https://doi.org/10.20965/jaciii.2016.p0992

Paper:

Analysis of Water Quality of Lake Hachiroko in Japan Using a Fuzzy Multiple Regression Model with ALOS AVNIR-2 Data

Dejian Wang*, **Yoichi Kageyama**,†**, **Makoto Nishida****, **and Hikaru Shirai****

*Dalian Nationalities University
18 Liaohexi Road, Kaifa Zone, Dalian 116600, China
**Graduate School of Engineering and Resource Science, Akita University
1-1 Tegata Gakuen-machi, Akita-shi, Akita 010-8502, Japan
E-mail: kageyama@ie.akita-u.ac.jp
†Corresponding author

[Received April 27, 2016; accepted August 30, 2016]

The distribution of water pollution is often assessed by remote sensing. In this study, we develop a fuzzy multiple regression model and analyze water quality using data collected by the Advanced Visible and Near Infrared Radiometer type-2 (AVNIR-2) of the Advanced Land Observing Satellite at different time points. We conduct a fuzzy multiple regression analysis of the AVNIR-2 data and direct measurements of the local water quality of Lake Hachiroko in Japan. The relationship between the AVNIR-2 and water quality data are analyzed by solving both min and max problems. We compare the estimated water quality maps with the actual distributions in the study area, and determine that the proposed method enables us to derive water quality conditions effectively from the AVNIR-2 data. Furthermore, by comparing maps created using AVNIR-2 data collected at different times, we obtain results revealing temporal changes in water quality. In addition, we compare maps created using the fuzzy multiple regression and fuzzy regression models. We demonstrate that the former offers a greater number of solutions and provides more details about water quality.

Keywords: remote sensing, ALOS, AVNIR-2, fuzzy multiple regression analysis, water quality

1. Introduction

The water quality of rivers and lakes is often checked to monitor the level of water pollution. A typical investigation involves extracting water samples directly from several locations. Although this type of method is well suited to the collection of water quality data for a relatively small area, difficulties exist in applying it to the monitoring of water quality over a large area.

Therefore, remote sensing has been used to analyze water quality. This method is especially useful given its ability to obtain measurements instantaneously, its wide-area coverage, and its periodicity [1–9]. For example, studies of the water quality of Lake Garda in Italy were previously undertaken using Hyperion [6], and Lake Kasumigaura in Japan, was studied using Landsat thematic mapper (TM) data [7]. In addition, temporal changes in water quality, such as at Lake Chilka in India, have been studied using three sets of IRS-IA satellite data [8], and in the northwestern part of the Baltic Sea using MERIS data [9]. An algorithm that employs a neural network has been used to monitor water quality based on remote sensing data [10, 11]. The aforementioned studies prove the usefulness of a neural network in studying water quality.

However, analysis of water quality using remote sensing data by means of conventional methods involves certain problems. First, extensive observation data are required. For example, in one study [12], observation data were taken from 138 locations at Taihu Lake in China. When a neural network is used, many training samples are required to build models [11]. In one study [11], 55 samples of observation data were collected, and in another [10], 136 samples were required. In addition, obtaining good water quality analysis results is difficult because of the effect of specific disturbances and uncertainties (e.g., atmospheric, surface, and water-wave effects, as well as noise in the measurement system) on remote sensing data [5]. Moreover, an increasing number of studies have included water quality measurements of chlorophyll-*a* (Chl *a*) and total suspended solid (TSS) [9, 12–14] than other measurements. Although these studies have confirmed that Chl *a* and TSS are useful in analyzing water quality, developing methods of analysis based on water quality parameters other than Chl *a* and TSS, and creating water quality estimate maps are necessary. To overcome these problems, new water quality analysis methods using remote sensing data must be developed.

In our study, in order to estimate the water quality of Lake Hachiroko in Japan based on surface information used as point information (such as that related to low water quality), and to estimate quality while considering any uncertainties in the data, we applied a fuzzy regression model to analyze the water quality of Lake Hachiroko based on Landsat TM data [5]. Although conventional methods used to determine water quality distribution typ-

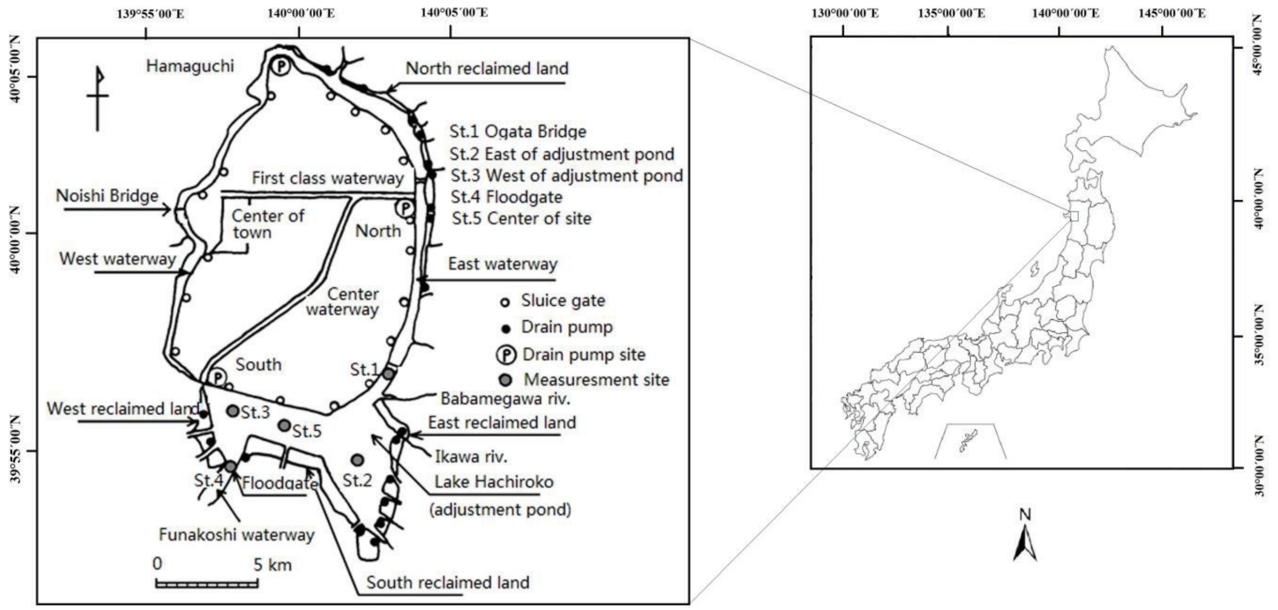


Fig. 1. Overview of Lake Hachiroko and the water quality measurement sites (St. 1 to 5).

ically employ water quality measurements obtained from many observation points (> 20), we considered the applying fuzzy regression analysis to such water quality monitoring data obtained from only five locations. The results showed that a water quality distribution map can be created using only a smattering of water quality data.

Analyses using remote sensing data present the challenge of constructing maps to estimate water quality when data are obtained under less-than-optimal conditions such as in the presence of cloud cover, and given the regression period of remote sensing data [2, 15]. Methods have been proposed to estimate water quality using data obtained from the active sensor on the Japanese Earth Resources Satellite-1 Synthetic Aperture Radar (SAR) [15] and the Advanced Land Observing Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR) [16]. All of these can obtain data without being constrained by weather conditions such as cloud cover. The results have clearly shown that the textures computed from these data can be applied to water quality analysis. To extrapolate water quality conditions over periods in which TM data cannot be obtained, numerical simulations were performed [2].

