Characterizing effects of landscape and morphometric factors on water quality of reservoirs using a self-organizing map

Young-Seuk Park a, b, *, Yong-Su Kwon a, Soon-Jin Hwang c, Sangkyu Park d

a Department of Biology, Kyung Hee University, Seoul 130-701, Republic of Korea
b Department of Life and Nanopharmaceutical Sciences, Kyung Hee University, Seoul 130-701, Republic of Korea
c Department of Environmental Health Science, Konkuk University, Seoul 143-701, Republic of Korea
d Department of Biological Science, Ajou University, Suwon 443-709, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 6 November 2013
Received in revised form 27 January 2014
Accepted 28 January 2014
Available online

Keywords:
Reservoir classification
Reservoir water quality
Reservoir morphometry
Land cover types
Self-organizing map (SOM)
Theoretical path model
Agricultural reservoir

A B S T R A C T

Understanding the pattern of reservoir water quality in relation to morphometry and other landscape characteristics can provide insight into water quality management. We investigated the water quality of 302 reservoirs distributed nationwide in Korea by classifying them using a self-organizing map (SOM), examining how hydrogeomorphometry variables are related to reservoir water quality, and evaluating the effects of variables at different categories including geology, land cover, hydromorphology, and physicochemistry on reservoir water quality through a theoretical path model. The SOM classified the reservoirs into six clusters, from least to most polluted, with differences in physicochemical and hydrogeomorphometry variables between clusters. Water quality exhibits strong relationships with the proportions of urban, agricultural, and forest land cover types in the watersheds. Finally, our results revealed that hydrogeomorphometry of reservoirs and percentages of land cover types within watersheds have a considerable impact on the water quality of adjacent aquatic ecosystems.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The water quality of agricultural reservoirs has a considerable influence on the quality of agricultural products (e.g. rice) that utilize water from reservoirs. Reservoir water chemistry is determined primarily by nutrient loading through hydrological processes, which in turn are affected by various factors including morphometry and land use/land cover (Rasmussen et al., 1989; D’Arcy and Carignan, 1997; Fergus et al., 2011). Moreover, landscape and land use/land cover types within watersheds are closely related to human activity (Lee et al., 2009; Catherine et al., 2010; Cross and Jacobson, 2013), and hydrogeomorphological characteristics of reservoirs influence reservoir transport processes, particularly through regulation of nutrient concentrations stemming from nutrient loading (Håkanson, 2005; Zampella et al., 2007; Bremigan et al., 2008; Fergus et al., 2011). For example, Jones et al. (2004) revealed the importance of landscape and morphological variables in determining nutrient concentrations in Missouri reservoirs, and Bremigan et al. (2008) demonstrated the importance of integrating land use/land cover, hydrological and morphological features, and food web interactions in the investigation of critical interactions and feedbacks among the physical, chemical, and biological components of linked land–water ecosystems.

Aquatic ecosystems in Asian countries including Korea are strongly influenced by severe rainfall and drought. The monsoon season in Korea usually occurs from late June to July accounting for more than 40% of the annual precipitation (Jung et al., 2001). Flood and drought periods in this region are strongly influenced by the periods and intensity of monsoon. Due to unbalanced precipitation throughout the year, this region inevitably needs reservoirs for flood control, irrigation, and water supply. Therefore, large numbers of reservoirs have been developed in this region, reflecting the historical importance of agriculture in relation to the Asian monsoon climate (Hwang et al., 2003). Although there are more than 18,700 reservoirs in Korea, there is limited limnological information in understanding the relationships between hydrogeomorphological characteristics and water quality of reservoirs.

Understanding the current states of target aquatic ecosystems can assist in the development of suitable management strategies. Classification of reservoirs is useful for setting reference conditions,
detecting trends in water chemistry, and simplifying reservoir management by grouping together reservoirs for which similar management strategies are likely to have similar results (Martin et al., 2011). Many previous studies have proposed the classification of reservoirs based on differences in various environmental factors such as trophic status and water chemistry (Carlson, 1977; Kratzer and Brezonik, 1981; OECD, 1982; Søndergaard et al., 2005), and hydrologic and geomorphic features (Winter, 2001; Wolock et al., 2004; Martin et al., 2011). For such approaches, one or two variables are typically used for classification. Therefore, these approaches are relatively simple to apply, require minimal data, and are generally easy to understand. However, the results obtained do not reflect the complex properties of reservoirs or the factors controlling these complexities. To overcome the limitations of classification using simple variables, multivariate approaches have been adopted to classify reservoirs based on chemical, morphometric, and geographical factors using techniques such as principal component analysis (Søndergaard et al., 2005; Navarro et al., 2009), fuzzy logic (Lu et al., 1999; Icaga, 2007), classification and regression trees (Martin et al., 2011), hierarchical linear models (Cheruvelli et al., 2008), and self-organizing maps (Lu and Lo, 2002; Simeonova et al., 2010). However, a synthetic understanding of the structure and function of ecosystems in a landscape context requires incorporation of a broader suite of individual response variables, hydrological and morphological features, and ecosystem processes at multiple spatial scales (Bremigan et al., 2008; Martin et al., 2011).

