18	41.86	2.35
LS factor (m)	0–0.93	1,003,897	18.55	20	46.51	2.51	21	48.84	2.63
	0.93–3.72	1,129,732	20.88	18	41.86	2.01	16	37.21	1.78
	3.72–6.33	1,112,813	20.56	3	6.98	0.34	3	6.98	0.34
	6.33–8.93	1,115,981	20.62	1	2.33	0.11	2	4.65	0.23
	8.93–47.46	1,049,305	19.39	1	2.33	0.12	1	2.33	0.12
Distance from lineament (m)	0–85.82	1,144,797	21.15	15	34.88	1.65	14	32.56	1.54
	85.82–193.09	1,158,897	21.41	8	18.60	0.87	11	25.58	1.19
	193.09–321.82	1,070,062	19.77	7	16.28	0.82	5	11.63	0.59
	321.82–522.06	1,049,599	19.39	9	20.93	1.08	10	23.26	1.20
	522.06–1,823.62	988,373	18.26	4	9.30	0.51	3	6.98	0.38
Distance from channel network (m)	0–49.71	1,072,193	19.81	16	37.21	1.88	20	46.51	2.35
	49.71–109.36	1,236,809	22.85	15	34.88	1.53	11	25.58	1.12
	109.36–175.64	1,085,313	20.05	8	18.60	0.93	10	23.26	1.16
	175.64–271.74	1,058,045	19.55	4	9.30	0.48	2	4.65	0.24
	271.74–845.04	959,368	17.73	0	0	0	0	0	0
Depth to groundwater (m)	0–6	697,317	12.89	7	16.28	1.26	8	18.60	1.44
	6–12	1,492,489	27.58	25	58.14	2.11	24	55.81	2.02
	12–18	1,062,904	19.64	10	23.26	1.18	9	20.93	1.07
	18–24	788,619	14.57	1	2.33	0.16	1	2.33	0.16
	24–30	1,370,399	25.32	0	0	0	1	2.33	0.09

(continued)

Table 2 | continued

Factor	Class	No. of pixels in domain ^a	% of domain	$T \geq 2.61^b$			$SPC \geq 4.875^b$		
				No. of data 1	% of data 1	FR of data 1	No. of data 1	% of data 1	FR of data 1
Slope aspect	Flat	355,792	6.57	4	9.30	1.41	3	6.98	1.06
	North	594,932	10.99	7	16.28	1.48	3	6.98	0.63
	Northeast	641,589	11.86	7	16.28	1.37	10	23.26	1.96
	East	658,668	12.17	6	13.95	1.15	5	11.63	0.96
	Southeast	597,515	11.04	4	9.30	0.84	3	6.98	0.63
	South	542,960	10.03	2	4.65	0.46	2	4.65	0.46
	Southwest	638,521	11.80	0	0	0	4	9.30	0.79
	West	710,517	13.13	5	11.63	0.89	8	18.6	1.42
	Northwest	671,234	12.40	8	18.60	1.50	5	11.63	0.94
Hydrogeology	Unconsolidated clastic rock	845,761	15.63	16	37.21	2.38	16	37.21	2.38
	Intrusive igneous rocks	2,301,033	42.52	20	46.51	1.09	18	41.86	0.98
	Coordinated carbonate rock	82,527	1.52	1	2.33	1.53	0	0	0
	Nonporous volcanic rock	3,886	0.07	0	0	0	0	0	0
	Clastic sedimentary rock	131,034	2.42	1	2.33	0.96	1	2.33	0.96
	Carbonate rocks	42,674	0.79	0	0	0	0	0	0
	Metamorphic rocks	2,004,813	37.05	5	11.63	0.31	8	18.60	0.50
Land use	Paddy field	500,200	9.24	16	37.21	4.03	14	32.56	3.52
	Field	694,102	12.83	13	30.23	2.36	13	30.23	2.36
	Grass land	34,097	0.63	0	0	0	0	0	0
	Mixed forest	3,626,815	67.02	7	16.28	0.24	7	16.28	0.24
	Other barren	2,166	0.04	0	0	0	0	0	0
	Residential area	127,960	2.36	3	6.98	2.95	3	6.98	2.95
	Transportation	41,306	0.76	1	2.33	3.05	2	4.65	6.09
	Industrial area	24,569	0.45	1	2.33	5.12	3	6.98	15.37
	Public facilities area	15,554	0.29	0	0	0	1	2.33	8.09
	Commercial area	6,505	0.12	0	0	0	0	0	0
	Inland wetland	477	0.01	0	0	0	0	0	0
	River	282,667	5.22	1	2.33	0.45	0	0	0
	Inland water	55,310	1.02	1	2.33	2.28	0	0	0
Soil texture	High infiltration rate	2,227,069	41.15	20	46.51	1.13	20	46.51	1.13
	Moderate infiltration rate	950,195	17.56	9	20.93	1.19	7	16.28	0.93
	Low infiltration rate	1,795,572	33.18	7	16.28	0.49	9	20.93	0.63
	Very slow infiltration rate	202,533	3.74	5	11.63	3.11	7	16.28	4.35
	Water	236,359	4.37	2	4.65	1.06	0	0	0
Convergence index	Concave(-)	2,644,781	48.87	26	60.47	1.24	29	67.44	1.38
	0	13,359	0.25	0	0	0	0	0	0
	Convex(+)	2,753,588	50.88	17	39.53	0.78	14	32.56	0.64