Data collected by the Advanced Visible and Near Infrared Radiometer type-2 (AVNIR-2) of the ALOS can also be used for water quality analysis [17]. Because the ground resolution of AVNIR-2 data is higher than that of TM data (30 m), AVNIR-2 data can be used for detailed analysis of water quality that was previously not possible because of the insufficient resolution of TM data. This method yields more solutions and provides detailed water quality results than possible with TM data [17]. In other words, the results of analysis show that the fuzzy regression model is useful for estimating water quality.

However, because fuzzy regression analysis considers only one of the features of remote sensing data, analyzing specific bands and water quality data is difficult. To solve this problem, we applied a method that uses a fuzzy multiple regression model. Because the fuzzy multiple regression model considers two types of data, namely, band and water quality data, we believe that obtaining more detailed water quality conditions is possible than with the simpler fuzzy regression analysis.

In this study, we applied a fuzzy multiple regression model developed using both water quality data (obtained from five locations) and AVNIR-2 data, to analyze the water quality in Lake Hachiroko. In addition, we compared the solutions with the actual local water quality conditions.

2. Study Area and Materials

2.1. Study Site Description

Lake Hachiroko, also known as Hachirogata, is located approximately 20 km north of Akita and is a brackish-water lake with a center latitude of 40°N and a longitude of 140°E. Until 1956, it extended 12 km east–west and 27 km north–south, and had a total area of 22,024 ha. The average depth was approximately 4 m, and even the deepest parts of the lake did not exceed 4.5 m. Based on a reclamation plan, parts of the lake were drained in 1956, and by May 1966, these parts had dried sufficiently to allow agriculture and settlement. In all, 17,239 ha, that is, 78.3% of the initial area of Hachirogata, was reclaimed.

Currently, Hachirogata has a surface area of only 4565 ha and consists of an east waterway, west waterway, and adjustment pond. Agricultural water for Ogata passes through 19 gates, and after it irrigates farmlands, the water is discharged into the lake through the main waterway and two drain pump sites at the north and south ends. Water from the lake is released into the Sea of Japan intermittently through floodgates located at the end of the adjustment pond. **Fig. 1** provides an overview of

Lake Hachiroko and the water quality measurement sites: Ogata Bridge, east of the adjustment pond, west of the adjustment pond, the floodgate, and the center of the site. These sites are designated St. 1 to 5, respectively.

2.2. Pollution and Water Quality Situations

More than 20 rivers, such as the Mitanegawa and Babamegawa, flow into Lake Hachiroko. Recent years have witnessed many water quality problems, such as blooms of green algae [18].

Given its increased levels of water pollution, Lake Hachiroko was listed as a "designated lake" under the "Act on Special Measures Concerning Conservation of Lake Water Quality" [19] in 2007. Therefore, understanding the details of the lake's surface water quality and the effect of seasons on its quality is necessary.

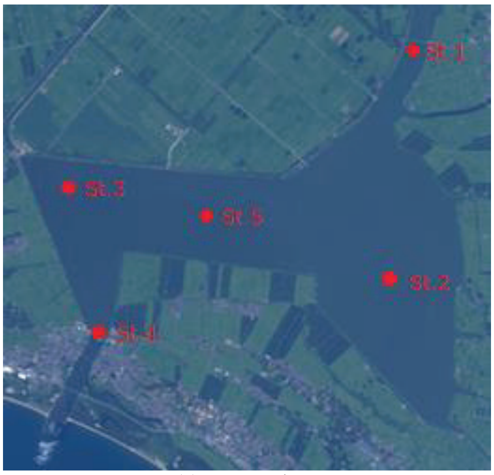
In addition, a water quality expert has expressed the following salient points regarding the study area [5]:

1) Murky water containing agricultural and domestic wastewater flows from the southern and northern drain pump sites.
2) An increase in pollution has occurred in the eastern part of Lake Hachiroko and near the floodgate, where contaminants accumulate because of poor water circulation.
3) Murky water containing domestic wastewater flows from the Babamegawa River, as well as from the drain pump site to the east of Lake Hachiroko.
4) Pollution in the west waterway, including the Noishi Bridge, is worsening because of poor water circulation.
5) Little movement of the lake water has occurred, except for a certain amount caused by wind.

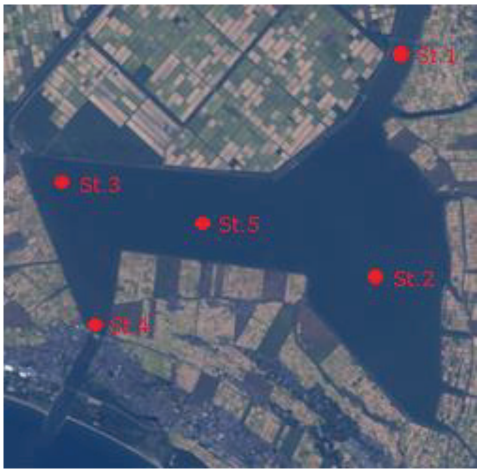
Points 2) and 3) are particularly relevant to the target area of this study.

2.3. AVNIR-2 Data for Analysis

AVNIR-2 data are obtained from Bands 1 to 4, which lie in the visible and near-infrared regions. The wavelengths of these bands are 0.42 to 0.50 μm (visible blue), 0.52 to 0.60 μm (visible green), 0.61 to 0.69 μm (visible red), and 0.76 to 0.89 μm (near-infrared), respectively. ALOS was finished in 2011, and because the regression period of ALOS was 46 days [20] but water quality measurements were recorded only once a month, in this study, data collected on 26 August 2006 (hereafter, "August data"), and September 20 (hereafter, "September data"), were used in the analysis, as shown in **Fig. 2**.



(a) August data.



(b) September data.

Fig. 2. AVNIR-2 data used (RGB; Bands 3, 2, and 1) and water quality measurement sites (St. 1 to 5).

2.4. Water Quality Measurements

Life environment parameters are constantly reviewed by the Akita Prefectural Government. From these, we selected six water quality parameters that can be used in our analysis to indicate levels of water pollution [18]: hydrogen-ion indicator (pH), dissolved oxygen (DO), chemical oxygen demand (COD), suspended solids (SS), total nitrogen (T-N), and total phosphorus (T-P). The results suggested that these water quality parameters reflect the overall water quality conditions. These water quality parameters were used both individually and in combination with the AVNIR-2 band data and applied to our fuzzy multiple regression analysis.

We selected water quality measurements recorded on 23 August 2006 and 26 September 2006, as listed in **Table 1**. Measurements were recorded at the water surface, and no rainfall occurred in the study area on the dates of the observations. The wind speed on both dates was recorded as approximately 1 m/s by the Japan Meteorological Agency [18].

3. Methods

The aim of this study is to clarify the relationship between AVNIR-2 data and the directly measured water

Table 1. Water quality measurements recorded by the Akita Prefectural Government [18].

