Analysis of Water Quality Conditions of Lake Hachiroko, Japan, with Fuzzy C-means Using Terra ASTER Data

Kai Matsui, Yoichi Kageyama*, and Hiroshi Yokoyama

Akita University, 1-1 Tegata Gakuen-machi, Akita City, Akita 010-8502, Japan

*Corresponding Author: kageyama@ie.akita-u.ac.jp

Abstract

Lake Hachiroko, Japan, has many water quality issues, evident from phenomena such as green algae blooms. Understanding the details of the surface water quality of the lake and the effect of seasons on its quality is necessary. In our previous studies, we conducted fuzzy regression analysis of remote sensing data and direct measurements of water quality. The results showed that estimation maps of water quality were well created, using only five data points of water quality parameters. To obtain maps that are in good agreement with experimental data, remote sensing data and water quality values should be acquired simultaneously. However, performing such simultaneous observations can affect the preparation of the water quality estimation maps. We overcame this obstacle by using fuzzy c-means clustering (FCM). With this method, we considered the effects of specific disturbances and uncertainties on the remote sensing data. The results indicated that FCM was particularly effective in determining suspended solids during water quality analysis. However, FCM has the issue that the initial values given to the classes used in FCM affect the FCM results. Thus, it is necessary to study the setting of the best initial values so as to obtain good estimation results. In this study, we evaluated the setting of the best initial values of FCM. In addition, based on the water quality data obtained from the study area, we considered to generate estimation maps that preset ranges of water quality levels. To evaluate the usefulness of FCM, we compared the estimation maps with the water quality data and the result of the fuzzy regression analysis.

Keywords: Remote Sensing, Water Pollution, Green Algae.

1. Introduction

Changes in water quality in rivers and lakes caused by water pollution are generally investigated by extracting water samples directly from several locations. However, this method is not appropriate for understanding the water quality conditions over a large area.

Therefore, remote sensing techniques have been used to analyze water quality, and these techniques are especially suitable for understanding the water quality conditions over a large area⁽¹⁻⁴⁾. In our previous studies, to estimate the water quality parameters in Lake Hachiroko, Japan, surface information, such as information on poor water quality, was obtained at different observation points. Because remote sensing data may have uncertainties, we applied a fuzzy regression model⁽³⁾ and a fuzzy multiple regression model for analysis⁽⁴⁾. The conventional methods for estimating water quality use water quality monitoring data sampled from a large number of observation points. However, we considered the possibility of applying fuzzy regression analysis to water quality monitoring data sampled from only a few observation points using the Advanced Visible and Near Infrared Radiometer type-2 on the Advanced Land Observing Satellite⁽³⁾. The results showed that water quality estimation maps can be created using only a small amount of water quality data. Because fuzzy regression analysis considers only one feature of remote sensing data, the analysis may be difficult depending on the specific bands and the water quality data. To solve this problem, we applied a method that uses a fuzzy multiple regression model that considers two features of remote sensing data, band data and water quality data⁽⁴⁾. To obtain maps that show good agreement with experimental data using the fuzzy regression model, remote sensing data and water quality values should be acquired simultaneously. However,

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

simultaneous observations hamper the construction of water quality estimation maps. We challenged this obstacle by using fuzzy c-means clustering (FCM) to estimate water quality from only remote sensing data⁽⁵⁾. This method considers the effects of specific disturbances and uncertainties on remote sensing data. The results indicated that FCM was effective in estimating the level of pollution in Lake Hachiroko. In addition, we worked to understand the water quality parameters that have a strong relationship with the estimation maps by FCM, using multi-temporal remote sensing data⁽⁶⁾. As a result, it became clear that the estimation maps obtained by FCM were particularly useful for determining suspended solid (SS) amounts in the water quality analysis.

However, FCM has the issue that the initial values given to the classes used in FCM affect the results of FCM. Thus, we are required to study the setting of the best initial values that can obtain good estimation maps. On the other hand, in previous studies, we classified Lake Hachiroko into classes whose degrees of pollution were low and high using FCM. These initial values were set using histogram information obtained from remote sensing data. For this reason, evaluating the water condition in the estimation maps by FCM remained relative.

In this study, we studied the setting of the best initial values in FCM. In addition, based on the water quality data

obtained from the study area, we considered to generate estimation maps that preset ranges of water quality levels. To evaluate the usefulness of FCM, we compared the estimation maps with the water quality data and the result of the fuzzy regression analysis⁽³⁾.

