
Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model

Saro Lee · Kyo-Young Song · Yongsung Kim · Inhye Park

Abstract An artificial neural network model (ANN) and a geographic information system (GIS) are applied to the mapping of regional groundwater productivity potential (GPP) for the area around Pohang City, Republic of Korea. The model is based on the relationship between groundwater productivity data, including specific capacity (SPC) and its related hydrogeological factors. The related factors, including topography, lineaments, geology, and forest and soil data, are collected and input into a spatial database. In addition, SPC data are collected from 44 well locations. The SPC data are randomly divided into a training set, to analyse the GPP using the ANN, and a test set, to validate the predicted potential map. Each factor's relative importance and weight are determined by the back-propagation training algorithms and applied to the input factor. The GPP value is then calculated using the weights, and GPP maps are created. The map is validated using area under the curve analysis with the SPC data that have not been used for training the model. The validation shows prediction accuracies between 73.54 and 80.09%.

Such information and the maps generated from it could serve as a scientific basis for groundwater management and exploration.

Keywords Groundwater development · Hydrogeological factor · Back-propagation training · Geographic information systems · Korea

Introduction

Groundwater, one of the most important natural resources, supports human health, economic development, and ecological diversity. The use of groundwater has increased because of factors such as high obtainability, excellent quality, and low development cost (Todd and Mays 2005). Surface water accounts for 0.3 % of the fresh water that exists on earth. In comparison, groundwater amounts to 30 % of the fresh water (Gleick 1993). Moreover groundwater can be recharged with rainfall consistently. Therefore, systematic development and management planning is crucial for establishing stabilised, secure sources of water. In the US, groundwater accounts for 20 % of all water resources used (Kenny et al. 2009). On the other hand, the occupancy rate of groundwater is 11 % in the Republic of Korea (MLTM 2006), and public water supplies (e.g., town water supplies) use only 5 % groundwater. Considering that water usage in Korea increased by more than 210 % (MLTM 2009) between 1994 and 2008, the development and utilisation of groundwater at national level is not up to the expectation of the people.

Little research has been conducted on the distribution of groundwater productivity potential (GPP) worldwide, and analysis methods have not been systematically formulated. The lack of a systematic approach has led to high time and labour costs and distorted estimates of groundwater resources. Development of groundwater resources takes time and requires considerable finances, and has the potential for a high failure rate depending on the skill of the field investigator (Sander et al. 1996).

Received: 23 August 2011 / Accepted: 6 July 2012
Published online: 19 August 2012

© Springer-Verlag 2012

S. Lee (✉) · K.-Y. Song
Geological Mapping Department,
Korea Institute of Geoscience & Mineral Resources (KIGAM),
124, Gwahang-no, Yuseong-gu, Daejeon 305-350, Korea
e-mail: leesaro@kigam.re.kr
Tel.: +82-42-8683057
Fax: +82-42-8683413

Y. Kim
Geogreen21 Co., Ltd.,
901 E&C Venture Dream TowerII 197-10, Guro-Dong Guro-Gu,
Seoul 152-719,
Republic of Korea

I. Park
Department of Geoinformatics,
University of Seoul,
Siripdae-gil 13, Dongdaemun-gu, Seoul 130-743,
Republic of Korea

However, the success rate can be improved using a geographic information system (GIS) database of hydraulic characteristics of aquifers such as the results of pumping tests by previous research, hydrogeological information on groundwater using a GIS spatial analysis method, and systematic probability statistics models (Engman and Gurney 1991; Jha et al. 2007). A reliable analysis method and model to predict the GPP areas are urgently needed for systematic development, efficient management, and sustainable uses of groundwater resources.

Groundwater is a dynamic system, influenced by combinations and interactions of various factors, including weather, hydrology, surface topography, and geologic characteristics (Park et al. 2001). To understand aquifer productivity and composite mechanisms in groundwater systems, the physical characteristics of the related factors that configure the system should be identified. While it is not possible to directly understand groundwater distribution using remote sensing and GIS technologies without field surveys, groundwater potential can be inferred from surface attributes such as geology, soil texture, land use, and drainage systems (Todd 1980; Jha and Peiffer 2006).

GIS and remote sensing technologies have great potential for use in groundwater potential analyses. Many studies have applied these techniques along with thematic layers such as those representing the geomorphology, drainage pattern, lineaments, lithology, and soil (Jaiswal et al. 2003; Solomon and Quiel 2006; Kim et al. 2010; Jasmin and Mallikarjuna 2011). Some studies have used personal judgments or local information to assign weight to different thematic layers and their features (Madrucci et al. 2008; Pradhan 2009; Yeh et al. 2009; Dar et al. 2010; Saud 2010; Singh et al. 2011). Other studies have applied probabilistic models such as frequency ratio and weights-of-evidence modelling for groundwater potential mapping (Corsini et al. 2009; Oh et al. 2011; Lee et al. 2012). Oh et al. (2011) and Lee et al. (2012) applied frequency ratio and a weight-of-evidence model for groundwater potential mapping and sensitivity analysis in Pohang City, Korea. More sophisticated assessments have been conducted using numerical modelling, decision trees, fuzzy logic, and analytic hierarchy process analysis (Srivastava and Bhattacharya 2006; Vijay et al. 2007; Murthy and Mamo 2009; Chenini and Mammou 2010; Kiesel et al. 2010). Some researchers have also integrated GIS, remote sensing, and geophysical surveys to derive additional thematic layers of surface parameters such as resistivity, aquifer thickness, or fault maps (Israil et al. 2006; Srivastava and Bhattacharya 2006; Ranganai and Ebinger 2008; Kumar et al. 2009). However, such studies have limitations because they use indirect indicators such as yield, groundwater depth, resistivity, and spring location rather than hydraulic constants such as specific capacity (SPC) and transmissivity. Furthermore, past studies generally produced small datasets and did not validate the results by comparison with other datasets.

In this study, an artificial neural network model (ANN) was applied to analyse and validate the groundwater

potential using a GIS with different training and test datasets of SPC for the area around Pohang City. GIS approaches provide a way to introduce information and knowledge from other data sources into the decision-making process, and aid in the handling and manipulation of classified remote sensing data (Adinarayana and Krishna 1996). Use of a GIS enables quantitative assessment of the consequences of heterogeneity in environmental systems over a broad range of spatial and temporal scales. Systematic integration of several surface features that indicate groundwater potentiality is an important aspect in water-management studies. A database designed to support water-resource decisions must contain various thematic information because of the interdisciplinary nature of water problems. Conventional methods of exploration do not always account for the diverse factors affecting the presence of groundwater (Murthy 2000). For this study in Pohang City, the result was then compared with the previous research of Oh et al. (2011) and Lee et al. (2012). GIS lacks prediction ability as it has no built-in function for interpretation of multi-dimensional data relating to individual incidents, or for exploration of the spatial patterns associated with the likely occurrence of future events. Therefore, it is necessary to have an applied technique for identifying spatial correlations or other spatial patterns using probability, statistics, and data mining approaches. This study conducts a regional GPP analysis by applying an ANN model in tandem with application of GIS and remote sensing techniques. The ANN model can provide a geospatial assessment tool to calculate the probabilistic relationship between a dependent variable and independent variables, including multi-classified maps. Although ANN models have been applied to many geosciences areas, this approach has not yet been used to delineate groundwater potential.

Materials and methods

Preparation of the GPP map using GIS involved four major steps: (1) assembly of a spatial database (75 SPC data were used to create a spatial database using GIS; topography, geology, lineament, and soil data were collected and input into a spatial database); (2) processing the data (SPC data were randomly split for training and testing and then used for analysing and validating GPP maps using the ANN model, training locations for ANN analysis were extracted from SPC data, the training dataset and various factors were analysed, and each factor's relative importance and weight were quantitatively determined); (3) application of weights to generate a GPP map; (4) validation of the GPP map using test SPC data that were not used directly in the analysis (Fig. 1).

Study area and hydrogeological setting

The study focuses on the Pohang City area of Korea. This area has experienced rapid population growth and increased demand for groundwater reserves. Thus, it is