Measurement site	Water quality parameters					
	pH	DO (mg/ℓ)	COD (mg/ℓ)	SS (mg/ℓ)	T-N (mg/ℓ)	T-P (mg/ℓ)
Ogata Bridge (St.1)	9.0 8.7	9.1 11.0	19 12	29 23	0.79 0.40	0.22 0.13
East of adjustment pond (St.2)	9.1 8.8	11.0 11.0	14 10	28 16	0.49 0.41	0.24 0.12
West of adjustment pond (St.3)	9.6 8.7	13.0 9.9	15 11	26 18	0.54 0.49	0.18 0.20
Floodgate (St.4)	9.2 9.1	10.0 12.0	12 15	4 42	0.40 0.54	0.18 0.28
Centre of the site (St.5)	9.4 8.8	13.0 10.0	16 11	24 12	0.59 0.43	0.19 0.15

Upper figure: measurements recorded on 23 August 2006.
Lower figure: measurements recorded on 26 September 2006.

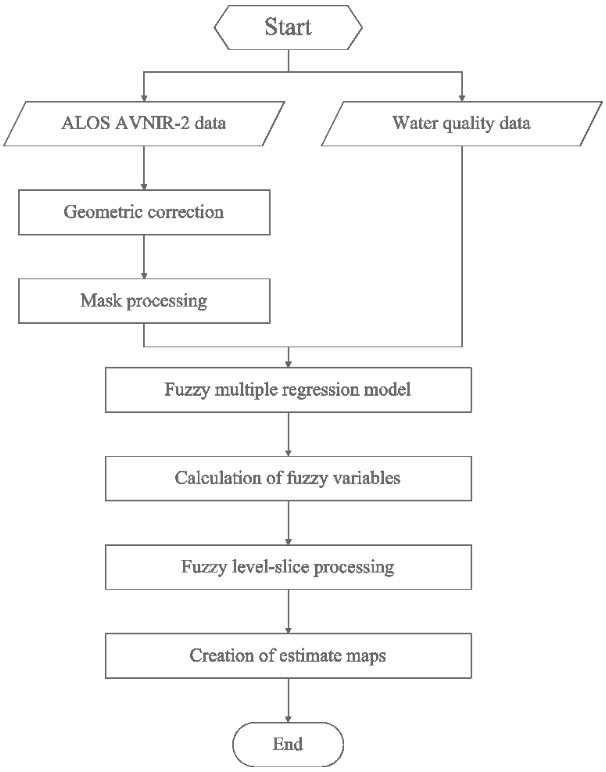


Fig. 3. Flow of water quality analysis.

quality data, as well as to analyze water quality conditions using a fuzzy multiple regression model. **Fig. 3** shows the process of water quality analysis using AVNIR-2 and water quality data.

3.1. Preprocessing

In general, remote sensing data acquired by satellites are distorted by factors such as slight variations in the satellite's attitude, the rotation of the Earth, and the curvature of the Earth's surface. Stretching of the remote sensing data to match the criteria image is referred to as "geometric correction," in that it applies corrections to the distortions previously described. In our study, 20 ground control points were selected and geometrically corrected using a secondorder conformal transformation. Resampling was performed by cubic convolution interpolation [21]. The target water area was then extracted using mask processing. **Fig. 4** shows the results from mask processing the AVNIR-2 data.

3.2. Fuzzy Multiple Regression Model

Because of the ground resolution of the sensors used to acquire remote sensing data, compensating for various disturbances and uncertainties in the remote sensing data [5] was necessary. The fuzzy set theory provides useful concepts and tools for addressing uncertainties. The fuzzy multiple regression model, which was developed based on fuzzy set theory, assumes that any difference between observation data and a model prediction indicates system fuzziness, thus revealing the relationship between input and output [22]. Remote sensing data include external disturbance components and sensing system noise. Therefore, we must consider fuzziness when processing data. We specified monitoring sites for AVNIR-2 data and assumed that the digital number (DN) of the pixels around them is a fuzzy number.

The fuzzy multiple regression model is based on the fuzzy regression model and computed using the following equations [5]:

$$Y(X_p) = A_0 + A_1 X_{p1} + \cdots + A_n X_{pn} \quad \dots\dots \quad (1)$$

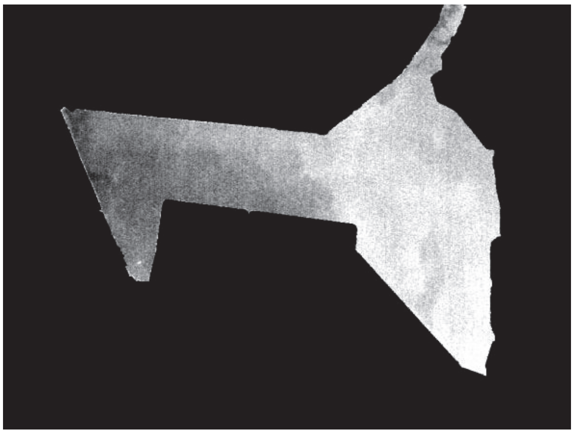
$$= (a(X_p), e(X_p))_L \quad \dots\dots \quad (2)$$

However, the fuzzy multiple regression model (with its two inputs and one output) as applied in this study can be given as follows:

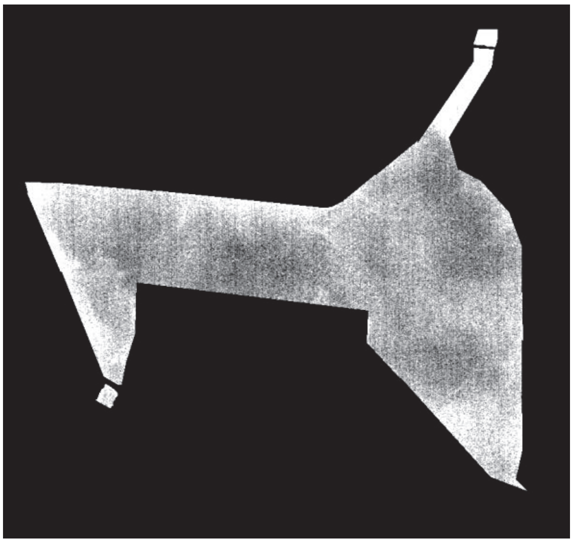
$$Y(X_p) = A_0 + A_1 X_{p1} + A_2 X_{p2} \quad \dots\dots \quad (3)$$

$$= (a_0, e_0)_L + (a_1, e_1)_L X_{p1} + (a_2, e_2)_L X_{p2} \quad (4)$$

where $Y(X_p)$ is an estimate (fuzzy number) that denotes the estimated fuzzy output interval; X_p is the output vector; and X_{p1} and X_{p2} are input vectors. Regression variable $A_i (i = 0, 1, 2)$ is a triangular fuzzy number. To calculate interval $Y(X_p)$ from Eqs. (3) and (4), $A_i (i = 0, 1, 2)$ must be known. In fact, the problem of fuzzy multiple regression analysis involves obtaining the fuzzy coefficient

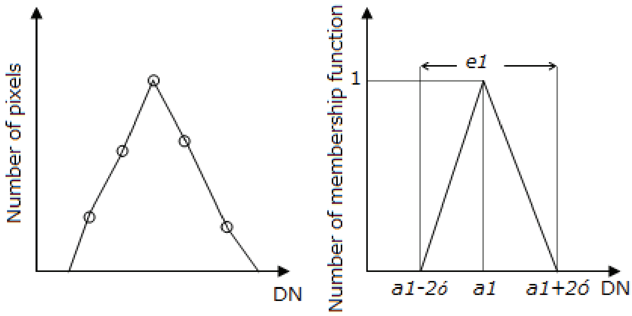


(a) August data, band 1.



(b) September data, band 1.

Fig. 4. Results of mask processing.