2. Study Area

Lake Hachiroko is located approximately 20 km north of Akita, Japan, and is a freshwater lake with a center latitude of 40°N and longitude of 140°E. The average depth of the lake is approximately 2.8 m and the maximum depth is 10 m. The lake has a surface area of 4732 ha and consists of an east waterway, a west waterway, and an adjustment pond, which are collectively known as Lake Hachiroko. More than 20 rivers flow into the lake. The annual water flow, including rainfall, into the lake is approximately 1.2 km³, and almost the same volume of water flows out from the floodgate to the Sea of Japan. Furthermore, water pollution has become a significant problem in the lake because of incoming agricultural drainage. Green algae blooms occur during the summer⁽⁷⁾. Thus, since 2007, the lake has been designated as part of the “Act on Special Measures Concerning Conservation of Lake Water Quality⁽⁸⁾.” Therefore, it is necessary to understand the details of the surface water quality of the lake and how these are affected by seasons. Fig. 1 shows an outline of Lake Hachiroko and the water quality measurement sites, denoted as St. 1 to St. 5. According to experts on water quality in the study area, it is likely that the east side of the adjustment pond, which receives polluted water from nearby rivers, will become polluted.

3. Materials

3.1 ASTER Data

The ASTER data used for analysis constitute three spectral bands at visible (green and red) and near-infrared wavelengths and have a resolution of 15 m. The approximate scene size is a 60 km wide swath. In this study, data recorded on August 5, 2012, were used for analysis, as shown in Fig. 2⁽⁹⁾.

Fig. 2. ASTER data (RGB: Bands 2, 3, and 1).
(citation: ASTER-VA image courtesy of
NASA/METI/AIST/Japan Spacesystems, and
U.S./Japan ASTER Science Team)

3.2 Water Quality Measurements

Environmental quality parameters are constantly reviewed and updated by the Akita Prefectural Government. From these, five water quality parameters were used for our analysis, namely biochemical oxygen demand (BOD), chemical oxygen demand (COD), SS, total nitrogen (T-N), and total phosphorus (T-P). We selected water quality measurements recorded on August 5, 2012⁽⁸⁾, as listed in Table 1. These data synchronize with the above remote sensing data. There was no rainfall when measurement of water quality and collection of remote sensing data were conducted⁽¹⁰⁾.

4. Data Analysis

First, we performed atmospheric correction to remove atmospheric effects from the ASTER data. Subsequently, mask processing was performed to detect the water areas in the study area. After that, we applied FCM to the extracted area to calculate the degrees of belonging to each class. Before performing FCM, the initial values in FCM were set using ASTER data and water quality data. To obtain the estimation maps, we applied level slice processing to the obtained degrees of belonging. In addition, to evaluate the usefulness of FCM, we compared the estimation maps with the water quality data and the result of the fuzzy regression analysis⁽³⁾. Fig. 3 shows the process of generation estimation maps by FCM.

Table 1. Water quality measurements⁽⁸⁾.
(August 5, 2012)

Measurement site	Water quality parameters				
	BOD (mg/l)	COD (mg/l)	SS (mg/l)	T-N (mg/l)	T-P (mg/l)
Ogata bridge	20.0	28.0	50	4.50	0.320
East of adjustment pond	17.0	26.0	47	2.40	0.260
West of adjustment pond	4.5	12.0	17	1.10	0.087
Floodgate	3.0	8.1	8	0.730	0.067
Center of site	22.0	40.0	63	6.50	0.380

Fig. 3. Flowchart of generation estimation maps by FCM.

4.1 Preprocessing

4.1.1 Atmospheric Correction

Scattering and absorption effects of the atmosphere affect approximately 90% of the luminance values of remote sensing data acquired in the visible wavelength range. Therefore, atmospheric correction⁽¹¹⁾ of the remote sensing data obtained in the visible wavelength region was performed based on the assumption of negligible water-leaving radiance in near-infrared wavelengths.

4.1.2 Mask Processing

Mask processing is used in image processing to extract only a specific part and exclude other parts from an image. By performing mask processing on all data used for analysis, target water areas were extracted.

Table 2. Slice level.

Slice level	Water quality parameters				
	BOD (mg/l)	COD (mg/l)	SS (mg/l)	T-N (mg/l)	T-P (mg/l)
1	-1.0	-1.0	-1	-0.10	-0.005
2	1.0-2.0	1.0-3.0	1-5	0.10-0.20	0.005-0.010
3	2.0-3.0	3.0-5.0	5-15	0.20-0.40	0.010-0.030
4	3.0-5.0	5.0-8.0	15-25	0.40-0.60	0.030-0.050
5	5.0-8.0	8.0-12.0	25-35	0.60-1.00	0.050-0.100
6	8.0-	12.0-	35-	1.00-	0.100-

4.2 Estimation Maps Generated by FCM

4.2.1 FCM

Remote sensing data include specific disturbances such as noise from the measurement system, and it is necessary to consider each instance of fuzziness in the data. FCM, a kind of clustering method, was used to consider fuzziness in the remote sensing data. A flowchart of the FCM is shown in Fig. 4. We classified the study area into two classes (Class 1 and Class 2) using FCM. C_1 and C_2 were defined from the center value of Class 1 and Class 2, respectively. The weight coefficient was set at 5 because it changes from 2 to 10 at an interval of 0.1. The clustering process stops when the number of moving pixels between the two classes reaches 1% or less of the total number of pixels.