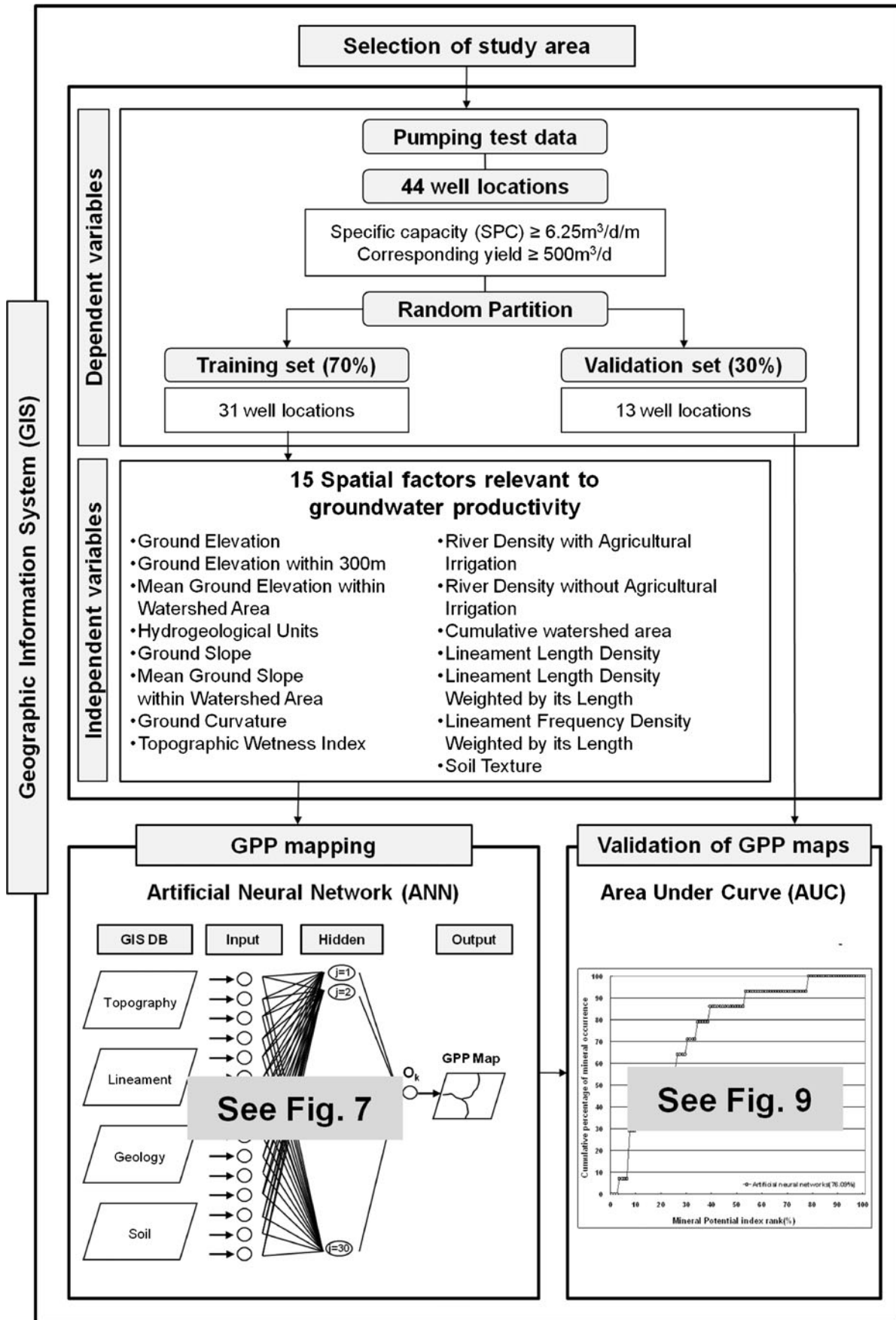


Fig. 1 Study flow for groundwater productivity potential (GPP) mapping

appropriate for evaluating groundwater potential. The study area lies between 35°50'07"N and 36°16'34"N latitude and 129°00'31"E and 129°34'57"E longitude (Fig. 2), and covers 891.44 km² on a digital topographic map (NGII 1997) with a 1:5000 scale (Table 1). Elevation in the area ranges from 0 to 919.3 m above sea level (a.s.l.), with an average of 144.2 m—SD(SD) =143.64 m). The terrain gradient computed from a 30 × 30 m digital elevation model (DEM) extracted from the 1:5,000-scale topographic map ranged from 0 to 50.4° with a mean value of 14.4° and a SD of 10.70°.

The bedrock geology of the study area consists mainly of Cretaceous sedimentary, granitic, and volcanic rocks; Tertiary granitic, volcanic, and sedimentary rocks; and Quaternary basalt (Hwang et al. 1996; Kim et al. 1998) (Fig. 3). The study area has two major faults: the Yangsan and Ulsan faults that trend N10°–20°E and N10°–20°W, respectively. In the study area, Cretaceous volcanic and granitic rocks are distributed mainly in the west, Tertiary sedimentary in the east, and Tertiary volcanic and granitic rocks in the south.

The study area was divided into two areas. The north-eastern area consists of coastal hills produced by Tertiary deposition, and the north-western area is a rugged mountain range consisting of volcanic rock and granite. The Taebaek mountain range (Mt. Bade, 645.0 m a.s.l.; Hyangnobong, 929.9 m a.s.l.; Satgatbong, 718.0 m a.s.l.; and Mt. Bihak, 762.0 m a.s.l.) lies along the Yangsan fault

zone. The main mountain system in the study area trends mostly N10°–20°E parallel to the Yangsan fault zone; the other small systems trend N60°–70°W parallel to faults that deviate from the Yangsan fault. Precipitation is most abundant from June to September. The mean annual rainfall between 1973 and 2009 was 1,119 mm.

Spatial database

To map GPP, the first stage was data collection and construction of the spatial database from which relevant factors were extracted (Table 1). Groundwater productivity data such as SPC, yield, well depth, well diameter, and depth/elevation of the water table were collected from the national groundwater survey in the Pohang area (MLTM 2003), the national groundwater monitoring network construction report (MLTM 1995, 1998, 2001), and the rural groundwater survey report (MIFAFF 1985–2005). SPC data were also used for the GPP. Fig. 4 describes spatial distribution and Table 2 shows statistics of the SPC data used in this study.

The SPC value is derived by the pumping rate divided by drawdown using pumping test results; pumping tests last more than 24 h. The well depth is distributed from 50 to 280 m, mean depth is 133 m, and SD is 45 m. Each discharge rate stands between 13 and 830 m³/d, mean discharge rate is 272 m³/d, and SD is 162 m³/d. Drawdown during each pumping test is between 3.49 m

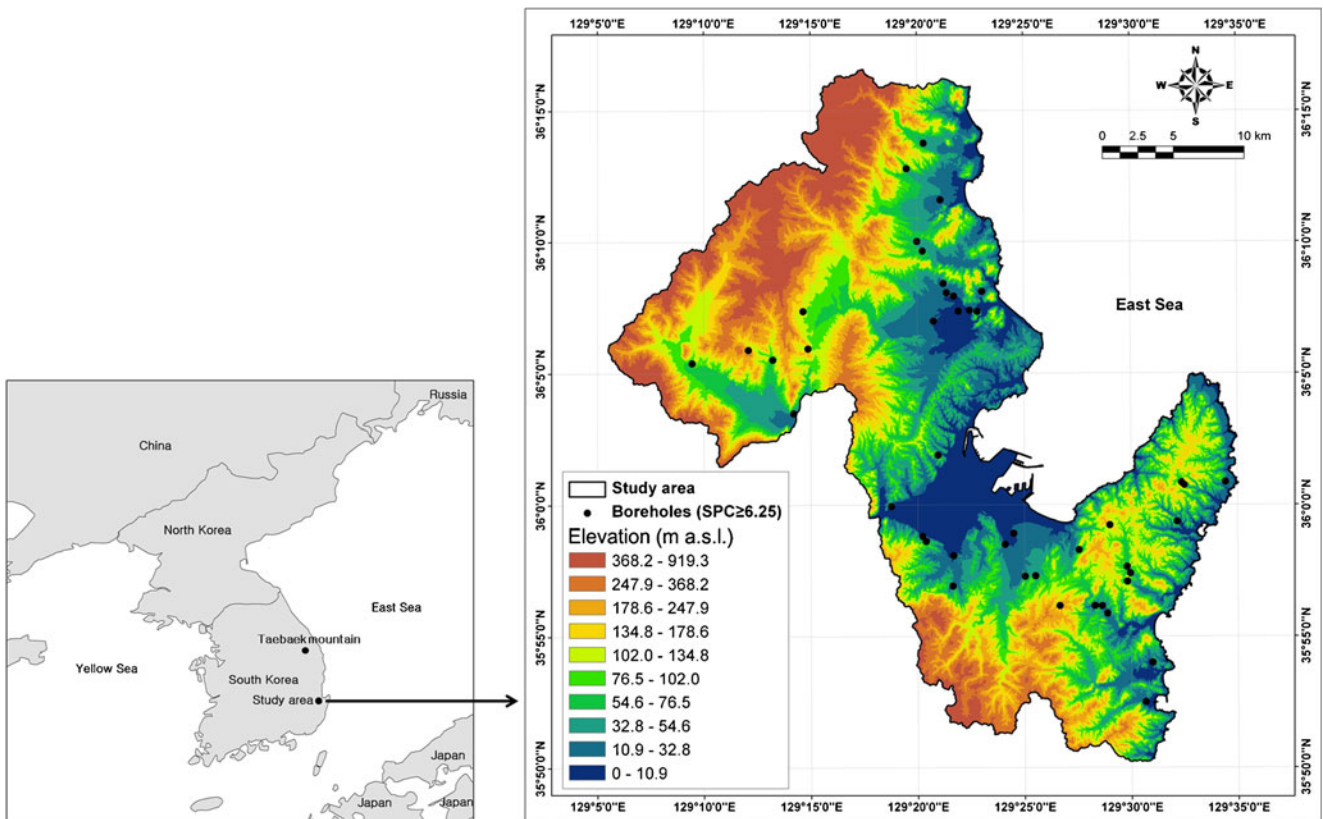


Fig. 2 Study area in Korea, showing land-surface elevation from a digital elevation model (DEM; Landsat TM, 2003). Where data ranges are given as, for example, 0–10.9, this means 0 to less than 10.9

Table 1 Hydrogeological factors related to groundwater productivity in this study. *Res.* refers to resolution

Category	Factors	Description	Unit	Data type	Scale
Topology ^a	Ground elevation	Ground elevation, obtained from DEM (Res. 30 m) for local ground elevation property. Negative relationship, in general	m a.s.l.	Grid	1:5,000
	Ground elevation within 300 m	Mean ground elevation in cells within radius 300 m, obtained from DEM for regional ground elevation property. Negative relationship, in general	m a.s.l.		
	Mean ground elevation within watershed area	Mean ground elevation within watershed area, obtained from DEM for hydrological regional ground elevation property. Negative relationship, in general	m a.s.l.		
	Ground slope	Ground slope, obtained from DEM (Res. 30 m) for local ground slope property. Negative relationship, in general	%		
	Mean ground slope within watershed area	Mean ground slope within watershed area, obtained from DEM for hydrological regional ground slope property. Negative relationship, in general	%		
	Ground curvature	Ground curvature, obtained from DEM (Res. 30 m) for local ground curvature property. Negative relationship, in general	–		
	Topographic wetness index	Topographic wetness index, obtained from DEM (Res. 30 m) for local ground wetness property. Positive relationship, in general	–		
	River density with agricultural irrigation	Density of river line feature for surface drainage property. Positive relationship, in general	km/km ²		
	River density without agricultural irrigation	Distance from river line feature for water supply from river. Negative relationship, in general	km/km ²		
Cumulative watershed area	Watershed area with accumulation through flow direction for water body scale. Positive relationship, in general	km ²			
Lineaments ^b	Lineament length density	Density of lineament length. Positive relationship, in general	km/km ²	Line	–
	Lineament length density weighted by its length	Density of lineament length weighted by its length. Positive relationship, in general	km/km ²		
	Lineament frequency density weighted by its length	Density of lineament frequency weighted by its length. Positive relationship, in general	count·km/km ²		
Geology ^c	Hydrogeological units	Hydrogeological units classified from bedrock lithology by similar hydrological property	–	Line	1:5,000
Soil ^d	Soil texture	Soil texture property for groundwater infiltration	–	Polygon	1: 25,000

^a Topological factor was extracted from digital topographic map (1:5,000 scale) by National Geographic Information Institute (NGII 1997)

^b The Landsat TM image of study area was acquired in March 2003

^c The geological map (1:50,000 scale) produced by the Korea Institute of Geoscience & Mineral Resource (KIGAM 1964)

^d The detailed soil map (1:25,000 scale) produced by National Academy of Agricultural Science (NAAS 1977)

and 137.50 m, mean drawdown is 41.17 m, and SD is 34.69 m. Groundwater level is measured 2 days before the pumping test and the drawdown is measured automatically using a pressure transducer (DIVER), during pumping and the recovery period.