(a) Measured value. (b) Triangular fuzzy number.

Fig. 5. Measured value and triangular fuzzy number of DN.

$A_i(i=0,1,2)$ in a fuzzy multiple regression model [21]. Because $A_i(i=0,1,2)$ is a triangular fuzzy number, $a_i(i=0,1,2)$ is assumed to be the center of each $A_i(i=,1,2)$, and $e_i(i=0,1,2)$ is assumed to be the width of each $A_i(i=0,1,2)$. In other words, the main objective of fuzzy multiple regression analysis is to obtain six variables (a_0, e_0, a_1, e_1, a_2, e_2).

In this study, the variables (a_0, e_0, a_1, e_1, a_2, e_2) were calculated using the output vector X_p and the input vectors X_{p1} and X_{p2} by means of the following equations [22]:

$$a(X_p) = a_0 + a_1 \cdot a(X_{p1}) + a_2 X_{p2} \quad \dots\dots (5)$$

$$e(X_p) = e_0 + e_1 \cdot |e(X_{p1})| + e_2 |X_{p2}| \quad \dots\dots (6)$$

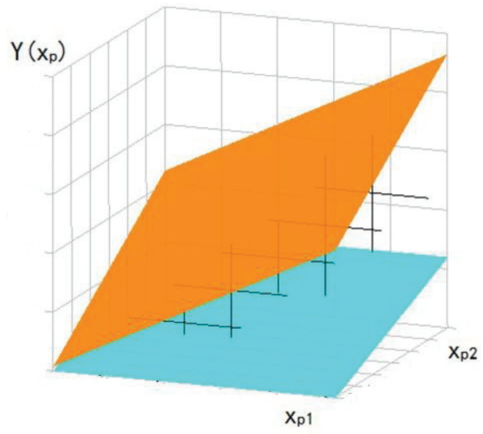
where X_p is derived from any band of the AVNIR-2 data; X_{p1} is taken from the AVNIR-2 data for Bands 1 to 4, with the exception of the X_p band; and X_{p2} is the set of measured water quality data. Because X_p and X_{p1} are assumed to be triangular fuzzy numbers, $a(X_p)$ is the center of X_p, and $e(X_p)$ is the width of X_p, which is calculated as the mean DN and twice the standard deviation ($\acute{o}$) of 25 pixels around each water quality measurement site. Furthermore, $a(X_{p1})$ and $e(X_{p1})$ can be obtained in a similar manner. **Fig. 5** shows an example of the manner in which the triangular fuzzy number is determined from the measured value. In addition, our preliminary experiments confirmed that the classification results obtained using the triangular membership were better than those obtained using the normal distribution membership functions.

In fuzzy multiple regression models, linear programming problems known as "min" and "max" problems can be formulated using interval data [22]. Here, the formulation of the min problem requires the use of a linear regression model with a minimum width that includes all interval data. Conversely, the formulation of the max problem requires the use of a linear regression model with a maximum width that is included within all interval data. In this study, we assigned the output vector X_p and the input vectors X_{p1} and X_{p2} to Eqs. (5) and (6). Thus, we calculated six variables (a_0, e_0, a_1, e_1, a_2, e_2) from the min and max problems. **Fig. 6** shows an example of a fuzzy multiple regression model and reflects the relationships between input and output vectors in the min and max problems.

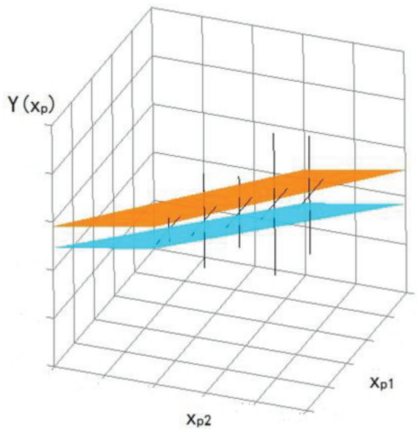
One processing run of the fuzzy multiple regression model uses a combination of the DN for band data and the values of one type of water quality measurement for five sites. For example, for our analysis using a combination of Band 2 data as X_p, Band 4 data as X_{p1}, and TN as X_{p2}, the DNs for Band 2 and 4 data are listed in **Table 2**, and the TN measurements are listed in **Table 1**. We substituted these values into Eqs. (5) and (6). In other words, we set the mean DN and twice the standard deviation of Band 2 as $a(X_p)$ and $e(X_p)$; the mean DN and twice the standard deviation of Band 4 as $a(X_{p1})$ and $e(X_{p1})$; and the TN measurements as X_{p2}. Thereafter, we calculated the coefficients by solving the min problem. The values a_0, e_0, a_1, e_1, a_2, and e_2 were found to be 81.566, 6.917, 0.211, 0.000, 0.000, and 0.000, respectively. Furthermore, $a_i(i=0,1,2)$, and $e_i(i=0,1,2)$ are not simple numbers but rather the coordinates of the center and the width of the triangular fuzzy number in the fuzzy multiple regression model, respectively.

3.3. Fuzzy Level-Slice Processing

The fuzzy output interval obtained using the fuzzy multiple regression model shows that the DN corresponds to



(a) Min problem.



(b) Max problem.

Fig. 6. Example of a fuzzy multiple regression model.

the forecast range obtained by measuring the water quality parameters and the AVNIR-2 band data. It has been demonstrated that water quality estimate maps generated using fuzzy levelslice processing can yield intermediate levels of water quality that are comparable to those obtained by conventional levelslice processing [5]. To investigate the water quality in greater detail, estimate maps were created by fuzzy levelslice processing in this study. This technique employs simplified fuzzy reasoning [23].

We assumed that the DN could correspond to specific water quality conditions set in an optional range. The production rule for estimating water quality for a given pixel is as follows:

$$\begin{array}{l} Rule\ \ 1: Y_1 \rightarrow Z_1 \\ \text{------------------------} \\ Rule\ \ n: Y_n \rightarrow Z_n \\ Input: S \\ \hline Output: Z_0 \end{array} \quad \ldots\ldots\ldots\ldots\ldots (7)$$

where $Y_i(i=1,\ldots,n)$ represents an estimated fuzzy set of the DN in proportion to the slice level. $Z_i(i=1,\ldots,n)$ represents the regression variables in each rule, which are calculated from the attributes of both the band data and water quality data in the proposed model. In addition, the values of the slice levels in the band data are calculated from the DN of each band data, and the values of the slice levels in the water quality data are set according to the environmental standard values for lakes recorded by the Akita Prefectural Government. S is the input for the DN. Z is the output, and is given as follows:

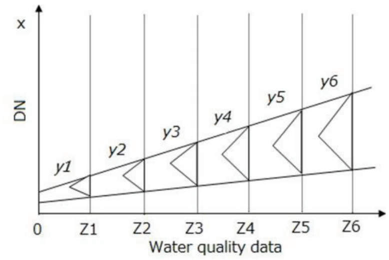
$$Z_0 = \frac{\sum_{i=1}^{n} h_i Z_i}{\sum_{i=1}^{n} h_i} \quad \ldots\ldots\ldots\ldots\ldots (8)$$

When the input S is known, h_i is the ratio for obtaining Z_i. The rule number corresponds to the slice number. In this study, we used six slices. **Fig. 7** shows an example of a fuzzy regression model and fuzzy set. For example, we assume six rules of DN that correspond to each water

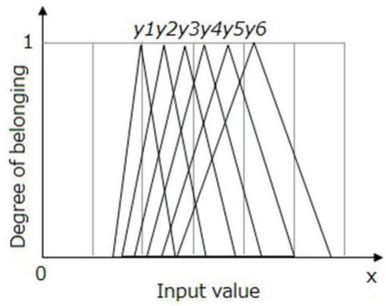
Table 2. DNs of 25 pixels around each water quality measurement site from the August data.