4.2.2 The Initial Values Setting in FCM

To perform classification by FCM, it is necessary to set initial values to C_1 and C_2 . The procedure of initial values setting of FCM is as follows:

- (1) Change the water quality data to the slice levels based on the environmental standard value⁽⁸⁾.
- (2) Select S_{C1} and S_{C2} which are the water quality measurement sites with the minimum and maximum slice level.
- (3) Take a sample of DN values in 50 points from S_{C1} and S_{C2} in the ASTER data.
- (4) Calculate the means of the sampled DN values, defined as mean DN.
- (5) Set the mean DN values as the initial values of C_1 and C_2 .

Slice levels are listed in Table 2. S_{C1} was set to Floodgate, whose all water quality parameters belonged to the lowest slice level. S_{C2} was set to East of adjustment pond. This is because the initial value there acquired the best estimation maps among the three sites whose all water

Fig. 4. Flowchart of the FCM.

quality parameters belonged to the highest slice level.

An example of initial values setting in FCM is shown in Fig. 5.

4.2.3 Level Slice Processing

To obtain estimation maps, level slice processing was performed. The degree of belonging of Class 2 was divided into L_{C1} to L_{C2} at equal intervals, where L_{C1} and L_{C2} are the slice levels in S_{C1} and S_{C2} .

5. Result and Discussion

5.1 Analysis of Water Quality Conditions

Table 3 lists the mean DN values that were set to the initial values of C_1 and C_2 in the FCM. The initial value of C_1 of band 1 data was a higher value than that of C_2 . As shown Table 1, the water quality value of Floodgate was a lower value than that of East of adjustment pond. Information of green algae on the surface of Lake Hachiroko was not obtained due to effects of the reflectance from the water surface in band 1. Therefore, band 1 data were removed from consideration.

Fig. 5. An example of initial values setting in FCM (Band 3 data, SS)

To evaluate the estimation maps by FCM, we compared the slice levels in the maps and the water quality data at each measurement site. When the each measurement sites in estimation maps has the same slice level as the water quality data, we determined as “in agreement.” Table 4 lists the comparison result of the estimation maps by FCM and the water quality data. In Table 4, “A” denotes being in agreement with the water quality data, and “B” denotes not being in agreement with the data.

The estimation maps for SS using Band 3 data are shown in Fig. 6. The L_{C1} and L_{C2} used to obtain estimation maps of the water quality parameter SS are three and six, respectively. The estimation maps are in agreement with the water quality data at Ogata bridge, East of adjustment pond, West of adjustment pond, and Floodgate. The estimation map near an estuary of East of adjustment pond, which received polluted water, is in agreement with the actual pollution condition because the slice level is six.

The estimation maps for BOD and T-P using Band 3 data are shown in Fig. 7. The L_{C1} and L_{C2} used to obtain estimation maps of the water quality parameter BOD are four and six, and the L_{C1} and L_{C2} used to obtain estimation

Table 3. Initial values of C_1 and C_2 for each band.

Initial value	Mean DN of band data		
	Band 1	Band 2	Band 3
C_1	104.042	97.352	20.300
C_2	98.024	103.858	30.840

C_1 : Mean DN value obtained from Floodgate

C_2 : Mean DN value obtained from East of adjustment pond

Table 4. Result of water quality analysis.

(a) Band 2 data.

Measurement site	Water quality parameters				
	BOD (mg/l)	COD (mg/l)	SS (mg/l)	T-N (mg/l)	T-P (mg/l)
Ogata bridge	A	A	A	A	A
East of adjustment pond	A	A	A	A	A
West of adjustment pond	A	B	A	B	A
Floodgate	A	A	A	A	A
Center of site	B	B	B	B	B

(b) Band 3 data.

Measurement site	Water quality parameters				
	BOD (mg/l)	COD (mg/l)	SS (mg/l)	T-N (mg/l)	T-P (mg/l)
Ogata bridge	A	A	A	A	A
East of adjustment pond	A	A	A	A	A
West of adjustment pond	A	B	A	B	A
Floodgate	A	A	A	A	A
Center of site	B	B	B	B	B

A: Estimation results are in agreement with water quality data.

B: Estimation results are not in agreement with water quality data.

maps of the water quality parameter T-P are five and six, respectively. These estimation maps are in agreement with the water quality data at Ogata bridge, East of adjustment pond, West of adjustment pond, and Floodgate. Therefore, we showed that the estimation maps obtained by FCM are in agreement with the water quality condition in Lake Hachiroko, which is also true in the case of the setting of the slice levels for classes of FCM being different.