^aTotal number of pixels is 5,411,728.^bTotal number of groundwater well pixels is 43.

wetness. Depth to groundwater had the highest groundwater potential, between 6 and 12 m. This implies that it is difficult to generate groundwater when an aquifer is either too shallow or too deep. Aspect showed the highest potential for groundwater generation on northerly aspects (north, northeast and northwest).

The neural network weights represent the contribution of each factor for GPP. For SPC, the convergence index showed the lowest average value of weight (0.0445). For T, valley depth showed the lowest average value of weight (0.0545). We quantitatively compared the relative weights of SPC and T by normalizing the data. Each value was divided by the lowest value. For T, distance from channel network displayed the highest contribution to GPP, with a normalized value of 1.74452. This was followed by land use (1.73675) and depth to groundwater (1.60861). Standard deviations ranged from 0.0003 to 0.0037. For SPC, hydraulic slope displayed the highest contribution to GPP, with a normalized value of 1.68006. This was followed by land use (1.54082) and TWI (1.52082). The standard deviations ranged from 0.0004 to 0.0082.

In GPP mapping, the predictor is a factor, and groundwater potential probability is defined as the productivity potential of groundwater. The probabilities in each of the three models were considered by the GPP index (GPPI). We have reclassified the GPPI to a potential productivity grade to get each GPP map. The index is divided into five classes according to the area occupied by: very high, high, medium, low and very low index ranges in 10%, 10%, 20%, 30% and 30%, respectively (Pradhan & Lee 2010; Naghibi & Pourghasemi 2015; Mohamed & Elmahdy 2017). This classification helped predict the GPP of each class as well as augmenting visualization of the predicted GPP area. As a result, the GPP maps were produced using the GPPI extracted from the FR and EBF and ANN models (Figure 5).

Validation

Next was to validate the generated GPP maps and compare the predictive accuracy of the ANN model to the FR and EBF models relying on two basic assumptions: the groundwater productivity area relates to spatial information and potential groundwater occurs under the same spatial conditions.

In this study, we determined the prediction rates for the validations by comparing the potential maps created using the FR and EBF and ANN models with the validation data set. We validated the GPP map by first sorting all GPPI cell values in descending order. Then, we divided these values into 100 classes at 1% cumulative intervals. More than 10% of the GPPI explained 30% of all groundwater potential wells by the three models, and 90% of groundwater was explained by an index rank >60%. We also utilized this procedure for all cells with a high likelihood of groundwater potential by comparing the 100 classes. Then, we made a graph by connecting two classified values. The AUC was used to assess the prediction accuracy qualitatively. For a quantitative comparison, the AUC values were recalculated as the total area (Lee & Sambath 2006).

For the T value and from the validation of the GPP maps (Figure 6), the ANN, FR and EBF approaches produced AUC values of 0.8219, 0.8115 and 0.8040, respectively. For the SPC value and from the validation of the GPP maps, the ANN, FR and EBF approaches produced AUC values of 0.8167, 0.8136 and 0.7989, respectively. This indicated that the AUC probability of the GPP maps for each model had corresponding accuracy. The ANN model had an AUC value of 0.8219, which was approximately 1.04% higher than that of the FR model and 1.79% higher than EBF model for the T value. For the SPC value, the ANN model had an AUC value of 0.8167, which was approximately 0.3% higher than that of the FR model and 1.79% higher than EBF model for the SPC value. In all models, 60% of the study area explained 90% of the groundwater potential wells. Both models showed high accuracy in GPP.