Measurement site	DN of band data			
	1	2	3	4
Ogata Bridge	125.640 (2.545)	90.640 (2.106)	52.360 (2.256)	55.840 (4.326)
East of adjustment pond	126.880 (2.445)	89.120 (2.099)	54.880 (1.677)	22.920 (1.132)
West of adjustment pond	124.040 (2.424)	83.280 (2.105)	51.200 (1.597)	20.960 (1.117)
Floodgate	125.200 (2.443)	87.560 (2.038)	54.520 (1.741)	24.280 (1.178)
Centre of the site	125.560 (2.413)	86.720 (2.177)	52.600 (1.589)	23.120 (1.075)

Upper figure: mean DN.
Lower figure: standard deviation.



(a) Fuzzy regression model.



(b) Fuzzy set.

Fig. 7. Example of a fuzzy regression model and fuzzy set.

Table 3. Results of water quality analysis obtained using August data.

	Fuzzy output	1			2			3			4		
	Band data	2	3	4	1	3	4	1	2	4	1	2	3
Water quality	pH	×	×	×	×	×	×	×	×	×	×	×	×
		×	×	\	×	×	\	×	×	\	\	\	\
	DO	×	×	×	×	○	○	×	○	×	×	○	○
		×	○	\	○	○	\	×	×	\	\	\	\
	COD	×	×	×	○	×	○	×	○	×	×	×	○
		×	×	\	×	×	\	×	○	\	\	\	\
	SS	○	○	×	○	○	○	×	×	×	×	×	×
		×	×	\	×	○	\	×	○	\	\	\	\
	T-N	○	○	×	○	○	○	×	×	×	×	×	×
		×	×	\	×	×	\	×	×	\	\	\	\
	T-P	○	○	×	×	×	○	○	○	×	×	×	×
		×	×	\	×	○	\	×	○	\	\	\	\

○: A normal solution.
×: A solution with negative DN and that cannot be used for water quality analysis.
\: A solution cannot be obtained or used.
Upper row: results of the min problem.
Lower row: results of the max problem.

Table 4. Results of water quality analysis obtained using September data.

	Fuzzy output	1			2			3			4		
	Band data	2	3	4	1	3	4	1	2	4	1	2	3
Water quality	pH	×	×	×	×	○	×	×	×	×	×	×	×
		×	×	\	×	○	\	×	×	\	\	\	\
	DO	×	×	×	×	×	×	×	×	×	×	×	×
		×	×	\	×	×	\	×	×	\	\	\	\
	COD	○	○	×	○	○	○	○	○	×	×	×	×
		×	×	\	○	○	\	○	○	\	\	\	\
	SS	○	○	×	○	○	×	○	×	×	○	×	○
		○	×	\	○	○	\	○	○	\	\	\	\
	T-N	○	×	×	○	×	×	○	○	×	×	○	×
		×	×	\	×	×	\	×	×	\	\	\	\
	T-P	○	×	×	×	×	×	×	×	×	×	×	×
		×	×	\	×	×	\	×	×	\	\	\	\

○: A normal solution.
×: A solution with negative DN and that cannot be used for water quality analysis.
\: A solution cannot be obtained or used.
Upper row: results of min problem.
Lower row: results of max problem.

quality measurement. In Eq. (7), Y_i is DN, and Z_i corresponds to the water quality measurements. The relation of the rules is shown in **Fig. 7(a)**. When the input S is known, we can calculate the ratio h_i for each rule n, as shown in **Fig. 7(b)**. We assign the values of h_i and Z_i in Eq. (8), and calculate the output Z_0 for the input S.

4. Results

4.1. Selection of Solution

The DN value of the ALOS AVNIR-2 data should be between and 255. In our study, when the slope of the fuzzy multiple regression model was determined to be negative, or when the interval was negative, the solution was not used for analysis. In particular, if a water quality estimate map was obtained when the coefficient was negative, a solution could not be obtained or used; this situation is indicated by "\" in the result tables. If a water quality estimate map was obtained but contained pixels with negative DNs, the solution could not be used for analysis; this situation is indicated by "×." For all events other than those marked "\" and "×," water quality estimate maps could be obtained and used for water quality analysis; these situations are indicated by "○."

Fig. 8. Water quality estimated with min problem using the August Band 1 and 3 data, and T-N.

Tables 3 and **4** list the results obtained in this study based on data from two bands. In these tables, the first row, "Fuzzy output," lists the band of the fuzzy output interval estimate, corresponding to $Y(X_p)$ in Eq. (5). The second row, "Band data," lists the band of the regression variable, which corresponds to X_{p1} in Eq. (5). Thus, the second column, "Water quality," is the regression variable, corresponding to X_{p2} in Eq. (5). For example, in **Table 3**, the results obtained from a combination of Band 1, Band 2, and pH, in the min problem, is marked "×."

In addition, to explain the relationships between the estimated maps and regression variables, we present the slice levels of the regression variables in the figures. The estimate map in **Fig. 8** reflects the slice levels of the DNs in Band 1 data. The slice levels in Band 3 data and water quality parameter TN are shown on the right side of this figure.

In this study, to evaluate the proposed method, we first calculated the solutions and then verified the effectiveness of the proposed method by comparing the calculated solutions with the local water quality data and actual water quality situations. Finally, we confirmed the consistency of the calculated results with the local water quality data.

4.2. Results of Water Auality Analysis for Each Level of Water Quality

Tables 3 and **4** list the results of the water quality analysis for each water quality measurement and the data for each band for the min and max problems. We found that when we used August data, the DO, COD, SS, TN, and TP values could be used to draw estimated maps with 32 patterns (indicated in **Table 3** by "○"), and when we used the September data, the pH, COD, SS, TN, and TP could be used to draw estimated maps with 31 patterns (indicated in **Table 4** by “○”).

5. Discussion

5.1. Discussion of Water Quality Estimate Maps

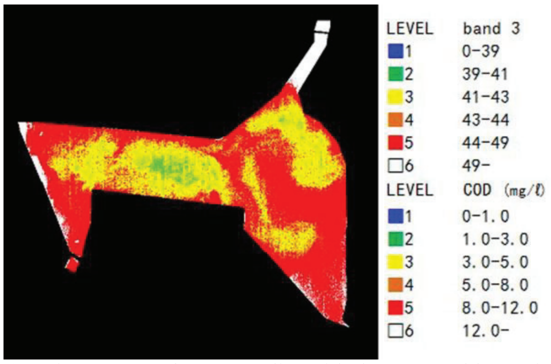
Figure 8 shows an example of a water quality estimate map obtained using August Band 1 data as an estimate, as well as Band 3 data and total nitrogen (TN) measurements as regression variables for the min problem. Near the Ogata Bridge, the T-N level of the water quality data, measured locally, was 0.79 mg/L, and in the estimate map, several pixels are classified as LEVEL 5 (red: T-N = 0.6 to 1.0 mg/L). This map is in good agreement with the measured water quality data. Near the floodgate, many pixels are classified as LEVEL 3 (yellow: T-N = 0.2 to 0.4 mg/L) and LEVEL 4 (orange: T-N = 0.4 to 0.6 mg/L), indicating that the water pollution was worse in this location than at the west end of the adjustment pond. This result is in agreement with Points 2) and 3) of the actual water quality situations, as detailed in Section 2.2.