According to the above results and Table 4, the estimation maps by FCM at four measurements sites are in agreement with the BOD, SS, and T-P of the water quality information. This finding indicates that the water quality parameters at occurrences of green algae and the DN values have a strong correlation.

Fig. 6. Estimation maps generated by FCM.
(Band 3 data, SS, $L_{C1} = 3$, $L_{C2} = 6$)

(a) BOD (Band 3 data, $L_{C1} = 4$, $L_{C2} = 6$)

(b) T-P (Band 3 data, $L_{C1} = 5$, $L_{C2} = 6$)

Fig. 7. Estimation maps generated by FCM.

5.2 Comparison with Fuzzy Regression Analysis Results

The estimation maps by fuzzy regression analysis⁽³⁾ for T-P, SS, and BOD using Band 3 data are shown in Fig. 8. The estimation map for T-P by fuzzy regression analysis is in agreement with the water quality data at Ogata bridge, East of adjustment pond, West of adjustment pond, and Floodgate. However, the estimation maps for SS and BOD

(a) T-P

(b) SS

(c) BOD

Fig. 8. Estimation maps by fuzzy regression analysis⁽³⁾.
(Band 3 data, Min problem)

by fuzzy regression analysis show a lower level than the water quality data at West of adjustment pond. Moreover, the estimation maps show a higher level than the water quality data at Floodgate. On the other hand, the estimation maps for SS and BOD by FCM are in agreement with the water quality data at West of adjustment pond and Floodgate. Therefore, the estimation maps generated by FCM are well associated with the water quality conditions.

The above result indicates that the estimation maps generated by FCM reveal the detailed water quality conditions in Lake Hachiroko.

6. Conclusions

In this research, we studied the setting of the initial values of FCM. Additionally, to evaluate the usefulness of FCM, we compared the estimation maps with the water quality data and the result of the fuzzy regression analysis. The following conclusions were reached:

- (1) The estimation maps generated by FCM were in agreement with the actual pollution conditions in Lake Hachiroko.
- (2) The estimation maps generated by FCM had a strong relationship with the BOD, SS, and T-P measured in Lake Hachiroko.
- (3) The water quality parameters at occurrences of green algae and the DN values showed a strong correlation.
- (4) The estimation maps generated by FCM showed detailed water quality conditions in Lake Hachiroko.

References

- (1) Claudia Giardino, Vittorio E. Brando, Arnold G. Dekker, Niklas Strömbeck, and Gabriele Candiani : “Assessment of Water Quality in Lake Garda (Italy) Using Hyperion”, *Remote Sensing of Environment*, Vol. 109, No. 2, pp. 183-195, 2007
- (2) Oyama Youichi, Matsushita Bunkei, Fukushima Takehiko, Matsushige Kazuo, and Imai Akio : “Application of Spectral Decomposition Algorithm for Mapping Water Quality in a Turbid Lake (Lake Kasumigaura, Japan) from Landsat TM Data”, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 64, No. 1, pp. 73-85, 2009
- (3) Dejian Wang, Yoichi Kageyama, Makoto Nishida, Hikaru Shirai, and Asana Motozawa : “Water Quality Analysis in Lake Hachiroko, Japan, Using ALOS AVNIR-2 Data”, *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 8, No. 6, pp. 627-633, 2013
- (4) Dejian Wang, Yoichi Kageyama, Makoto Nishida, Hikaru Shirai, and Asana Motozawa : “Analysis of Water Quality of Lake Hachiroko, Japan, Using Fuzzy Multiple Regression Model with ALOS AVNIR-2 Data”, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 20, No. 6, pp. 992-1003, 2016
- (5) Yoichi Kageyama, Asako Izumi, Makoto Nishida, and Hiroshi Yokoyama : “Application of Fuzzy C-means for Understanding Water Quality in Lake Hachiroko, Japan”, *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 11, No. 6, pp. 835-837, 2016
- (6) Kai Matsui, Yoichi Kageyama, Makoto Nishida, and Hiroshi Yokoyama : “Analysis of Water Quality Conditions with Fuzzy C-Means Using Terra Aster Data and Landsat-8 Data”, *Int. Symp. on Remote Sensing 2017 (Nagoya, Japan)*, P-138, pp. 895-898, 2017
- (7) <http://www.pref.akita.lg.jp/> [accessed July 17, 2018]
- (8) www.env.go.jp/ [accessed July 17, 2018]
- (9) <http://landbrowser.geogrid.org/landbrowser/> [accessed July 17, 2018]
- (10) <http://www.jma.go.jp/> [accessed July 17, 2018]
- (11) Mikio Takagi and Haruhisa Shimoda : “Handbook of Image Analysis [Revised Edition]”, University of Tokyo Press, 2004