DISCUSSION AND CONCLUSIONS

Around 20% of total global water use originates from groundwater sources (renewable or not), and this rate is rising rapidly. Investments in safe drinking water and sanitation, such as groundwater development, can contribute to the global economy and social community. Some areas in Korea, such as Okcheon, utilize only surface water. Because of the unusual meteorological phenomenon during the dry season, the surface water of these regions is

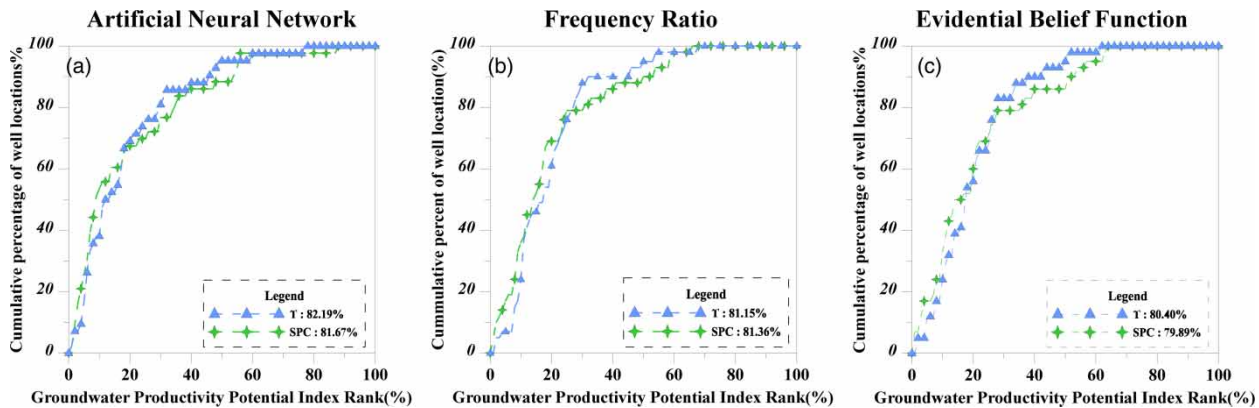


Figure 6 | ROC curves for the groundwater potential maps with T/SPC values produced by (a) ANN model, (b) FR model and (c) EBF model.

overly taxed. These areas need a stable water acquisition system that can provide high-quality, stable water for drinking and agriculture.

This study applied FR and EBF (probabilistic) and ANN (data mining) models to GIS to estimate the region's potential for groundwater productivity and compared the results of these models. We used 14 factors used in three models originally, and to produce appropriate results, we researched removing three low weight factors and using 14 factors. The three factors that were removed were distance from fault, lineament density and profile curvature. We selected three factors that consistently maintained the lowest values from the FR and ANN weight values. As a result, the result of ANN verification was higher than previously, and the T value was more accurate than the SPC value. The FR value has also been fine-tuned. The AUCs for the prediction data set were 82.19, 81.15 and 80.40 for the ANN, FR and EBF models for the T value, and 81.67, 81.36 and 79.89 for the ANN, FR and EBF models for the SPC value. The ANN model had better performance for both the T and SPC value than the FR and EBF models. The FR model displayed high prediction accuracy (>80%) for groundwater potential.

High GPP was estimated under a variety of conditions: gentle slopes, gentle hydraulic slopes, low relative slope positions and short slope lengths. It is interesting that high GPP was also estimated for steep slopes, steep hydraulic slopes, higher relative slope positions and longer slope lengths. This was primarily because of surface runoff processes such as rainfall runoff in the upper region and

accumulation in the lower regions positively influencing the aquifer. These variables were juxtaposed with distance from lineament and distance from channel network, which showed a negative correlation with GPP. The closer the channel, the greater the GPP because of surface water and groundwater exchange. Finally, the higher the linear density, the higher the GPP estimates.

The ANN model is simple and can process more data quickly, and the results are easy to understand. Furthermore, the ANN model can be calculated by factor weight that is given to factors that are important in GPP analysis according to significance ranking. The ANN model provides a more accurate analysis of GPPs than many other models and has the advantage of being able to easily process continuous and discrete data.