In general, groundwater productivity is governed by many hydrogeological factors including surface and bedrock lithology, structure, slope steepness and morphology, stream evolution, climate, soil, vegetation cover, land use, and human activity, but these relationships have not been statistically or quantitatively verified. In this study, 15 hydrogeological factors expected to be related to groundwater potential were reviewed using SPC. Finally, 15 hydrogeological factors (Table 1) were selected and applied to the ANN model.

The spatial database was constructed with a resolution of 30 × 30 m on the basis of Landsat Thematic Mapper (TM) imagery. The 15 hydrogeological factors were

converted to ArcGIS grid format, and the GRID set comprised 1,822 rows × 1,787 columns. In the study area, the total number of cells was 990,495 with SPC in 75 cells (62 cells for training and 13 cells for validation).

Topographic factors

A triangulated irregular network was made using the elevation value, and a DEM was made with 30 × 30 m resolution after elimination of internal drainage ('sink area') in the elevation grid using the FILL command in ArcGIS 9.0. One input factor, the 'ground elevation within 300 m', was given a mean value for each grid cell based on the value of neighbouring cells within a 300-m radius. Using the DEM, the slope, curvature, and topographic wetness index (TWI) were calculated. The mean ground elevation and slope within the watershed area were obtained after watershed delineation.

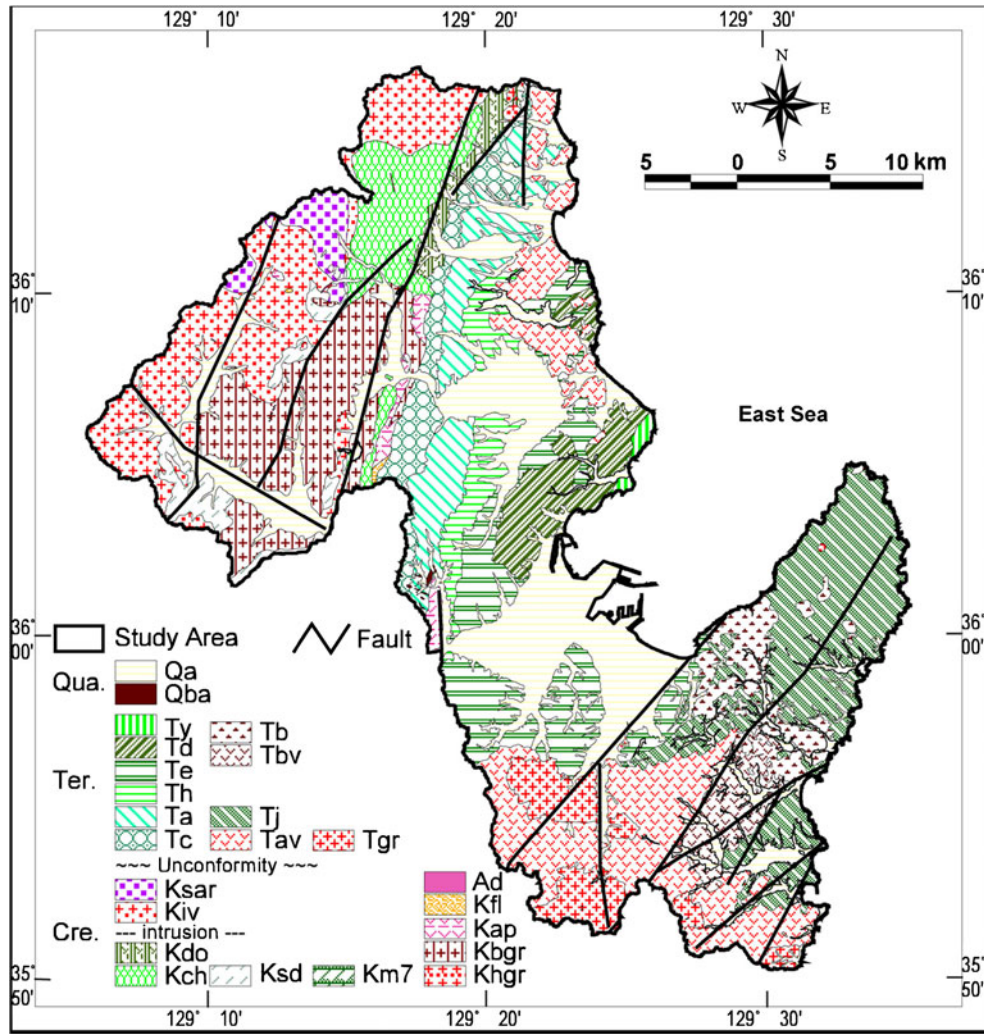


Fig. 3 Geological map of the study area

The river density and distance from the river were considered the prime indicators for selection of groundwater potential sites because they are related to the yield of groundwater. For the same permeable aquifer, the closer to the river, the higher the groundwater capacity. The topographic factors are shown in Fig. 5.

Hydrogeology and lineament factors

The study area had 24 geological units of bedrock. With the exception of alluvial units, these units were classified into six geological groups with similar hydrogeological properties such as porous volcanics, semi-consolidated sediments 1 and 2, non-porous volcanics, intrusives, and clastic sediments (Fig. 6). These hydrogeological units were identified by groundwater productivity statistics, spatial distribution, and stratigraphic history (Table 3). The semi-consolidated sediments 1, consisting predominantly of sandstone and conglomerate, were overlain by the semi-consolidated sediments 2, consisting mainly of mudstone and shale.

In an image, a lineament is defined as a straight or slightly curved surface feature of natural origin, interpreted directly

from the imagery (O’Leary et al. 1976; Koike et al. 1998). Lineaments and intersections of lineaments reflect rock structures through which water can percolate and travel as far as several kilometres within. Depending upon the terrain, the lineament may be a zone of influence. To predict ground control points in underground structures (Kane et al. 1996), lineaments are strongly related to discontinuities such as joints, faults, and folds. For these reasons, the feature ‘lineament’ was used for structural analysis, analysis of the relationship with lithology, and assessment of groundwater productivity. In this study, lineaments were detected through interpretation of Landsat TM imagery and hillshade maps from a DEM made by a structural geologist with much experience in such interpretation. To eliminate errors resulting from scale deviation and sun illumination, the DEM data were used. The hillshade maps from the DEM were used to detect the lineaments. In the hillshade maps, the sun altitude is 45°, and the sun azimuths are 0, 45, 90, 135, 180, 225, 270, and 315°. The Landsat TM image was then overlaid onto the hillshade maps, and the lineaments were detected.

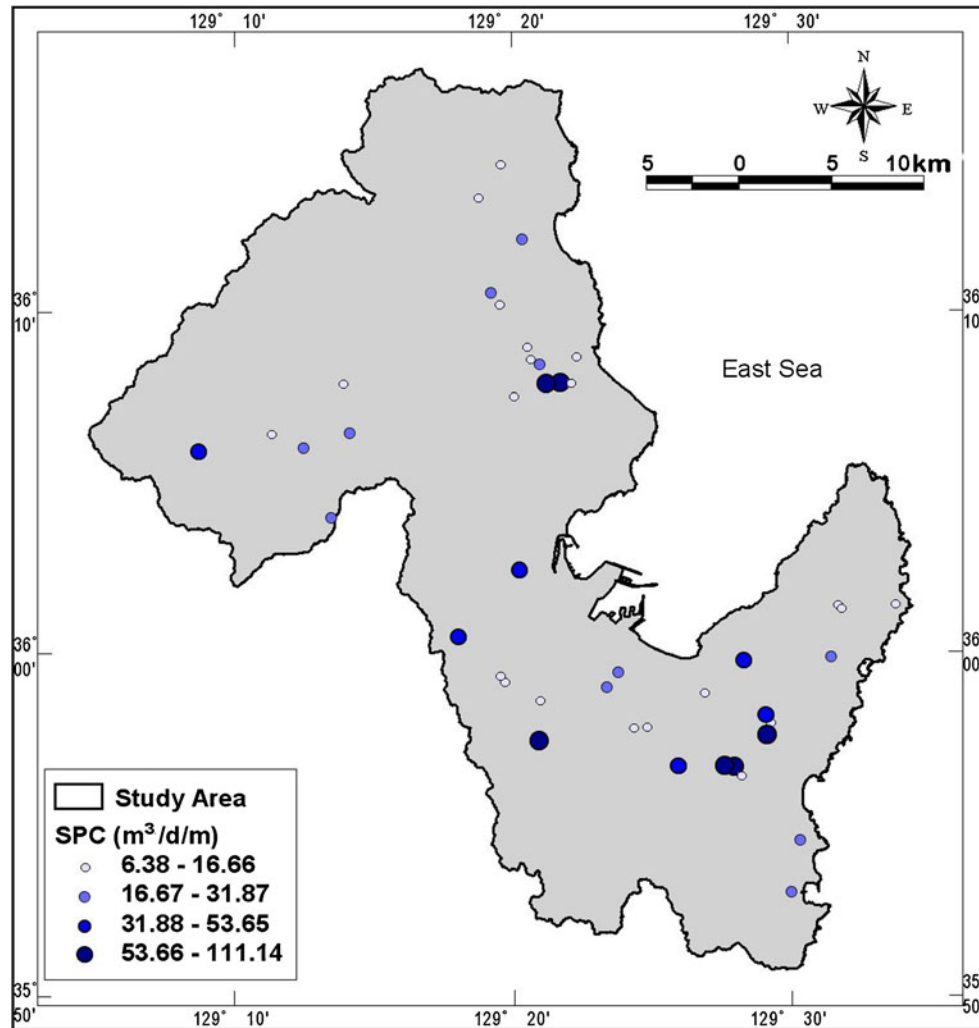


Fig. 4 Location and value range of groundwater productivity data (SPC specific capacity)

Finally, a digital topographic map is used with a scale of 1:5000 to exclude drainage and roads, both of which could be detected as lineament. In this study, lineament density was applied to spatial relationship regarding lineament factors. Lineament density is useful for understanding the local distribution of lineaments. Lineament density analysis considered the lineament frequency and length, and the densities of lineament length and frequency were also analyzed using the weight of lineament length in the study area (Fig. 6). The density distribution of lineaments was calculated using Hardcastle(1995)'s circular grid with 1.41 km radius and 2 km grid spacing. The circular grids overlapped about 30–50 % and each circle included 10~15 lineaments. In order to calculate the expectation value of lineament existence for each cell, the density value of the cells was divided by the sum of density of all cells.