The ecological data that are typically used to assess the ecological structure and function of reservoirs affected by many interrelated factors are bulky, non-linear, and complex and exhibit noise, redundancy, internal relationships, and outliers. Moreover, a wide range of variables and complex interactions exist between explanatory and response variables. Accordingly, non-linear analysis methods are preferable when dealing with such non-linear and complex ecological data. One particular non-linear approach, the use of artificial neural networks, has become increasingly popular in recent years. In particular, the self-organizing map (SOM; Kohonen, 2001) has arisen as a versatile tool for extracting information from complex data and may be applied effectively in the classification and association of samples and their variables. In ecological studies, since Chon et al. (1996) first applied the SOM technique to study the patterning of benthic communities in streams, the method has become popular in ecological studies and is used frequently for the classification of various ecological and environmental data: communities (Foody, 1999; Giraudeau and Lek, 2001; Park et al., 2003; Kwon et al., 2012; Bae et al., 2013), hydrosystems and landscapes (Tison et al., 2004), natural resource and ecosystem management (Park et al., 2003, 2013; Gevrey et al., 2004; Park and Chung, 2006), water resources (Lu and Lo, 2002; Zhang et al., 2008; Bae et al., 2013; Liukkonen et al., 2013), computational policy simulations for natural hazard mitigations (Samarasinghe and Strickert, 2013), surface temperature anomaly and solar activity (Friedel, 2012), and spatial and temporal variations of benzene (Strebel et al., 2013), Kalteh et al. (2008) and Chon (2011) reviewed the applications of the SOM techniques in ecological and environmental sciences.

Most previous studies have focused primarily on variables in single categories (such as water chemistry, hydrology, geology, landscape, reservoir morphology), or have conducted limited analyses to investigate the relationships between variables in different categories in terms of reservoir water quality. However, understanding reservoir features in multiple categories can improve predictive relationships because many variables are involved in mediating the responses of reservoirs to anthropogenic changes (Knoll et al., 2003; Jones et al., 2004; Bremigan et al., 2008).

Therefore, in the present study, we evaluated water quality and environmental variables at 302 reservoirs throughout Korea using multiple categories of related factors. The hierarchical framework was used to determine which environmental variables can best explain the observed water quality. Specifically, we aimed to define patterns of water quality of reservoirs throughout Korea, to identify the most dominant environmental variables influencing reservoir water quality, and to clarify the relative influence of regional (geographical and land cover) and local (morphological) variables constraining water quality in Korean reservoirs in the monsoon region.

2. Materials and methods
2.1. Data collection

We constructed a dataset composed of variables at multiple spatial scales and included geographical, land cover, hydromorphological, and water quality variables. Water quality data were obtained from the Water Quality Monitoring Networks of Reservoirs, operated by the Ministry of Environment in Korea (http://water.mene.go.kr/). From the database of the monitoring networks, 302 agricultural reservoirs were selected based on the availability of data. The dataset comprised six variables: chlorophyll-a (Chl-a), total suspended solids (TSS), dissolved oxygen (DO), chemical oxygen demand (COD), total nitrogen (TN), and total phosphorus (TP). Variables had been measured according to the protocol provided by the Ministry of Environment, Korea (Ministry of Environment, 2011). pH and DO were measured in situ using the appropriate sensors (i.e. a 632-pH meter and TOA DO-149, respectively). Other variables were analysed in the laboratory. Water samples had been preserved and analysed within the recommended time limit to avoid deterioration. BOD measures the amount of oxygen consumed by microorganisms over a period of 5 days during incubation in dark conditions at 20 °C. COD was determined from organic matter oxidized in a sulphuric acid medium by K2Cr2O7. TSS was determined by the weight difference imparted by filtration through a glass fibre filter, after drying at 105–110 °C for 2 h. Chl-a, TN, and TP were measured by absorptionmetry analysis.