In fact, we already know that the results of many data-driven techniques, including FR, EBF and even ANN, are very sophisticated and accurate, but still do not fully reflect the real world. However, since it is impossible to conduct a comprehensive survey of a wide area of research in reality, we can only use the existing survey data to verify it and trust its value. We verified the predictive value quantitatively through the AUC method and quantitatively analysed the importance of all data through weighted analysis. We also attempted to reduce this uncertainty using the most accurate scale of available data, recognizing the uncertainty of the input data with scale. Therefore, the results of this study can be trusted. This proposed GPP mapping method can be applied to groundwater use planning and management, such as regional groundwater development planning and

water system control based on systematic and objective planning. The application of GPP maps can help reduce potential well development costs, which is vital for countries with limited water sources (Elmahdy & Mohamed 2014; Falah et al. 2017).

ACKNOWLEDGEMENTS

This research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Minister of Science, ICT and Future Planning of Korea. This research (NRF-2016K1A3A1A09915721) was supported by Science and Technology Internationalization Project through National Research Foundation of Korea (NRF) grant funded by the Ministry of Education, Science and Technology (MEST). This research was also supported by Public Technology Development Project based on Environmental Policy (2016000210001) provided by Korea Environmental Industry and Technology Institute.

REFERENCES

- Al-Abadi, A. M., Pradhan, B. & Shahid, S. 2015 Prediction of groundwater flowing well zone at An-Najif Province, central Iraq using evidential belief functions model and GIS. *Environ. Monit. Assess.* **188** (10), 549.
- Atkinson, P. M. & Tatnall, A. 1997 Introduction neural networks in remote sensing. *Int. J. Remote Sens.* **18** (4), 699–709.
- Carranza, E. J. M. & Hale, M. 2003 Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines. *Ore Geol. Rev.* **22** (1), 117–132.
- Chen, X.-y., Chau, K.-w. & Wang, W.-c. 2015 A novel hybrid neural network based on continuity equation and fuzzy pattern-recognition for downstream daily river discharge forecasting. *J. Hydroinf.* **17** (5), 733–744.
- Dadgar, M. A., Zeaieanfirouzabadi, P., Dashti, M. & Porhemmat, R. 2017 Extracting of prospective groundwater potential zones using remote sensing data, GIS, and a probabilistic approach in Bojnourd basin, NE of Iran. *Arab. J. Geosci.* **10**, 114.
- Dempster, A. P. 1967 Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **38** (2), 325–339.
- Dempster, A. P. 1968 A generalization of Bayesian inference. *J. R. Stat. Soc. Series B Stat. Methodol.* **30**, 205–247.
- Elmahdy, S. I. & Mohamed, M. M. 2014 Groundwater potential modelling using remote sensing and GIS: a case study of the Al Dhaid area, United Arab Emirates. *Geocarto. Int.* **29** (4), 433–450.
- Elmahdy, S. I. & Mohamed, M. M. 2015 Probabilistic frequency ratio model for groundwater potential mapping in Al Jaww plain, UAE. *Arab. J. Geosci.* **8** (4), 2405–2416.
- Falah, F., Ghorbani Nejad, S., Rahmati, O., Daneshfar, M. & Zeinivand, H. 2017 Applicability of generalized additive model in groundwater potential modelling and comparison its performance by bivariate statistical methods. *Geocarto. Int.* **32** (10), 1069–1089.
- Ganapuram, S., Nagarajan, R. & Chandra Sekhar, G. 2016 Identification of groundwater drought prone zones in Pedda vagu and Ookachetti vagu watersheds, tributaries of the Krishna River, India. *Geocarto. Int.* **31** (4), 385–407.
- Garrett, J. H. 1994 Where and why artificial neural networks are applicable in civil engineering. *J. Comput. Civil Eng.* **8** (2), 129–130.
- Ghorbani Nejad, S., Falah, F., Daneshfar, M., Haghizadeh, A. & Rahmati, O. 2017 Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. *Geocarto. Int.* **32** (2), 167–187.
- Hou, E., Wang, J. & Chen, W. 2018 A comparative study on groundwater spring potential analysis based on statistical index, index of entropy and certainty factors models. *Geocarto. Int.* **33** (7), 754–769.
- Hwang, J., Kim, D. & Jung, H.-S. 2017 An efficient ship detection method for KOMPSAT-5 synthetic aperture radar imagery based on adaptive filtering approach. *Korean J. Remote Sens.* **33** (1), 89–95.
- IPCC 2013 *Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.
- Jenifer, M. A. & Jha, M. K. 2017 Comparison of Analytic Hierarchy Process, Catastrophe and Entropy techniques for evaluating groundwater prospect of hard-rock aquifer systems. *J. Hydrol.* **548**, 605–624.
- Jothibasu, A. & Anbazhagan, S. 2017 Spatial mapping of groundwater potential in Ponnaiyar River basin using probabilistic-based frequency ratio model. *Model. Earth Syst. Environ.* **3** (1), 33.
- Kang, J., Zhao, L., Li, R., Mo, H. & Li, Y. 2017 Groundwater vulnerability assessment based on modified DRASTIC model: a case study in Changli County, China. *Geocarto. Int.* **32** (7), 749–758.
- Kim, D. & Jung, H.-S. 2017 Oil spill detection from RADARSAT-2 SAR image using non-local means filter. *Korean J. Remote Sens.* **33** (1), 61–67.
- Kim, J.-C., Kim, D.-H., Park, S.-H., Jung, H.-S. & Shin, H.-S. 2014 Application of Landsat images to snow cover changes by volcanic activities at Mt. Villarica and Mt. Lliama, Chile. *Korean J. Remote Sens.* **30** (3), 341–350.
- Kim, N., Cho, J., Hong, S., Ha, K.-J., Shibasaki, R. & Lee, Y.-W. 2016 Statistical estimation of crop yields for the Midwestern United States using satellite images, climate datasets, and soil property maps. *Korean J. Remote Sens.* **32** (4), 383–401.