Soil factor

Regarding land surface data, the soil was used as a factor related to groundwater potential. Soil texture invariably

controls the penetration of surface water into an aquifer system. The soil texture of the study area was generated by a 1:25,000 scale soil map published by the National Institute of Agricultural Science and Technology (1977). Fourteen different categories of soil were extracted from the soil map: forest, grassland, gravel, gravelly sandy loam, fine gravelly sandy loam, sandy loam, loamy fine sand, fine sand, loam, gravelly loam, silt loam, gravelly silt loam, clay silty loam, and silty clay loam (Fig. 6f).

Theory

An artificial neural network (ANN) is a 'computational mechanism able to acquire, represent, and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping' (Garrett 1994). The ANN model offers a number of advantages, including ability to analyze complex patterns quickly and with a high degree of accuracy. In addition, the ANN makes no assumption about the nature of the distribution of the data. Consequently, when the relationship

Table 2 Statistics of groundwater productivity data for the study area

Type	No.	Min.	Max.	Average	Median
SPC (m ³ /d/m)	83	1.42	111.14	16.54	6.73

between the variables does not fit an assumed model, better results can be expected with artificial neural networks (Livingstone et al. 1997).

The back-propagation training algorithm is the most frequently used neural network method, and was also used in this study. The back-propagation training algorithm is trained using a set of examples of associated input and output values. The purpose of an ANN is to build a model of the data-generating process so that the network can generalise and predict outputs from inputs that it has not previously seen. This learning algorithm is a multi-layered

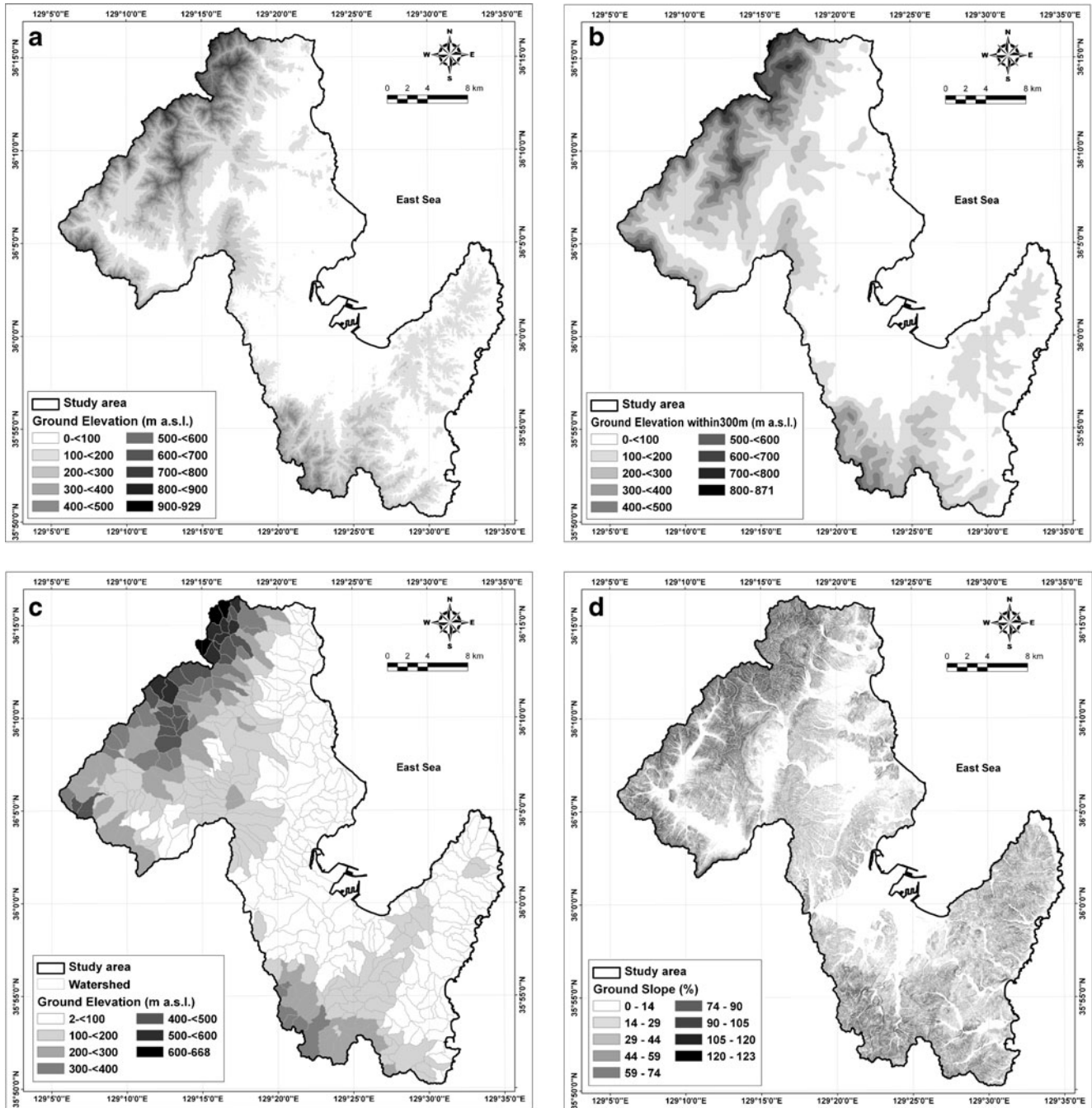


Fig. 5 Topographic factors. **a** Ground elevation, **b** ground elevation within 300 m (mean value for each grid cell based on the value of neighbouring cells within a 300-m radius), **c** mean ground elevation within watershed, **d** ground slope, **e** mean ground slope within watershed, **f** ground curvature, **g** topographic wetness index (TWI), **h** river density without agricultural irrigation, **i** distance from river, **j** cumulative watershed area

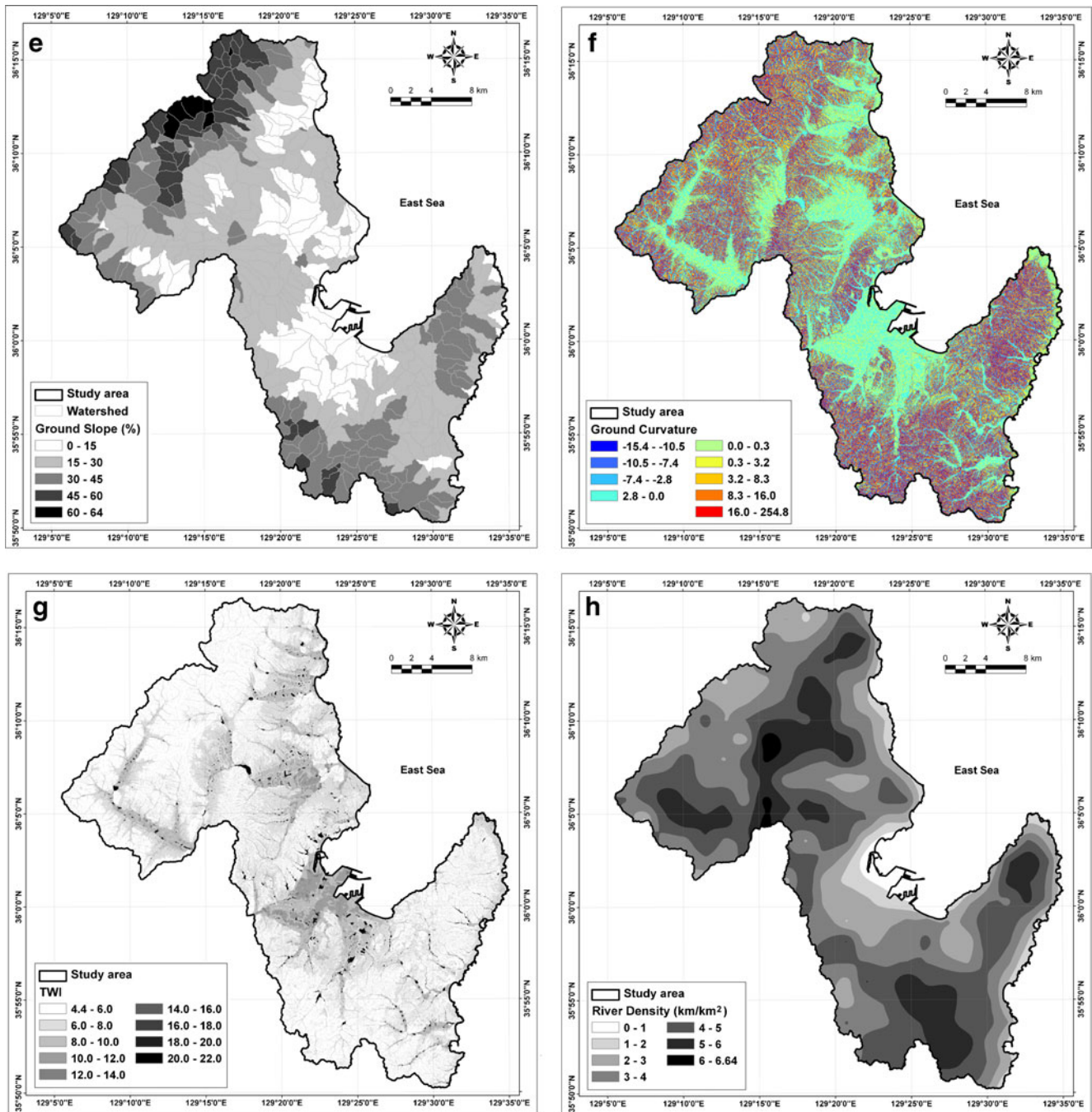


Fig. 5 (continued)

neural network composed of an input layer, hidden layers, and output layer. The hidden and output layer neurons process their inputs by multiplying each input by a corresponding weight, summing the product, and then processing the sum using a nonlinear transfer function to produce a result (Fig. 7). An ANN ‘learns’ by adjusting the weights between the neurons in response to the errors between the actual output values and the target output values. At the end of this training phase, the neural network provides a model that should be able to predict a target value from a given input value.