Figure 9(a) shows an example of a water quality estimate map obtained using September Band 2 data as an estimate, as well as Band 3 data and chemical oxygen demand (COD) measurement as regression variables for the max problem. Near the Ogata Bridge, the COD from the water quality data, measured locally, was 12.0 mg/L, and on the estimate map, several pixels are classified as LEVEL 6 (white: COD > 12.0 mg/L). This map is in good agreement with the measured water quality data. Near the floodgate, many pixels are classified as LEVEL 5 (red: COD = 8.0 to 12.0 mg/L) and LEVEL 6 (white: COD > 12.0 mg/L), indicating that the water pollution there was worse than elsewhere in the lake. This result is in agreement with Points 2) and 3) of the actual water quality situations (Section 2.2).

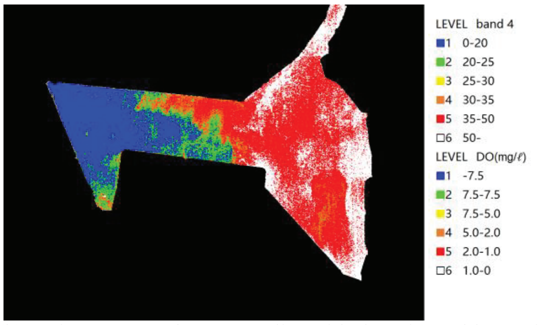
Thus, not only are the results shown in **Figs. 8** and **9** in good agreement with the actual water quality situations described in Points 2) and 3), but the other valid results (indicated in **Tables 3** and **4** by “○”) are as well. This indicates the efficacy of the proposed method. Some of the results are shown in **Fig. 9**. However, because the proportion of valid solutions was not very high (22.2% of the August data, and 21.5% of the September data), improving the efficiency of analysis by considering factors such as additional features of remote sensing data is necessary in a future study.

5.2. Comparison with Classification Results Obtained Using the Fuzzy Regression Model

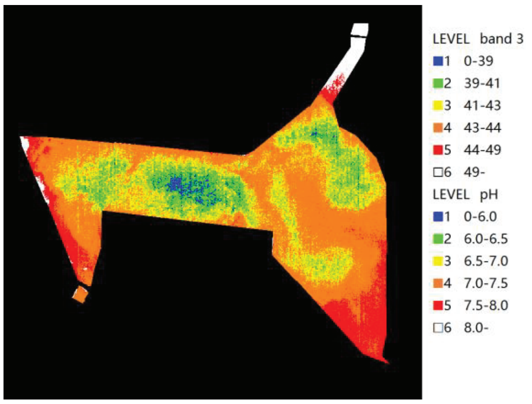
In this study, we used data acquired from ALOS AVNIR-2 and estimated the water quality of Lake Hachiroko using fuzzy multiple regression analysis. Wang et al. applied a fuzzy regression model to water quality analysis using the same AVNIR-2 data [17]. To determine the effectiveness of the proposed method, we compared the results obtained by the proposed method with those obtained using methods detailed in previous studies.



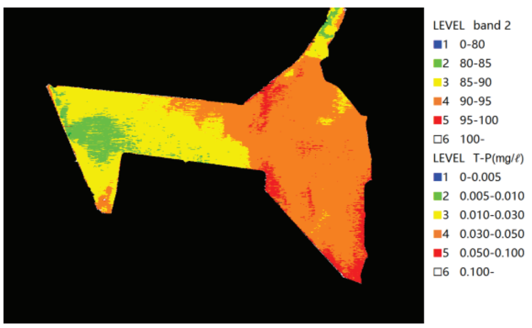
(a) Estimate map of water quality with the max problem using September band 2 data, band 3 data, and COD.



(b) Estimate map of water quality with the min problem using August band 2 and 4 data and DO.

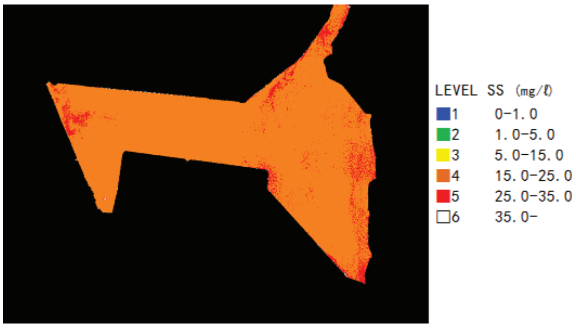


(c) Estimate map of water quality with the min problem using September band 2 and 3 data and pH.

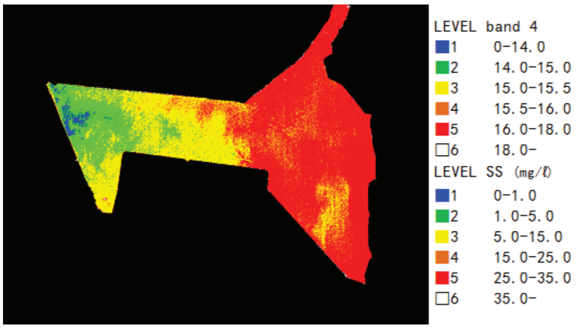


(d) Estimate map of water quality with the max problem using August band 3 and 2 data and T-P.

Fig. 9. Some results obtained by the proposed method.

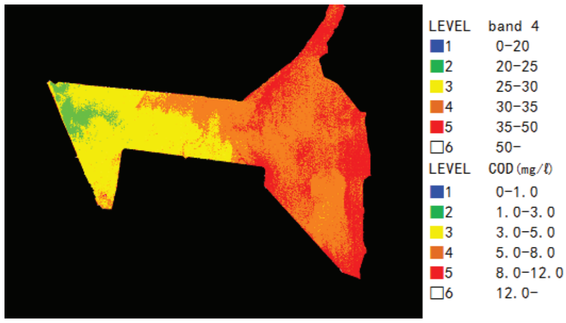


(a) Fuzzy regression model using AVNIR-2 band 2 data and SS.

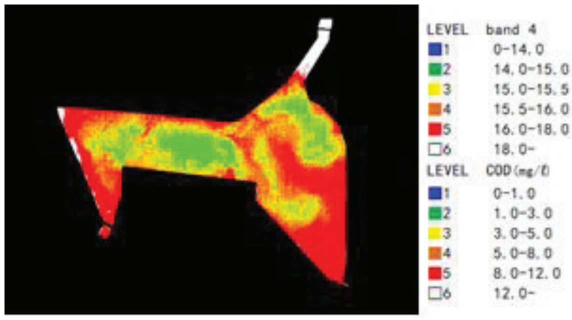


(b) Fuzzy multiple regression model using AVNIR-2 band 2 data, band 4 data, and SS.

Fig. 10. Maps of water quality estimated using August data.



(a) August data.



(b) September data

Fig. 11. Maps of water quality estimated using Band 2 and 4 data, and COD.