- Kim, D., Jung, H.-S. & Kim, J.-C. 2017 Comparison of snow cover fraction functions to estimate snow depth of South Korea from MODIS imagery. *Korean J. Remote Sens.* **33** (4), 401–410.
- Kim, J.-C., Lee, S., Jung, H.-S. & Lee, S. 2018 Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto. Int.* **33** (9), 1000–1015.
- Langbein, W. B. & Leopold, L. B. 1964 Quasi-equilibrium states in channel morphology. *Am. J. Sci.* **262** (6), 782–794.
- Lee, S. & Lee, C.-W. 2015 Application of decision-tree model to groundwater productivity-potential mapping. *Sustainability* **7** (10), 13416–13432.
- Lee, S. & Sambath, T. 2006 Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ. Geol.* **50** (6), 847–855.
- Lee, S. & Talib, J. A. 2005 Probabilistic landslide susceptibility and factor effect analysis. *Environ. Geol.* **47** (7), 982–990.
- Lee, S., Ryu, J.-H., Lee, M.-J. & Won, J.-S. 2006 The application of artificial neural networks to landslide susceptibility mapping at Janghung, Korea. *Math. Geol.* **38** (2), 199–220.
- Lee, S., Song, K.-Y., Kim, Y. & Park, I. 2012 Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. *Hydrogeol. J.* **20** (8), 1511–1527.
- Lee, S., Hwang, J. & Park, I. 2013 Application of data-driven evidential belief functions to landslide susceptibility mapping in Jinbu, Korea. *Catena* **100**, 15–30.
- Lee, M.-J., Song, W.-K., Wom, J.-S., Park, I. & Lee, S. 2014 Spatial and temporal change in landslide hazard by future climate change scenarios using probabilistic-based frequency ratio model. *Geocarto. Int.* **29** (6), 639–662.
- Lee, S., Jeon, S.-W., Oh, K.-Y. & Lee, M.-J. 2016a The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: a case study of Inje, Korea. *Open Geosci.* **8** (1), 117–132.
- Lee, S., Park, S. H. & Jung, H. S. 2016b Multi-temporal analysis of deforestation in Pyeongyang and Hyesan, North Korea. *Korean J. Remote Sens.* **32** (1), 1–11.
- Lee, S., Kim, J.-C., Jung, H.-S., Lee, M. J. & Lee, S. 2017a Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomat. Nat. Haz. Risk* **8** (2), 1185–1203.
- Lee, S., Lee, M.-J. & Jung, H.-S. 2017b Data mining approaches for landslide susceptibility mapping in Umyeonsan, Seoul, South Korea. *Appl. Sci.* **7** (7), 683.
- Lee, S., Hong, S.-M. & Jung, H.-S. 2018 GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. *Geocarto. Int.* **33** (8), 847–861.
- MLTM 2012 *National Groundwater Monitoring Network Construction Report 2012*. Ministry of Land, Transport and Maritime Affairs, Korea.
- MLTM 2016 *National Groundwater Monitoring Network in Korea Annual Report 2016*. Ministry of Land, Transport and Maritime Affairs, Korea.
- Mogaji, K. A. & Lim, H. S. 2018 Development of groundwater favourability map using GIS-based driven data mining models: an approach for effective groundwater resource management. *Geocarto. Int.* **33** (4), 1–26.
- Mohamed, M. M. & Elmahdy, S. I. 2017 Fuzzy logic and multi-criteria methods for groundwater potentiality mapping at Al Fo'ah area, the United Arab Emirates (UAE): an integrated approach. *Geocarto. Int.* **32** (10), 1120–1138.
- Mousavi, S. M., Golkarian, A., Naghibi, S. A., Kalantar, B. & Pradhan, B. 2017 GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. *AIMS Geosciences* **3** (1), 91–115.
- Naghibi, S. A. & Pourghasemi, H. R. 2015 A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Resour. Manage.* **29** (14), 5217–5236.
- Naghibi, S. A., Pourghasemi, H. R. & Abbaspour, K. 2017 A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theor. Appl. Climatol.* **131** (3–4), 967–984.
- Oh, H.-J., Kim, Y.-S., Choi, J.-K., Park, E. & Lee, S. 2011 GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *J. Hydrol.* **399** (3), 158–172.
- Oh, K.-y., Jung, H.-s. & Lee, M.-j. 2017 Accuracy Evaluation of DEM generated from satellite images using automated geopositioning approach. *Korean J. Remote Sens.* **33** (1), 69–77.
- Ozdemir, A. 2011a GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *J. Hydrol.* **411** (3–4), 290–308.
- Ozdemir, A. 2011b Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *J. Hydrol.* **405** (1–2), 123–136.
- Paola, J. & Schowengerdt, R. 1995 A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **16**, 3033–3058.
- Park, I., Kim, Y. & Lee, S. 2014a Groundwater productivity potential mapping using evidential belief function. *Groundwater* **52** (S1), 201–207.
- Park, I., Lee, J. & Saro, L. 2014b Ensemble of ground subsidence hazard maps using fuzzy logic. *Cent. Eur. J. Geosci.* **6** (2), 207–218.
- Park, S., Hamm, S.-Y., Jeon, H.-T. & Kim, J. 2017 Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS. *Sustainability* **9** (7), 1157.
- Pourghasemi, H. R. & Beheshtirad, M. 2015 Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran. *Geocarto. Int.* **30** (6), 662–685.