Two stages are involved in using a neural network for multi-source classification: the training stage, in which the internal weights are adjusted, and the classification stage. Typically, the back-propagation algorithm trains the network until some target minimal error between the desired and actual output values of the network is achieved. Once the training is complete, the network is used as a feed-forward structure to produce a classification for the entire dataset (Paola and Schowengerdt 1995).

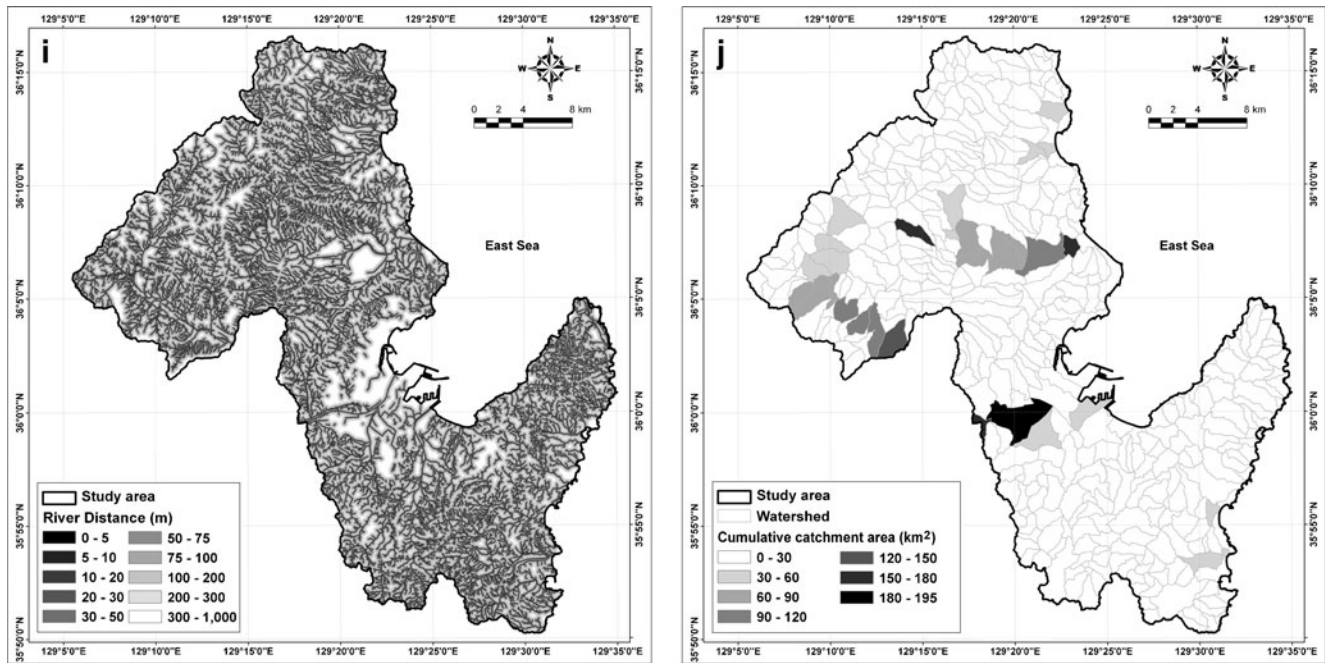


Fig. 5 (continued)

Procedure

A neural network consists of a number of interconnected nodes, where each node is a simple processing element that responds to the weighted inputs it receives from other nodes. The arrangement of the nodes is referred to as the network architecture (Fig. 7). The receiving node sums the weighted signals from all the nodes connected to it in the preceding layer. The input that a single node receives is weighted. The network used in this study consisted of three layers. The first was the input layer, where the nodes were the elements of a feature vector; the second was the internal or “hidden” layer; the third was the output layer that presented the output data. Each node in the hidden layer is interconnected to nodes in both the preceding and following layers by weighted connections (Atkinson and Tatnall 1997). Using the back-propagation training algorithm, the weights of each factor can be determined and may be used for the classification of data (input vectors) that the network has not seen before. Zhou (1999) described a method for determining the weights using back-propagation.

The 15 hydrogeological factors were used as the input data. Locations considered likely and unlikely to have GPP were selected as training sites. In an artificial neural network, the selection of training sites is important, and the selection of likely and unlikely GPP areas was carefully considered for training in this study. To select the training sites based on scientific and objective criteria, the SPC criterion of $6.25 \text{ m}^3/\text{d}/\text{m}$ was used. The average yield of well data is $274 \text{ m}^3/\text{d}$ and average well depth is 130 m. The median value of transmissivity (T) ($3.79 \text{ m}^2/\text{d}$) was taken as the criterion for division purposes. From the relationship between T and SPC, $T=0.9948 \times 0.7293 \text{ SPC}$,

$\text{SPC}=6.25 \text{ m}^3/\text{d}/\text{m}$ was taken as the criterion for division purposes. These criteria, $T=3.79 \text{ m}^2/\text{d}$ and $\text{SPC}=6.25 \text{ m}^3/\text{d}/\text{m}$, are equivalent to well yield $500 \text{ m}^3/\text{d}$, with supposed drawdown of $2/3$ of the total well depth. Then, based on $\text{SPC}=6.25 \text{ m}^3/\text{d}/\text{m}$, areas with SPC data greater than $6.25 \text{ m}^3/\text{d}/\text{m}$ were classified as the ‘likely GPP’ training dataset, and areas with SPC data less than $6.25 \text{ m}^3/\text{d}/\text{m}$ were classified into an ‘unlikely GPP’ dataset. Next, 70 % of SPC values were classified into a ‘likely GPP’ training dataset that was randomly selected and used for training. The remaining 30 % of SPC values were used for validation.

The back-propagation algorithm was then applied to calculate the weights between the input and hidden layers and between the hidden and output layers. A three-layered, feed-forward network was implemented using the MATLAB software package based on the framework provided by Hines (1997). Hines (1997) tried different numbers of hidden layers to analyse the relationship between the number of hidden layers and the training cycle to reach the error goal. In the case of twice the number of factors, the training cycle showed the lowest rate of success in reaching the goal. The number of hidden layers and the number of nodes in a hidden layer required for a particular classification problem are not easy to deduce. In this study, a $15 \times 30 \times 1$ (number of input, hidden, and output layers) structure was selected for the network, with input data normalised in the range of 0.1 to 0.9. The nominal and interval class group data were converted to continuous values ranging between 0.1 and 0.9. Therefore, the continuous values were not ordinal data but nominal data, and the numbers denote the classification of the input data. The learning rate was set at 0.01, and the initial weights were randomly set at values between 0.1 and 0.3. The weights

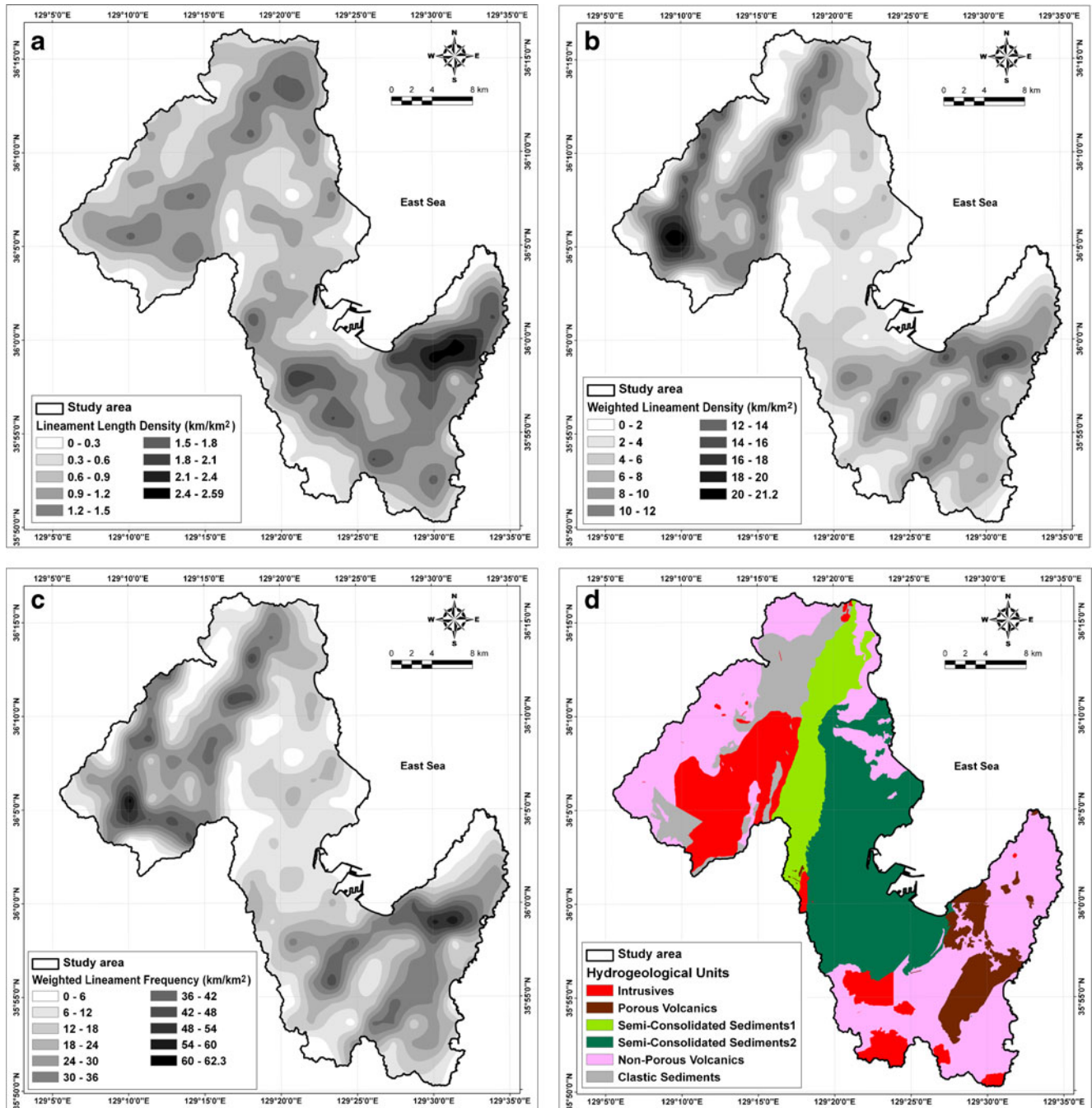


Fig. 6 Lineament, geologic and soil factors. **a** Lineament length density, **b** lineament length density weighted by its length, **c** lineament frequency weighted by its length, **d** hydrogeological units, **e** lineament distribution, **f** soil factors

calculated from 10 test cases were compared to determine whether the variation in the final weights depended on the selection of the initial weights.