Regarding the August data, we compared 13 patterns obtained from the solutions generated using the fuzzy regression model and 32 from those generated using the fuzzy multiple regression model. Regarding the September data, we compared eight patterns obtained from the solutions generated using the fuzzy regression model and 31 from those generated using the fuzzy multiple regression model. Furthermore, as shown in **Fig. 10**, for the same AVNIR-2 data and water quality measurements, the fuzzy regression model yielded two classification levels, whereas the fuzzy multiple regression model yielded five classification levels.

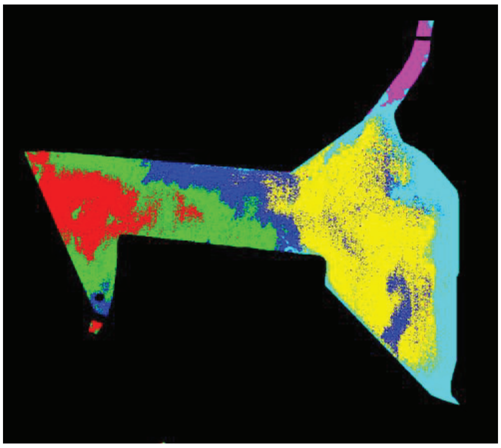
Although obtaining results for all the band data and water quality data is impossible, a comparison of our results with those obtained using the fuzzy regression model shows that our method yields more solutions and provides more details about water quality. In addition, the study of [17] used 24 combinations for analysis when employing the fuzzy regression model, whereas our study used 144 when employing the fuzzy multiple regression model. Therefore, because of the increased number of combinations, generating an increased number of useful solutions is possible and can thus further develop our analysis. Considering that we plan to develop this method to obtain water quality information for those periods for which data are not available and to improve the temporal resolution, having a greater number of solutions and detailed water quality conditions available will be useful.

5.3. Temporal Changes in Water Quality

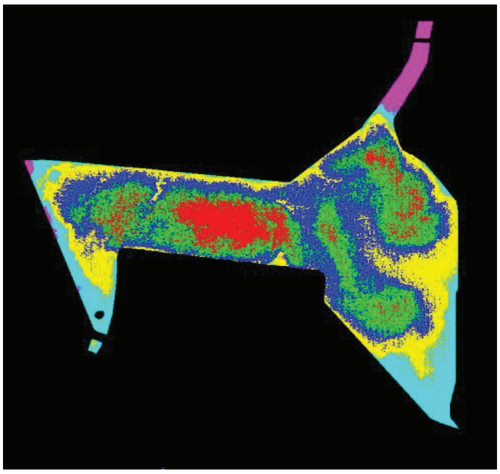
The results obtained using the ALOS AVNIR-2 data were found to be in good agreement with the measured water quality conditions for each period. However, the difference between the results obtained using the August and September data indicate temporal changes in water quality.

Temporal changes in the water quality of Lake Hachiroko were investigated by comparing the results obtained in this study with those obtained by using the fuzzy regression model [17]. We selected results that reflect the water quality conditions from both the August and September data sets within the same band and the same water quality measurements. The comparison showed that our results are similar to those of [17] and presented some new points. First, in the study of [17], the wavelength region of band 4 (0.76 to 0.89 μm) had a low reflectance at the water surface [24]. Therefore, only the AVNIR-2 data from Bands 1 to 3 were used for the analysis. In our study, we considered more features than those in the case involving fuzzy regression analysis and added an analysis of the water quality using AVNIR-2 Band 4 data, as shown in **Fig. 11**. We compared the temporal changes determined using data from AVNIR-2 Bands 1 to 4. Band 4 data provided more information on the temporal changes in water quality, suggesting that a wider scope of application was necessary to determine temporal changes in water quality.

Second, application of the fuzzy multiple regression model led to an increase in the number of solutions that can be used for comparison. Thus, the previously useless water quality parameter became useful. For example,



(a) August data.



(b) September data

Fig. 12. Classification results obtained by the k-means method.

in the study of [17], T-N could not reflect the temporal changes in water quality, but in this study, the T-N data were used. The results showed that we can compare the temporal changes in water quality by using a greater number of groups of ANVIR-2 band data and water quality measurements.

Furthermore, in the study of [17], only four sets of results were used to compare temporal changes in water quality. In this study, 13 sets (indicated in **Tables 3** and **4** by "○") were used. Because in both studies the same data (remote sensing data and water quality measurements) were used, a greater number of solutions means that more detailed information embedded in the remote sensing data can be estimated. Thus, the greater number of solutions reflects the high efficiency of the analysis. In addition, although measurements were few, expert knowledge [5] and the results of numerical simulations [2] reveal that gradual changes can reflect the actual situation.

We noticed some plume patterns in **Fig. 11(b)**. We plan to determine the reason for this phenomenon in a future study.

5.4. Comparison of our Results with those Obtained by the Conventional Method

In this study, August and September data were used to estimate the water quality of Lake Hachiroko by employing the fuzzy multiple regression model. To determine the effectiveness of the method, the k-means method was used for classification [21]. We also used the August and September data and set $k = 6$, a value determined to match the number of slice levels. **Fig. 12** shows the results classified by the k-means method. Considerable noise appeared in the results of the k-means method, and the relation between the classification results and water quality conditions was not immediately clear.

Conventional methods such as the k-means method, uses data from many bands simply and cannot yield a good classification result. By contrast, the proposed method considers the features of each band and uses the fuzzy multiple regression model. Thus, the proposed method is particularly useful for analyzing water quality conditions.

6. Conclusion

In this study, water quality conditions were analyzed by applying a fuzzy multiple regression model to the ALOS AVNIR-2 data. We reached the following conclusions:

1) The method proposed in this study, namely, the fuzzy multiple regression model, can accurately estimate the water quality in Lake Hachiroko, Japan.

2) Given that the estimated maps obtained using the fuzzy multiple regression model were more detailed, we suggest that detailed water quality conditions can be obtained by combining data from multiple bands.

3) The proposed method is useful for monitoring temporal changes in water quality.

The results obtained using the proposed method are in good agreement with the water quality measurements and actual water quality situations, thus confirming the effectiveness of our method.

In the future, we plan to increase the amount of data analyzed and examine the overall applicability of the proposed method.

Acknowledgements

The authors thank Drs. C. Ishizawa and T. Takahashi, Akita Univ., and Ms. C. Kasai for their help in conducting the experiments. The authors also thank the Akita Prefectural Government for providing water quality data.

References:

[1] S. Thiemann and H. Kaufmann, "Lake Water Quality Monitoring using Hyperspectral Airborne Data – A Semiempirical Multisensor and Multitemporal Approach for the Mecklenburg Lake District, Germany," Remote Sensing of Environment, Vol.81, pp. 228-237, 2002.

[2] Y. Kageyama and M. Nishida, "Water Quality Analysis based on Remote Sensing Data and Numerical Model," J. of Geography, Vol.109, No.1, pp. 27-36, 2000.

[3] K. Oki and Y. Yasuoka, "Estimation of Annual Total Nitrogen Load in Lake Basin with Remote Sensing – Case Study at Lake Kasumigaura," J. of the Remote Sensing Society of Japan, Vol.17, No.1, pp. 22-35, 1997.

[4] G. Campbell, S. R. Phinn, A. G. Dekker, and V. E. Brando, "Remote Sensing of Water Quality in an Australian Tropical Freshwater Impoundment using Matrix Inversion and MERIS Images," Remote Sensing of Environment, Vol.115, pp. 2402-2414, 2011.