Additionally, we used eight geographical and morphometric factors (altitude, bank height (B_height), bank width (B_width), circumference (distance around reservoirs), reservoir length, surface area, pondage, and catchment area) and percentages of various land cover types in corresponding reservoir basins and investigated their relationships with water quality. Altitude, reservoir length, and circumference were measured from a geographic information system (GIS) map, whereas bank height, bank width, surface area, pondage, and catchment area were obtained from the Korea Agriculture and Rural Infrastructure Corporation (KARICO, 2000). We divided circumference by surface area to calculate the complexity of shorelines. The percentages of each land cover type in each catchment area were obtained from the Water Management Information System (WAMIS, http://www. wamis.go.kr) operated by the Ministry of Land, Transport, and Maritime Affairs of Korea, and we used five land cover categories: urban, grassland, forest, paddy field, and dry field.

2.2. Modelling procedure
2.2.1. Trophic states

We calculated Spearman rank correlation coefficients to evaluate the relationships between the selected environmental variables. The trophic states of reservoirs were categorized into five groups based on the standardization of Kratzer and Brezonik (1981) using Carlson’s trophic state index (TSI) (Carlson, 1977). TSI with Chl-a (TSIChl-a), TP (TSITP), and TN (TSITN) were calculated as Eqs. (1)–(3) (Carlson and Simpson, 1996):

\[
\text{TSI}_{\text{Chl-a}} = 9.81 \times \ln(\text{concentration of Chl-a}) + 30.6
\]

\[
\text{TSITP} = 14.42 \times \ln(\text{concentration of TP}) + 4.15
\]

\[
\text{TSITN} = 14.43 \times \ln(\text{concentration of TN}) + 54.45
\]

where concentrations of each chemical must be in units of mg/L.

2.2.2. Self-organizing map

A self-organizing map (SOM) was used to characterize the spatial distribution patterns of water quality in the 302 study reservoirs. The SOM approximates the probability density function of the input data and can be used for efficient clustering and visualization of high-dimensional data in a two-dimensional lattice (Kohonen, 2001). Then, the SOM averages the input dataset in weight vectors through a learning process to remove noise.

A typical SOM consists of two layers: an input layer and an output layer. All of the units (or neurons) in the input layer are connected to all of the units in the output layer and each connection is associated with a weight, known as the connection intensity. First, the input layer receives input data; then, the dissimilarity (i.e. the
Euclidean distance between the weight vector and the input vector) is calculated. Here, we used six chemical variables as input data in the present study: Chl-\(a\), TSS, DO, COD, TN, and TP. The output layer typically consists of \(N\) output neurons arranged in a two-dimensional grid to improve visualization, and each output neuron represents a computational unit in the learning process. The optimum arrangement for the output layer is thought to be a hexagonal lattice, because this configuration does not favor the horizontal and vertical directions as much as a rectangular array (Kohonen, 2001; Park et al., 2003). There are no strict rules regarding the number of output neurons selected. In the present study, we trained the SOM with different map sizes and selected 84 (\(N = 12 \times 7\)) as the appropriate number of output neurons based on a heuristic equation (\(S/N\) of number of samples) (Vesanto and Alhoniemi, 2000).

During the typical learning process for a SOM, each input vector is assigned to one output neuron through SOM calculation. Each output neuron has a vector of coefficients or weights associated with the input data; these weights establish a link between the input units and their associated output units. The learning algorithm can be described as follows: when an input vector (in this case, values of six chemical factors in a sample) is presented to the SOM, the neurons in the output layer compete to be selected as the best matching unit (BMU), which is represented by the minimum distance between the weight vector and input vector. The BMU and its neighbors, which are predefined in the algorithm, update their weight vectors according to the sequential learning rules of SOM. The learning process is continued until a stopping criterion is met; this typically occurs when the weight vectors stabilize or when a particular number of iterations is completed. The training is usually done in two phases: a rough training to achieve ordering with a large neighborhood radius of BMU, followed by a fine tuning with a small radius. Gaussian function was used to determine the neighborhood radius size of BMU, resulting in a more smooth formation of topography in the map and faster convergence of the weight vectors. This learning process results in training the network to classify the input vectors by the weight vectors. Detailed descriptions of the computation and ecological application of the SOM algorithm can be found in Kohonen (2001) and Park et al. (2003), respectively.