The model was trained for 2000 epochs, and the root mean square error (RMSE) value used for the stopping criterion was set at 0.01. If this RMSE value was not achieved, then the maximum number of iterations was terminated at 2000 epochs. When the latter case occurred, the maximum RMSE value was 0.366. The final weights between layers acquired during training of the neural network and the contribution or

importance of each of the 15 factors was used to predict GPP. Finally, the weights were applied to the entire study area, and GPP maps were created for each training case.

Results

Weight determination and GPP mapping

The final weights between layers acquired during training of the neural network and the contribution or importance

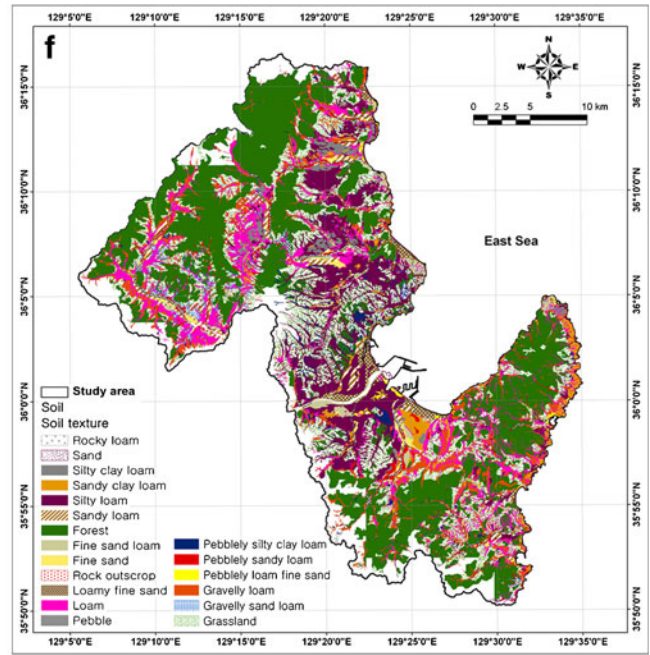
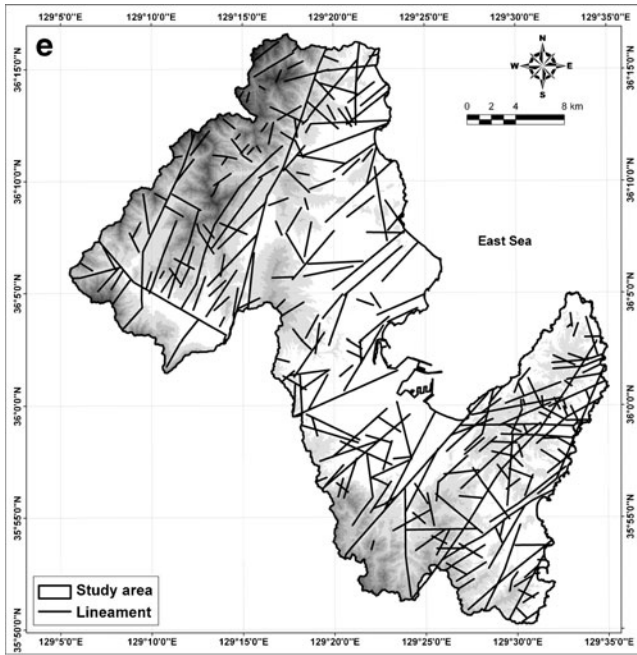


Fig. 6 (continued)

of each of the 15 factors used to predict GPP are shown in Table 4. The results were not the same because the initial weights were assigned random values. Therefore, in this study, the calculations were repeated 10 times to allow the results to achieve similar values. The SD of the results

Table 3 Geology of bedrock and hydrogeological units of the study areat

Age	Geological units	Hydrogeological units	
Quaternary	Qba	Basalt	
Tertiary	Yonil series	Ty	Yonam form.
		Td	Duho form.
		Te	Idong form.
		Th	Hunghae form.
		Ta	Hakrim form.
	Janggi series	Tc	Chunbuk Cong.
		Tb	Basalt
		Tbv	Alkalic volc.
		Tj	Janggi series
		Tav	Acidic volc.
Cretaceous	Yucheon Group	Tgr	Granite
		Ksar	Acidic volc.
	Bulgugsa Series	Kiv	Volc.
		Ad	Dyke
	Hayang group	Kfl	Syenite
		Kap	Acidic porphyrite
		Kbgr	Biotite granite
		Khgr	Hornblende granite
		Kdo	Dochunri member
		Kch	Cheonggyeri member
Ksd	Daegu form.		

ranged from 0.006 to 0.012; therefore, random sampling did not have a large effect on the results. For easy interpretation, the average values were calculated, and these values were divided by the average of the weights of the factor that had a minimum value. In the weight, the ‘hydrogeological unit’ was the maximum value (1.430) and the ‘curvature’ value was the minimum value (1.000).

The weights were applied to the entire study area, and the GPP map was created (Fig. 8). The GPP values were classified by equal areas and grouped into four classes (% of area) of GPP rank for easy visual interpretation: very high (5 %), high (10 %), medium (15 %) and low (70 %). The minimum and maximum GPP values were 0.0764 and 0.9075, respectively. The mean and SD were 0.4560 and 0.1648, respectively.

Validation

The GPP map is expected to effectively predict future GPP areas. This map can be further validated using new exploration locations as they occur. Here, the result of the GPP analysis was validated using test SPC data that were not used for the analysis. Validation was performed by comparing the test SPC data with the GPP map. For this, success rate curves were calculated for quantitative prediction, and the area under the curve was calculated. The rate shows how well the model and factors predict the GPP, and the area beneath the curve thus qualitatively assesses the prediction accuracy. To obtain the relative ranking for each prediction pattern, the calculated index values of all pixels in the study area were sorted in descending order. The ordered pixel values were then divided into 100 classes with accumulated 1 % intervals. The rate validation results appear as a line in Fig. 9. As a result, the 80–100 % class (20 %) in which the

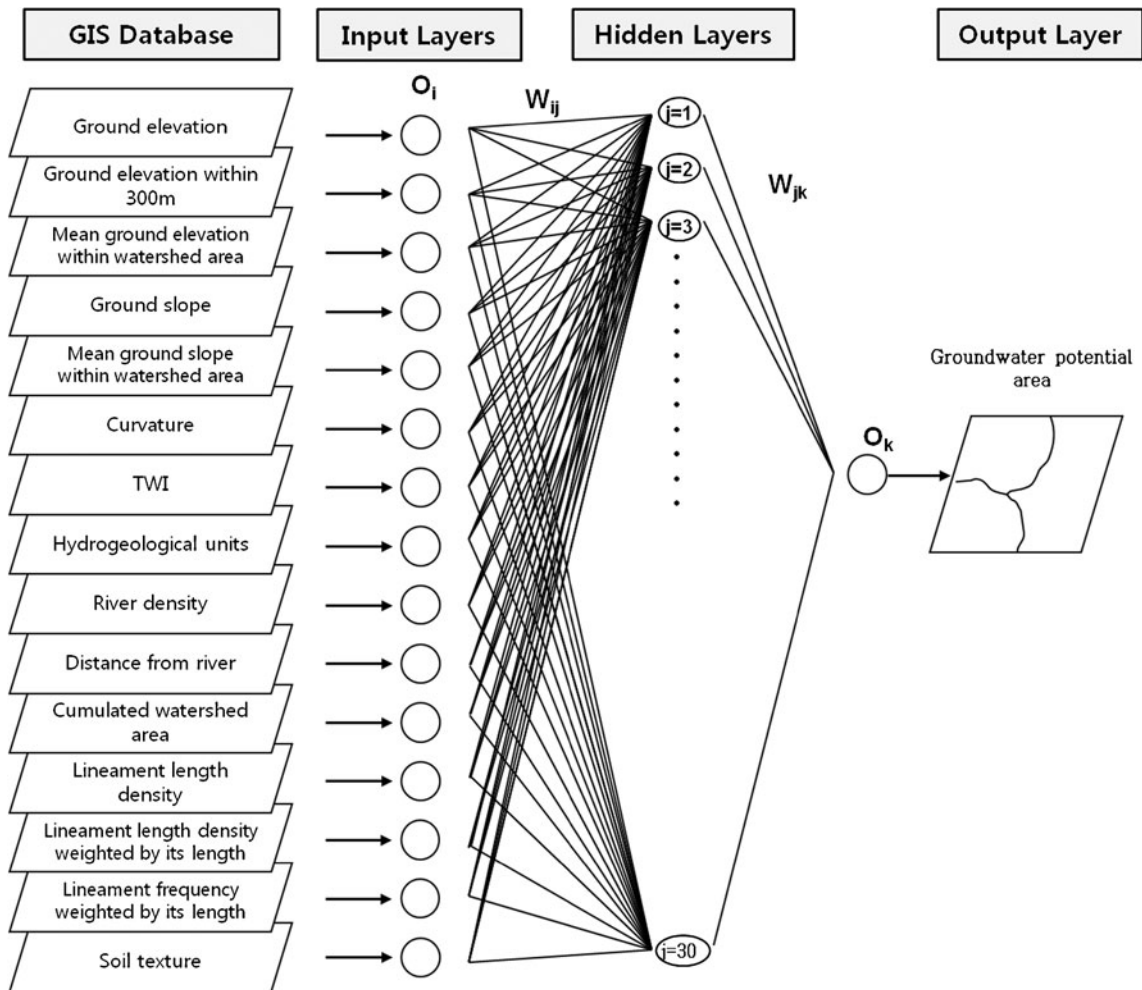


Fig. 7 Architecture of artificial neural network

groundwater potential index had a high rank could explain between 46 and 62 % of all occurrences of $SPC \geq 6.25 \text{ m}^3/\text{d}$

m. The graph showing the best prediction accuracy was selected from among the 10 runs.