[5] M. Nishida and K. Otsuka, "Application of Fuzzy Regression Model on Water quality Analysis with Satellite Image Data and Drawing of Estimation Map," Trans. IEE Japan, Vol.115-C, pp. 381-388, 1995.

[6] C. Giardino, V. E. Brando, A. G. Dekker, N. Strömbeck, and G. Candiani, "Assessment of Water Quality in Lake Garda (Italy) using Hyperion," Remote Sensing of Environment, Vol.109, pp. 183-195, 2007.

[7] O. Youichi, M. Bunkei, F. Takehiko, M. Kazuo, and I. Akio, "Application of Spectral Decomposition Algorithm for Mapping Water Quality in a Turbid Lake (Lake Kasumigaura, Japan) from Landsat TM Data," ISPRS J. of Photogrammetry and Remote Sensing, Vol.64, pp. 73-85, 2009.

[8] S. Sudhakar and D. K. Pal, "Water Quality Assessment of the Lake Chilka," Int. J. of Remote Sensing, Vol.14, No.14, pp. 2575-2579, 1993.

[9] E. T. Harvey, S. Kratzer, and P. Philipson, "Satellite-based Water Quality Monitoring for Improved Spatial and Temporal Retrieval of Chlorophyll-*a* in Coastal Waters," Remote Sensing of Environment, Vol.158, pp. 417-430, 2015.

[10] N. Chang and B. Vannah, "Comparative Data Fusion between Genetic Programing and Neural Network Models for Remote Sensing Images of Water Quality Monitoring," 2013 IEEE Int. Conf. on Systems, Man, and Cybernetics, pp. 1046-1051, 2013.

[11] Q. Shen, B. Zhang, J. Li, H. Zhang, and M. Chen, "Neural network modeling for retrieval of water quality of Lake Taihu from field spectral measurement," 2008 Congress on Image and Signal Processing, Vol.5, pp. 178-182, 2008.

[12] C. Le, Y. Li, Y. Zha, D. Sun, C. Huang, and H. Zhang, "Remote Estimation of Chlorophyll *a* in Optically Complex Waters based on Optical Classification," Remote Sensing of Environment, Vol.115, pp. 725-737, 2011.

[13] R. M. Cavalli, G. Laneve, L. Fusilli, S. Pignatti, and F. Santini, "Remote Sensing Water Observation for Supporting Lake Victoria Weed Management," J. of Environmental Management, Vol.90, pp. 2199-2211, 2009.

[14] S. Chen, L. Han, X. Chen, D. Li, L. Sun, and Y. Li, "Estimating Wide Range Total Suspended Solids Concentrations from MODIS 250-m Imageries: An Improved Method," ISPRS J. of Photogrammetry and Remote Sensing, Vol.99, pp. 58-69, 2015.

[15] D. Wang, Y. Kageyama, M. Nishida, and H. Shirai, "Algorithm to Analyze Water Quality Conditions of Lake Hachiroko using Textures of JERS-1 SAR Data," Int. J. of the Society of Materials Engineering for Resources, Vol.18, No.2, pp. 51-58, 2012.

[16] D. Wang, Y. Kageyama, M. Nishida, H. Shirai, and C. Kasai, "Water Quality Analysis of Lake Hachiroko, Japan, using ALOS PALSAR Data," Int. J. of the Society of Materials Engineering for Resources, Vol.20, No.2, pp. 175-180, 2014.

[17] D. Wang, Y. Kageyama, M. Nishida, H. Shirai, and A. Motozawa, "Water Quality Analysis in Lake Hachiroko, Japan, Using ALOS AVNIR-2 Data," IEEJ Transcripts on Electrical and Electronic Engineering, Vol.8, No.6, pp. 627-633, 2013.

[18] http://www.pref.akita.lg.jp/ [accessed April 21, 2016]

[19] http://www.env.go.jp/ [accessed April 21, 2016]

[20] http://www.eorc. jaxa.jp/ALOS/en/doc/fdata/ALOS_HB_RevC_EN.pdf [accessed April 21, 2016]

[21] M. Takagi and H. Shimoda, Handbook of Image Analysis [Revised Edition], University of Tokyo Press, 2004.

[22] H. Ishibuchi, "Fuzzy Regression Analysis," J. of the Japan Society for Fuzzy Theory and Systems, Vol.4, pp. 52-60, 1992.

[23] M. Mizumoto, "Fuzzy Reasoning (1)," J. of the Japan Society for Fuzzy Theory and Systems, Vol.4, pp. 256-264, 1992.

[24] Y. Kageyama, K. Miura, M. Nishida, and D. Ishiyama, "Analysis of Seasonal Change for Water Conditions in Lake Hosenko, Japan, Using ALOS AVNIR-2 Data," IEEJ Transcripts on Electrical and Electronic Engineering, Vol.7, pp. 225-227, 2012.

Name:
Dejian Wang

Affiliation:
Network and information center, Dalian Nationalities University

Address:
18 Liaohexi Road, Kaifa Zone, Dalian 116600, China

Brief Biographical History:
2003 Received M.E. degree, Akita University
2003- Joined ADK Fuji System Co., Ltd.
2010- Joined Dalian Nationalities University
2014 Received Dr. E. degree, Akita University

Main Works:
• Remote sensing

Name:
Yoichi Kageyama

Affiliation:
Graduate School of Engineering Science, Akita University

Address:
1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan

Brief Biographical History:
1997 Received M.E. degree, Akita University
1997- Research Associate, Department of Computer Science and Engineering, Akita University
2001 Received Dr. E. degree, Akita University
2001- Lecture, Akita University
2004- Associate Professor, Akita University
2013- Professor, Graduate School of Engineering and Resource Science, Akita University

Main Works:
• Remote sensing, human information, image processing, and image recognition

Membership in Academic Societies:
• The Institute of Electrical Engineers of Japan
• The Information Processing Society of Japan (IPSJ)
• The Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)

Name:
Makoto Nishida

Affiliation:
Graduate School of Engineering Science, Akita University

Address:
1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan

Brief Biographical History:
1974- Joined Toyota Motor Co. Ltd.
1975- Research Associate, Akita University
1987- Lecture, Akita University
1988- Visiting Researcher, Clarkson University
1992- Associate Professor, Akita University
1996- Professor, Department of Computer Science and Engineering, Akita University
2007- Director, General Information Processing Center, Akita University
2008- Dean, Faculty of Engineering and Resource Science, Akita University
2010- Dean, Graduate School of Engineering and Resource science, Akita University
2011- Executive Director/Vice President, Akita University
2014- Vice President, Akita University
2016- Professor, Akita University

Main Works:
- Remote sensing, image information applications, and knowledge-based information systems

Membership in Academic Societies:
- The Institute of Electrical Engineers of Japan
- The Information Processing Society of Japan (IPSJ)
- The Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
- The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Hikaru Shirai

Affiliation:
Graduate School of Engineering and Resource Science, Akita University

Address:
1-1 Tegata Gakuen-Machi, Akita 010-8502, Japan

Brief Biographical History:
2013 Received the M.E., Graduate School of Engineering and Resource Science, Akita University
2013- Joined Akita Electronics Systems Co., Ltd.
2014- Ph.D. course, Graduate School of Engineering and Resource Science, Akita University

Main Works:
- Remote sensing

Membership in Academic Societies:
- The Institute of Electrical Engineers of Japan
- The Information Processing Society of Japan (IPSJ)
- The Japan Society of Material Cycles and Waste Management