Measuring quantitative performance of models is vital to establish an appropriate level of their confidence (Bennett et al., 2013). In the present study, two criteria, the quantization error for resolution and the topographic error for topology preservation, were used to evaluate the quality of the maps. The quantization error is the average distance between each data vector and its BMU for measuring map resolution (Kohonen, 2001), and the topographic error represents the proportion of all data vectors for which first and second BMUs are not adjacent for the measurement of topology preservation (Kivivuoto, 1996). Thus, these error values were used as indicators of the accuracy of mapping in preserving the topology (Kohonen, 2001; Park et al., 2003, 2006). After the learning process of SOM, a hierarchical cluster analysis was conducted according to the Ward linkage method using Euclidean distance to define the cluster boundaries in the SOM units. The multi-response permutation procedure (MRPP) was applied to evaluate the significance of the clusters defined in the SOM. Additionally, analysis of variance (ANOVA) and Tukey’s honestly significant difference (HSD) test were conducted to compare differences in variables between the clusters defined in the SOM. The elements of weight vectors for different input variables were visualized to characterize their contribution with respect to the clusters on the trained SOM in grey scale.

To train the SOM, we used the functions and parameters provided in the SOM toolbox (http://www.cis.hut.fi/projects/somtoolbox) for MATLAB version 6.1, which was developed at the Laboratory of Information and Computer Science at Helsinki University of Technology. MRPP was conducted using the software program PC-ORD. Other statistical analyses were conducted using the software program STATISTICA.

### 2.2.3. Theoretical path model

A theoretical path model (TPM) was used to describe the directed dependencies between a set of variables (Donovan, 1984) based on Pearson correlation coefficients and coefficient of determination (\(R^2\)), for environmental variables from multiple stepwise regression models (Li et al., 2012). Only a single important indicator was employed for each category in the TPM. The environmental variables were categorized into four groups, from large to small scale: geographical factors, land cover types, morphometric factors, and physicochemical factors. The concentration of Chl-\(a\) was used as a dependent factor, and TPMs were conducted using SPSS.

### 3. Results

#### 3.1. Characteristics of environmental variables

Altitude exhibits the strongest correlation with COD (\(P < 0.01\)), followed by Chl-\(a\), TP, and TSS (\(P < 0.01\)) (Table 1). Percentage of forest area correlates negatively with nutrient variables, whereas paddy field area exhibits a positive correlation with all chemical variables except DO.

The majority of study reservoirs are eutrophic (Table 2). Among 302 reservoirs, 71.7% were nutrient-rich (65.0% and 6.7% in eutrophic and hypertrophic states, respectively) based on TSITN (Fig. 1a), whereas 61.4% were nutrient-rich (46.4% eutrophic and 15.0% hypertrophic) (Fig. 1b) based on TSITP. Meanwhile, a total of 59.8% of reservoirs (49.3% eutrophic and 10.5% hypertrophic) were nutrient-rich based on TSIChl-\(a\) (Fig. 1c), showing a pattern similar to that based on TSITP. Trophic states of oligotrophic and ultraoligotrophic conditions were less than 20% in all three TSIs, with the lowest proportion in TSITN (6.7%).

#### 3.2. Classification of reservoirs

After the training the SOM, the final quantization and topographic errors are 0.142 and 0.013, respectively, indicating that the SOM was smoothly trained in topology. Subsequently, the SOM units were classified into six clusters (I–VI) based on dissimilarities in the dendrograms of hierarchical cluster analysis; these clusters are significantly different in terms of water quality (MRPP, \(A = 0.40, P < 0.01\)) (Fig. 2), although clusters I and II, clusters III and IV, and clusters V and VI are relatively similar.