Table 4 Weights of hydrogeological factors considered in the GPP analysis based on each case

Factor	No.										Average	SD	Weight
	1	2	3	4	5	6	7	8	9	10			
Ground elevation	0.061	0.063	0.064	0.070	0.066	0.073	0.051	0.077	0.066	0.064	0.065	0.007	1.100
Ground elevation within 300 m	0.063	0.064	0.074	0.070	0.081	0.078	0.046	0.054	0.048	0.062	0.064	0.012	1.074
Mean ground elevation within watershed area	0.056	0.052	0.070	0.062	0.069	0.059	0.069	0.061	0.068	0.050	0.061	0.007	1.033
Hydrogeological units	0.093	0.087	0.082	0.076	0.084	0.086	0.096	0.071	0.090	0.086	0.085	0.007	1.430
Ground Slope	0.075	0.061	0.069	0.063	0.084	0.067	0.076	0.062	0.071	0.057	0.068	0.008	1.149
Mean ground slope within watershed area	0.060	0.070	0.063	0.070	0.062	0.054	0.075	0.062	0.066	0.071	0.065	0.006	1.096
Ground curvature	0.051	0.060	0.055	0.050	0.058	0.061	0.066	0.071	0.065	0.058	0.059	0.007	1.000
Topographic wetness index	0.062	0.070	0.049	0.075	0.045	0.055	0.052	0.065	0.065	0.070	0.061	0.010	1.023
River density with agricultural irrigation	0.066	0.057	0.066	0.066	0.049	0.059	0.069	0.073	0.071	0.055	0.063	0.008	1.063
River density without agricultural irrigation	0.071	0.083	0.072	0.070	0.068	0.080	0.070	0.057	0.068	0.085	0.072	0.008	1.215
Cumulative watershed area	0.059	0.062	0.067	0.053	0.085	0.066	0.064	0.061	0.078	0.060	0.066	0.009	1.102
Lineament length density	0.077	0.072	0.074	0.070	0.057	0.081	0.076	0.074	0.063	0.084	0.073	0.008	1.222
Lineament length density weighted by its length	0.054	0.083	0.064	0.063	0.057	0.066	0.065	0.069	0.062	0.082	0.067	0.009	1.120
Lineament frequency density weighted by its length	0.078	0.059	0.064	0.070	0.070	0.058	0.056	0.078	0.064	0.059	0.066	0.008	1.102
Soil texture	0.075	0.059	0.068	0.074	0.064	0.059	0.068	0.064	0.057	0.059	0.064	0.006	1.084

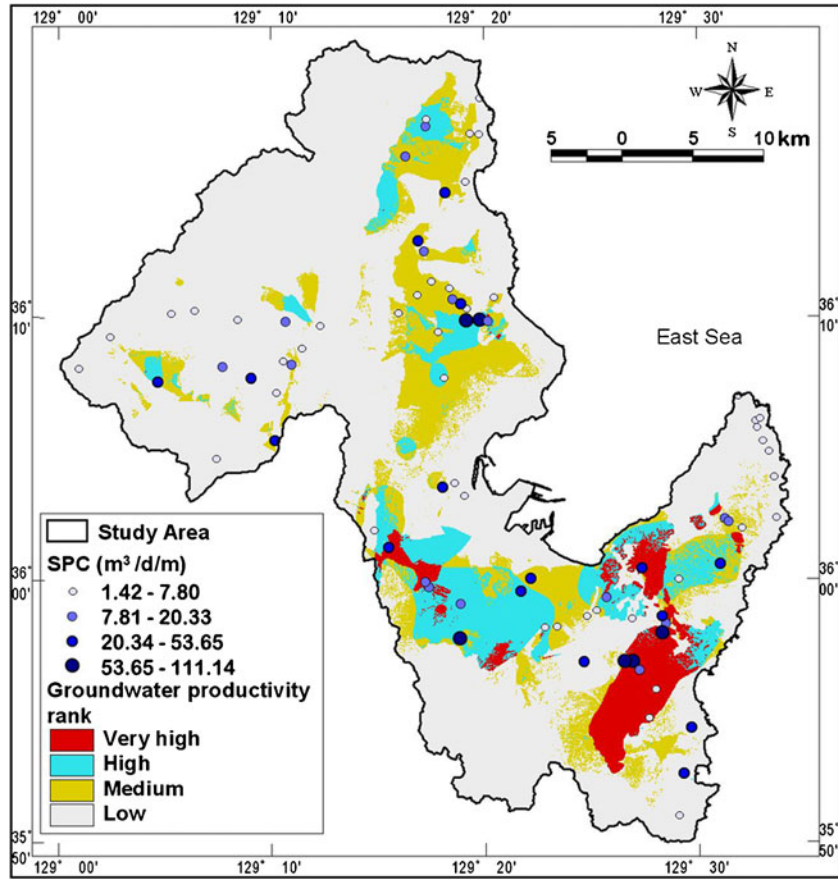


Fig. 8 Groundwater productivity potential (GPP) map

To compare the result quantitatively, the areas under the curve (Lee and Dan 2005) were re-calculated as if the

total area were one, which indicates perfect prediction accuracy. The resulting area ratios, suggesting the

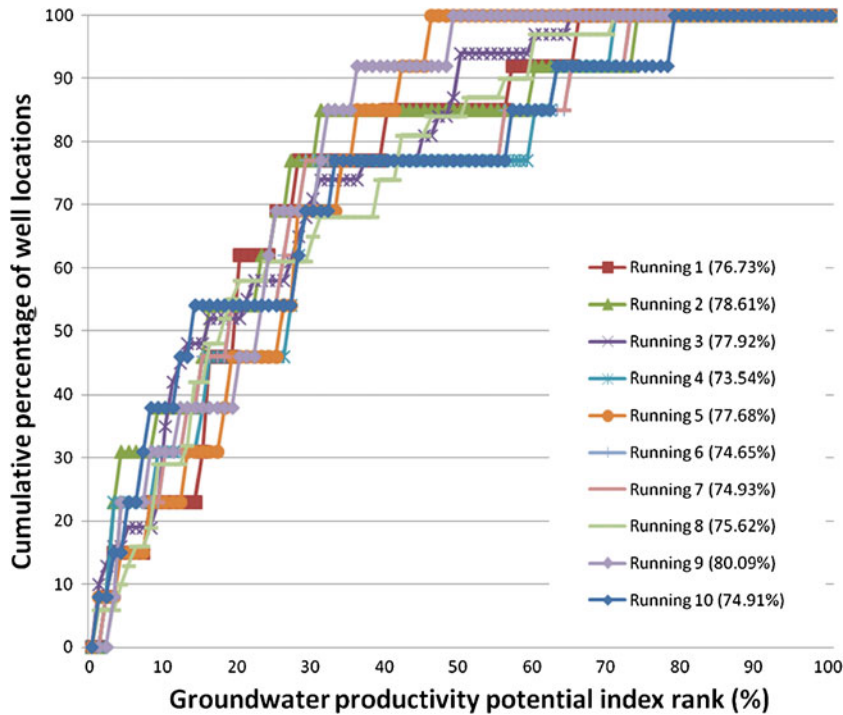


Fig. 9 Success rate curve of GPP index by ANN

prediction accuracy, were between 0.7354 and 0.8009; thus, the prediction accuracy was between 73.54 and 80.09 %, with an average accuracy of 76.47 %.

Discussion

In this study, the GPP maps were made using an ANN and repeated 10 times. Training sites were extracted from SPC data. The validation showed between 73.54 and 80.09 % prediction accuracy, with an average of 76.47 %. The accuracies were similar, and they were satisfactory considering the map scale and input data accuracy. Additionally, using the artificial neural network, the relative importance and weight of factors were calculated. The ‘hydrogeological units’ showed the highest weight value (1.430), followed by the ‘lineament length density’ with a value of 1.222. The ‘curvature’ showed the lowest value at 1.000, and the ‘topographic wetness index (TWI)’ was 1.023. These results indicate that the ‘hydrogeological unit’ was the most important factor and was 1.4 times more important than ‘curvature’ in GPP mapping. In particular, most of the very high potential was distributed in areas of porous volcanics, and low-potential areas were distributed in hydrogeology units of non-porous volcanics and intrusives in this study (see Fig. 8 and Fig. 6d).

Oh et al. (2011) and Lee et al. (2012) applied the frequency-ratio and weight-of-evidence method for groundwater potential mapping and sensitivity analysis in the same area. A comparison of the result maps developed by Oh et al. (2011), case 1 ($SPC \geq 6.25 \text{ m}^3/\text{d}/\text{m}$) of Lee et al. (2012) and Fig. 8 revealed that the overall distributions of the three result maps showed similar patterns. In the central part, the distribution of groundwater potential indices showed a pattern similar to that of Oh et al. (2011) and Lee et al. (2012), and consisted mainly of very high and high indices. The northeast part of the site was expected to represent high and very high potential based on the frequency-ratio model; however, that area was expected to have high and medium potential based on the ANN model. In contrast, the southeast part was expected to have very high and high potential using the results from this study. These differences in spatial distribution can be explained by correlations among factors used in the studies. For these three models, the frequency-ratio model demonstrated 77.78 % accuracy; the weight-of-evidence model described 71.20 %; the ANN model exhibited 76.47 % accuracy on average. The frequency ratio model showed the best result of groundwater productivity. However, the highest prediction accuracy of 10 times training using ANN model showed 80.09 %. In some cases, the ANN model can demonstrate higher accuracy than the frequency ratio model.

Conclusion

Groundwater plays an increasingly important role in both private and public water supplies worldwide. Some areas in Korea rely on a water-supply system that uses only

surface water. These areas need an alternative, stable water acquisition system that can provide high-quality reliable drinking water. Therefore, this study developed an ANN approach that used GIS to estimate a region’s potential for groundwater resources.