- Pradhan, B. & Lee, S. 2010 Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Software* **25** (6), 747–759.
- Rahmati, O., Pourghasemi, H. R. & Melesse, A. M. 2016 Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena* **137**, 360–372.
- Shafer, G. 1976 *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey.
- Sokeng, V. J., Kouamé, F., Ngatcha, B. N., N'da, H. D., You Akpa, L. & Rirabe, D. 2016 Delineating groundwater potential zones in Western Cameroon Highlands using GIS based Artificial Neural Networks model and remote sensing data. *Int. J. Innovation Appl. Stud.* **15** (4), 747–759.
- Tahmassebpour, N., Rahmati, O., Noormohamadi, F. & Lee, S. 2016 Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arab. J. Geosci.* **9**, 79.
- Taormina, R. & Chau, K.-w. 2015 Neural network river forecasting with multi-objective fully informed particle swarm optimization. *J. Hydroinform.* **17** (1), 99–113.
- WEF 2016 *The Global Risk Report 2016*. World Economic Forum, Geneva, Switzerland.
- WWDR 2017 *The United Nations World Water Development Report 2017. Wastewater: The Untapped Resource*. UNESCO, Paris, France.
- Zabihi, M., Pourghasemi, H. R., Pourtaghi, Z. S. & Behzadfar, M. 2016 GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environ. Earth Sci.* **75** (8), 1–19.
- Zandi, J., Ghazvinei, P. T., Hashim, R., Yusof, K. B. W., Arrifin, J. & Motamedi, S. 2016 Mapping of regional potential groundwater springs using logistic regression statistical method. *Water Resour.* **43** (1), 48–57.
- Zeinivand, H. & Ghorbani Nejad, S. 2018 Application of GIS-based data-driven models for groundwater potential mapping in Kuhdasht region of Iran. *Geocarto. Int.* **33** (6), 651–666.

First received 5 March 2018; accepted in revised form 6 June 2018. Available online 12 July 2018