We have visualized the contribution of each environmental variable in grey scale, with dark and light representing the highest and lowest values, respectively (Fig. 2c). These values were calculated during the SOM learning process; this visualization describes the discriminatory powers of input variables on the clusters defined in the SOM. Clusters V and VI (particularly VI) are characterized by

---

**Table 1**

Spearman correlation coefficients between geographical, morphometric, and physicochemical variables, and percentages of various land cover types.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Physicochemical variables</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographical variable</td>
<td>Dissolved oxygen (DO)</td>
<td>-0.15*</td>
<td>-0.45**</td>
<td>-0.22**</td>
<td>-0.05</td>
<td>-0.37**</td>
</tr>
<tr>
<td>Morphometric variables</td>
<td>Chemical oxygen demand (COD)</td>
<td>0.08</td>
<td>0.55**</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.08*</td>
</tr>
<tr>
<td></td>
<td>Total suspended solid (TSS)</td>
<td>0.09</td>
<td>0.25**</td>
<td>0.10</td>
<td>-0.01</td>
<td>0.22**</td>
</tr>
<tr>
<td></td>
<td>Total nitrogen (TN)</td>
<td>0.03</td>
<td>0.24**</td>
<td>0.14</td>
<td>0.07</td>
<td>0.23**</td>
</tr>
<tr>
<td></td>
<td>Total phosphorus (TP)</td>
<td>0.04</td>
<td>0.19**</td>
<td>0.07</td>
<td>0.05</td>
<td>0.19**</td>
</tr>
<tr>
<td></td>
<td>Chlorophyll-(a) (Chl-(a))</td>
<td>0.06</td>
<td>0.32**</td>
<td>0.16**</td>
<td>0.08</td>
<td>0.29**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.04</td>
<td>0.19**</td>
<td>0.07</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>0.10</td>
<td>0.03</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Percentages of land cover types</td>
<td></td>
<td>0.05</td>
<td>0.37**</td>
<td>0.26**</td>
<td>0.24**</td>
<td>0.42**</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>0.00</td>
<td>-0.58**</td>
<td>-0.27**</td>
<td>-0.21**</td>
<td>-0.46**</td>
</tr>
<tr>
<td></td>
<td>Forest</td>
<td>0.03</td>
<td>0.53**</td>
<td>0.25**</td>
<td>0.18**</td>
<td>0.42**</td>
</tr>
<tr>
<td></td>
<td>Paddy field</td>
<td>0.06</td>
<td>0.46**</td>
<td>0.16**</td>
<td>0.09</td>
<td>0.32**</td>
</tr>
<tr>
<td></td>
<td>Grassland</td>
<td>0.13</td>
<td>0.26**</td>
<td>0.26**</td>
<td>0.30**</td>
<td>0.44**</td>
</tr>
</tbody>
</table>

*\(P < 0.05\), **\(P < 0.01\).
the highest concentration of nutrients, whereas clusters I and II (particularly I) exhibit the lowest values for most variables; clusters III and IV exhibit intermediate values. Moreover, TSS exhibits a maximum value in cluster V. These patterns were also observed in field-measured data, which exhibit significant differences in each variable between clusters (Tukey’s HSD test, \( P < 0.05 \)) (Table 2).

TSITP and TSITN are significantly higher in cluster VI than in the other clusters and are lowest in cluster I (Tukey’s HSD test, \( P < 0.05 \)) (Fig. 3); TSITN exhibits similar patterns. Meanwhile, the TN:TP ratio is highest in cluster I and lowest in cluster VI (Tukey’s HSD test, \( P < 0.05 \)).

Altitude is significantly higher in clusters I and III and noticeably lower in cluster VI (Tukey’s HSD test, \( P < 0.05 \)) (Fig. 4). Similarly, bank height is higher in clusters I, II, and III than in clusters IV, V, and VI. Conversely, bank width is lowest in cluster I and highest in cluster VI. The circumference, length, and surface area of the reservoirs are also lowest in cluster I, whereas shoreline complexity is highest in cluster I and lowest in cluster VI. Finally, catchment area is relatively high in cluster VI, although there are no significant differences between clusters.

Percentages of land cover types for each reservoir basin vary significantly between clusters (Tukey’s HSD test, \( P < 0.05 \)) (Fig. 5). Cluster I, which is characterized by the lowest values of all variables, also exhibits a considerably high proportion of forest area and low proportions of urban, paddy field, dry field, and grassland areas; conversely, cluster VI exhibits the highest percentages of urban, grassland, and paddy field areas.