The primary value of the results is that even with some incomplete data sets and broad assumptions, the method proved to be a robust and useful tool for estimating and mapping productivity potential mapping. Also, Oh et al. (2011) conducted a sensitivity analysis using the frequency-ratio model. According to the result of sensitivity analysis (Table 4), in order of influence, soil texture, cumulative watershed area, and distance from river had a positive influence on the groundwater-potential maps. In contrast, using all factors (including ground elevation, curvature, ground elevation within 300 m) had a negative influence on the groundwater-potential maps. In the ANN model, the weights of factors represent the importance of each factor. The calculated weights indicate that the hydrogeological units ‘river density without agricultural irrigation’ and ‘lineament length density’ were more important factors than ground curvature, topographic wetness index, and mean ground elevation within the watershed area, in terms of their effect on the groundwater-potential maps in this study. The result maps from each model reflected the influences of important factors.

The proposed method of GPP mapping can be applied in planning and managing groundwater use, such as in regional groundwater development planning, determination of promising groundwater development areas in emergency situations, and control of the water-supply system based on a systematic, objective, and scientific decision model. In addition, the GPP map can greatly help planners seeking suitable locations at which to implement exploration. The method used in this study is also valid for generalised planning and assessment purposes. However, the method might be less useful at the site-specific level, at which local geologic and geographic heterogeneities may prevail. For the method to be more generally applied, more groundwater productivity data are needed, and the method must be applied to more case studies.

Acknowledgements This research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Ministry of Knowledge and Economy of Korea.

References

- Adinarayana J, Krishna NR (1996) Integration of multi-seasonal remotely-sensed images for improved landuse classification of a hilly watershed using geographical information systems. *Int J Remote Sens* 17:1679–1688
- Atkinson PM, Tatnall ARL (1997) Neural networks in remote sensing. *Int J Remote Sens* 18:699–709
- Chenini I, Mammou AB (2010) Groundwater recharge study in arid region: an approach using GIS techniques and numerical modeling. *Comput Geosci* 36:801–817
- Corsini A, Cervi F, Ronchetti F (2009) Weight of evidence and artificial neural networks for potential groundwater spring

- mapping: an application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology* 111:79–87
- Dar IA, Sankar K, Dar MA (2010) Remote sensing technology and geographic information system modeling: an integrated approach towards the mapping of groundwater potential zones in Hardrock terrain, Mamundiyan basin. *J Hydrol* 394:285–295
- Engman ET, Gurney RJ (1991) *Remote sensing in hydrology*. Chapman and Hall, London
- Garrett JH (1994) Where and why artificial neural networks are applicable in civil engineering. *J Comput Civil Eng* 8(2):129–130
- Gleick PH (1993) *Water in crisis: a guide to the world's fresh water resources*. Oxford University Press, New York
- Hardcastle KC (1995) Photolineament factor: a new computer-aided method for remotely sensing the degree to which bedrock is fractured. *Photogramm Eng Rem S* 61:739–747
- Hines JW (1997) *Fuzzy and neural approaches in engineering*. Wiley, New York
- Hwang JH, Kim DH, Cho DL, Song KY (1996) Explanatory note of the Andong sheet. Korea Institute of Geoscience and Mineral Resources, Daejeon, Korea
- Israil M, Al-hadithi M, Singhal DC (2006) Application of resistivity survey and geographical information system (GIS) analysis for hydrogeological zoning of a piedmont area, Himalayan foothill region, India. *Hydrogeol J* 14:753–759
- Jaiswal RK, Mukherjee S, Krishnamurthy J, Saxena R (2003) Role of remote sensing and GIS techniques for generation of groundwater prospect zones towards rural development: an approach. *Int J Remote Sens* 24:993–1008
- Jasmin I, Mallikarjuna P (2011) Review: satellite-based remote sensing and geographic information systems and their application in the assessment of groundwater potential, with particular reference to India. *Hydrogeol J*. doi:10.1007/s10040-011-0712-7
- Jha MK, Peiffer S (2006) *Applications of remote sensing and GIS technologies in groundwater hydrology: past, present and future*. Bayreuth University Press, Bayreuth, Germany
- Jha MK, Chowdhury A, Chowdary VM, Peiffer S (2007) Groundwater management and development by integrated remote sensing and geographic information systems: prospects and constraints. *Water Resour Manag* 21:427–467
- Kane WF, Peters DC, Speirer RA (1996) Remote sensing in investigation of engineered underground structures. *J Geotech Eng* 122:674–681
- Kenny JF, Barber NL, Hutson SS, Linsey KS, Lovelace JK, Maupin MA (2009) Estimated use of water in the United States in 2005. *US Geol Surv Circ* 1344, 52 pp
- Kiesel J, Fohrer N, Schmalz B, White MJ (2010) Incorporating landscape depressions and tile drainages of a northern German lowland catchment into a semi-distributed model. *Hydrol Process* 24:1472–1486
- Kim DH, Hwang JH, Park KH, Song KY (1998) Explanatory note of the Busan sheet. Korea Institute of Geoscience and Mineral Resources, Daejeon, Korea
- Kim NW, Lee JW, Lee J, Lee JE (2010) SWAT application to estimate design runoff curve number for South Korean conditions. *Hydrol Process* 24:2156–2170
- Koike K, Nagano S, Kawaba K (1998) Construction and analysis of interpreted fracture planes through combination of satellite-image derived lineaments and digital elevation model data. *Comput Geosci* 24:573–583
- Korea Institute of Geoscience and Mineral Resource (1964) *Geological map, 1:50,000*. KIGMR, Daejeon, Korea
- Korean Ministry for Food, Agriculture, Forestry and Fisheries (1985–2005) *Rural groundwater survey report*. KMFAFF, Seoul
- Korean Ministry of Land, Transport and Maritime Affairs (1995) *National groundwater monitoring network construction report*. KMLMA, Seoul
- Korean Ministry of Land, Transport and Maritime Affairs (1998) *National groundwater monitoring network construction report*. KMLMA, Seoul
- Korean Ministry of Land, Transport and Maritime Affairs (2001) *National groundwater monitoring network construction report*. KMLMA, Seoul
- Korean Ministry of Land, Transport and Maritime Affairs (2003) *Report of groundwater in Pohang area*. KMLMA, Seoul
- Korean Ministry of Land, Transport and Maritime Affairs (2006) *National groundwater monitoring network construction report*. KMLMA, Seoul
- Korean Ministry of Land, Transport and Maritime Affairs (2009) *National groundwater monitoring network construction report*. KMLMA, Seoul
- Kumar MG, Bali R, Agarwal AK (2009) Integration of remote sensing and electrical sounding data for hydrogeological exploration: a case study of Bakhar watershed, India. *Hydrolog Sci J* 54:949–960
- Lee S, Dan NT (2005) Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides. *Environ Geol* 48:778–787
- Lee S, Kim YS, Oh HJ (2012) Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping. *J Environ Manage* 96:91–105
- Livingstone DJ, Manalack DT, Tetko IV (1997) Data modelling with neural networks: advantages and limitations. *J Comput Aid Mol Des* 11:135–142
- Madrucci V, Taioli F, de Araujo CC (2008) Groundwater favorability map using GIS multicriteria data analysis on crystalline terrain, São Paulo State, Brazil. *J Hydrol* 357:153–173
- Murthy KSR (2000) Ground water potential in a semi-arid region of Andhra Pradesh: a geographical information system approach. *Int J Remote Sens* 21:1867–1884
- Murthy KSR, Mamo AG (2009) Multi-criteria decision evaluation in groundwater zones identification in Moyale–Teltele subbasin, South Ethiopia. *Int J Remote Sens* 30:2729–2740
- National Academy of Agricultural Science (1977) *Detailed soil map, 1:25,000*. NAAS, Seoul
- National Geographic Information Institute (1997) *Digital topographic map, 1:5,000*. NGII, Seoul
- O'Leary DW, Friedman JD, Pohn HA (1976) Lineament, linear, lineation; some proposed new standards for old terms. *Geol Soc Am Bull* 87:1463–1469
- 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* 399:158–172
- Paola JD, Schowengerdt RA (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 YJ, Lee KK, Kim JM (2001) Effects of highly permeable geological discontinuities upon groundwater productivity and well yield. *Math Geol* 32:605–618
- Pradhan B (2009) Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. *Cent Eur J Geosci* 1:120–129
- Ranganai RT, Ebinger CJ (2008) Aeromagnetic and Landsat TM structural interpretation for identifying regional groundwater exploration targets, south-central Zimbabwe Craton. *J Appl Geophys* 65:73–83
- Sander P, Chesley MM, Minor TB (1996) Groundwater assessment using remote sensing and GIS in a rural groundwater project in Ghana: lessons learned. *Hydrogeol J* 4:40–49
- Saud MA (2010) Mapping potential areas for groundwater storage in Wadi Aurnah Basin, western Arabian Peninsula, using remote sensing and geographic information system techniques. *Hydrogeol J* 18:1481–1495
- Singh CK, Shashtri S, Singh A, Mukherjee S (2011) Quantitative modeling of groundwater in Satluj River basin of Rupnagar district of Punjab using remote sensing and geographic information system. *Environ Earth Sci* 62:871–881
- Solomon S, Quiel F (2006) Groundwater study using remote sensing and geographic information system (GIS) in the central highlands of Eritrea. *Hydrogeol J* 14:729–741

- Srivastava P, Bhattacharya A (2006) Groundwater assessment through an integrated approach using remote sensing, GIS and resistivity techniques: a case study from a hard rock terrain. *Int J Remote Sens* 27:4599–4620
- Todd DK (1980) *Groundwater hydrology*. Wiley, New York, pp 111–163
- Todd DK, Mays LW (2005) *Groundwater hydrology*. Wiley, New York
- Vijay R, Panchbhai N, Gupta A (2007) Spatio-temporal analysis of groundwater recharge and mound dynamics in an unconfined aquifer: a GIS-based approach. *Hydrol Process* 21:2760–2764
- Yeh HF, Lee CH, Hsu KC, Chang PH (2009) GIS for the assessment of the groundwater recharge potential zone. *Environ Geol* 58:185–195
- Zhou W (1999) Verification of the nonparametric characteristics of backpropagation neural networks for image classification. *IEEE T Geosci Remote* 37:771–779