3.3. Theoretical path model

The TPM showed that forest area (a land cover type variable) and bank height (a hydro-morphological variable) are the factors with the greatest influence on Chl-\( \alpha \) in each category, exhibiting significant negative correlations with Chl-\( \alpha \) (\( -0.51 \) and \(-0.50 \), respectively; both \( P < 0.01 \)) and positive correlations with altitude (\( 0.54 \) and 0.66, respectively; both \( P < 0.01 \)) (Fig. 6). Moreover, COD exhibits negative correlations with forest area and bank height (\( -0.58 \) and \(-0.55 \), respectively; both \( P < 0.01 \)). Finally, Chl-\( \alpha \) concentration (as a water quality factor) exhibits a negative relationship with altitude, forest area, and bank height but a positive relationship with COD concentration, resulting in a high explanation power (\( R^2 = 0.52 \)) (Fig. 6).

4. Discussion

The present study has revealed a close relationship between water quality and landscape/morphometric variables of reservoirs. In particular, our results demonstrate that SOM and TPM techniques are legitimate methods for the classification of reservoirs using various water quality factors and can be used to characterize the relationships between numerous variables.

The geographical location of reservoirs is of critical importance in determining their water quality, particularly because water quality can be determined by factors including climate and hydrology, pollutant sources, land cover/use in the associated catchment area, and reservoir morphometry (Ahearn et al., 2005; Håkanson, 2005; White and Greer, 2006; Lee et al., 2009). In the present study, altitude was found to exhibit a positive relationship with bank height but negative relationships with bank width, circumference, reservoir length, surface area, and various chemical variables. That is, reservoirs located at relatively high altitudes with high bank height and low bank width exhibit lower nutrient concentrations than those at low altitudes, possibly owing to lower nutrient loadings from the associated catchment areas. Conversely, reservoirs at low altitudes are usually surrounded by urban and agricultural fields, which play primary roles in degrading water quality in adjacent aquatic systems, e.g. by altering soil surface conditions, increasing the impervious area, and generating pollution (White and Greer, 2006).

The results of our study highlight the importance of morphometric characteristics of reservoirs in determining water quality. Bank width, circumference, reservoir length, and surface area are...
smaller in cluster I (oligotrophic to mesotrophic) than in cluster VI (hypertrophic state), whereas bank height and shoreline complexity are highest in cluster I and lowest in cluster VI (Fig. 4). The morphometric characteristics of lakes/reservoirs are known to be helpful in anticipating change within a system and predicting how such change will affect the resident organisms (Håkanson, 2005). Moreover, various pollutants that originate primarily from human land use activity outside aquatic systems are transported to reservoirs through the land–water interface. Recently, the geometric shape (edge) of reservoirs was demonstrated empirically to be significant in determining their water quality (Cadenasso and Pickett, 2000; Hwang et al., 2007). As such, reservoir morphometry, including shoreline complexity, can affect material transport into reservoirs and regulate nutrient concentrations imparted by loading, primary production, and secondary production.

Recently, modelling methods such as fuzzy logic and the SOM technique have been used in the classification and assessment of water quality in reservoirs, offering improvements upon traditional methods such as Carlson’s TSI and the OECD classification (Lu and Lo, 2002; Søndergaard et al., 2005; Icaga, 2007; Navarro et al., 2009; Simeonova et al., 2010; Nikoo and Mahjour, 2013). In addition, the SOM approach has become a focus of particular interest across various ecological studies and has been applied in data mining for long-term ecological data (Hyun et al., 2005; Kangur et al., 2007), hazard rating for natural resources (Park and Chung, 2006; Park et al., 2013), prediction of water resources (Lee and Scholz, 2006; Ahn et al., 2011), and dimension reduction of large datasets (Park et al., 2006; Griebeler and Seitz, 2006). Recently Voyslavov et al. (2012, 2013) reported a tool combining the SOM and Hasse diagram techniques for use in surface water quality assessment. The SOM method, which includes the most popular unsupervised learning algorithm available, reduces a high-dimensional space to fewer dimensions (Kohonen, 2001). Moreover, the SOM technique is effective in the clustering, visualization, and abstraction of complex data using a non-linear projection of multivariate data into lower dimensions. This method is more

Fig. 2. (a) Classification of 302 reservoirs through the SOM learning process with six environmental variables. b) A dendrogram of the hierarchical cluster analysis according to the Ward linkage method using Euclidean distance, defining six clusters. c) Distribution pattern of six environmental variables used in SOM training. Dark and light units represent high and low levels, respectively, of each environmental variable.

Fig. 3. Differences in trophic state and TN:TP ratio between six different clusters defined in the SOM. a) TSI with TP (TSI_{TP}), b) TSI with TN (TSI_{TN}), c) TSI with Chl-a (TSI_{Chl-a}), and d) TN:TP ratio. Alphabetic letters indicate significant differences between clusters based on Tukey’s HSD multiple comparison test (P < 0.05).
efficient than algebraic methods in dealing with large amounts of noisy data and has been recommended for use in exploratory approaches to datasets in which unexpected structures might be found. In present study, the SOM method was used to classify agricultural reservoirs using environmental variables throughout Korea, and to characterize the relationships between water quality variables and hydrogeomorphical variables.

The distribution pattern of environmental variables in each cluster in the SOM is indicated by the gradient distribution map (Fig. 2c). Cluster VI displays the highest concentration of nutrients;
in particular, this cluster exhibits significantly higher concentrations of Chl-a and TP than any other clusters (Table 2). DO, COD, and TN also exhibit similar patterns, indicating a highly polluted state. Conversely, altitude is considerably lower in cluster VI than in other clusters (Fig. 4). Reservoirs in cluster VI exhibit relatively high bank width and low bank height in the lowlands. Generally, increasing catchment area is thought to increase pollutant loading, thus increasing the possibility of reservoir eutrophication (Kalff, 2001; Hwang et al., 2003), although pollutant loading can also be affected by the percentage of various land cover types within a given catchment (Hwang et al., 2007). Moreover, the rate of application of agricultural fertilizer in Korea is very high ($2.1 \times 10^6$ tons y$^{-1}$), particularly for TN ($4.6 \times 10^5$ tons N y$^{-1}$). Thus, the background concentration of nitrogen in Korean aquatic systems is typically higher than that in other countries (Kim et al., 2001). Differences in the optimum N:P ratio at different trophic states may be related to both morphometric factors and biological factors such as the presence and abundance of phytoplankton species (Kim et al., 2007). The nutrient limitation can also be affected by the monsoon climate, particularly in June and July in South Korea, and flooding events can increase phosphorus discharge and cause decreases in the TN:TP ratio in highly polluted reservoirs in summer.

Our study supports the limnological theory that nutrient concentrations are determined primarily by external inputs and modified by morphology and hydrology (Edmondson, 1961; Jones et al., 2004). Agricultural areas (including paddy fields and dry fields) are greater sources of nutrients than other dominant types of land use: in the present study, chemical variables such as COD, TP, and Chl-a were found to display strong positive correlations with paddy field and dry field areas but negative correlations with forest area (Table 1). Urban and grassland areas were also found to be correlated positively with chemical variables. Previous studies have consistently revealed nutrient export from agricultural areas to be several times greater than that from grassland and forest areas, and nutrient export from urban areas often equals or exceeds that from agriculture (Frink, 1991; Jones et al., 2004). Nutrients from agricultural lands originate from non-point pollution sources, which are known to be the dominant sources of water quality impairment in most countries. The Ministry of Environment of Korea estimated the proportion of water pollutant loading from non-point sources to be 42–69% in 2003 and predicted that this proportion will rise to 65–70% by 2015; this indicates the importance of non-point sources for water pollution (Ministry of Environment, 2006). Therefore, our results suggest that efforts to improve reservoir water quality should focus on minimizing non-point nutrient sources.

5. Conclusions

We characterized the effects of landscape factors and morphometric factors on reservoir water quality at the national scale in South Korea using SOMs and TPM methods. The studied reservoirs were classified into six clusters, revealing differences in trophic states and physicochemical, landscape, and morphometric variables between clusters. The results indicate that reservoir water quality is strongly dependent on the altitude and morphometry of reservoirs, which can be represented by factors such as bank height, bank width, circumference, surface area, and complexity of shoreline. Water quality also exhibits a strong relationship with proportions of urban, agricultural, and forest land cover types within reservoir watersheds. In particular, the theoretical path model revealed the influence of altitude, percentage of land cover types, and morphometric variables on reservoir water quality. Thus, our results support the assertion that the morphometry of reservoirs and land cover types within watersheds have significant impacts on the water quality of adjacent aquatic ecosystems.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF- 2013R1A1A2009494).

References


Fig. 6. (a) Schematic diagram of relationships between water quality variables, and (b) theoretical path models showing the relationships between variables at various scales. The bold and normal letters represent $R^2$ values obtained from the multiple stepwise regression models and Pearson correlation coefficients, respectively. The solid lines indicate positive relationships between variables, while the dotted lines indicate negative ones.