USE OF UNMANNED AERIAL VEHICLE DATA IN NEAR-INFRARED REGION TO ESTIMATE WATER QUALITY OF MIHARU DAM RESERVOIR, JAPAN

Yoichi Kageyama (1), Shin Totsuka (1),
Masato Ishikawa (2), Bunyu Kobori (2), Daisuke Nagamoto (3)

1 Akita University, 1-1, Tegata Gakuen-machi, Akita-shi, Akita 010-8502, Japan
2 Civil Engineering & Eco-Technology Consultants Co., Sendai TB Building 6F, 4-3-10, Tsutsujigaoka, Miyagino-ku, Sendai-shi, Miyagi 983-0852, Japan
3 Civil Engineering & Eco-Technology Consultants Co., KN Building 2F Kita 3-jo Higashi 3-1-30, Chuo-ku, Sapporo, Hokkaido, 060-0033, Japan
Email: kageyama@ie.akita-u.ac.jp

KEY WORDS: Blue-green algae, Fuzzy regression model, Fuzzy c-means

ABSTRACT: Lake Sakurako is a reservoir of the Miharu dam in Fukushima Prefecture, Japan. The water quality of the small lake becomes significantly worse during the summer, owing to the occurrence of blue-green algae. Taking into account the limited water quality monitoring data available for the lake, we previously used a fuzzy regression analysis (FRA) of water quality measurements and water conditions that appear in near-infrared (NIR) data collected by unmanned aerial vehicles (UAVs). Then, fuzzy c-means (FCM), which shows relative differences in water quality, was also applied for the analysis. Furthermore, we investigated a noise removal process using a non-local mean (NLM) filter and demonstrated that the process provides more detailed information regarding the lake's water quality. However, a comparison of classification results with respect to differences in analysis methods has not yet been conducted. Therefore, this paper describes the differences in classification results obtained by both FRA and FCM using an NLM filter. Water quality data sampled at 20 points synchronized with UAV data were acquired. Five water quality parameters were directly measured. The analysis method adopted is comprised of preprocessing, NLM filtering, FRA, fuzzy level slices, and FCM. FRA assumes that the differences between observation data and the model prediction indicate the system fuzziness, thus revealing the relation between the input and the output. FRA was carried out for each combination of data: the UAV NIR data and the measurements of water quality parameters. For FCM, first, the study area was divided into two classes: C1 and C2. The initial point of C1 was selected from an average value of 2% from the minimum value of the study area, and the initial point of C2 was selected from an average value of 2% from the maximum value of the study area. Then, the degree of belonging to C2 was divided into preset levels. The results suggest that the application of both FRA and FCM using an NLM filter to understand the water quality is more effective than the simple FRA method. When the presence of blue-green algae was high, it became clear that FRA using an NLM filter can estimate the water quality more accurately than FCM on a 256-level gray scale.

1. INTRODUCTION

Water pollution in lakes has become a problem owing to industrial activities and changes in daily life. It is, therefore, necessary to both estimate the water quality in the environment and preserve it. The water quality of lakes is tested regularly to monitor the levels of water pollution. A typical investigation method involves taking water samples directly from certain points. Although this approach is well-suited to collecting water quality data for a relatively small area, there are difficulties in applying it for monitoring water quality over large areas.

Lake Sakurako is a reservoir of the Miharu dam in Fukushima Prefecture, Japan. The water quality of the small lake becomes significantly worse during the summer, owing to the occurrence of blue-green algae. Taking into account the limited water quality monitoring data available for the lake, we previously used a fuzzy regression analysis (FRA) of water quality measurements and water conditions that appear in near-infrared (NIR) data collected by unmanned aerial vehicles (UAVs) (Kageyama et al., 2016a). As a result, the analysis of the UAV data reflected the water quality conditions satisfactorily, indicating that an analysis using NIR data is the most effective method for water quality estimation from UAV data. Then, fuzzy c-means (FCM), which shows relative differences in water quality, was also applied for the analysis (Kageyama et al., 2018). Furthermore, we investigated a noise removal process using a non-local mean (NLM) filter and demonstrated that the process provides more detailed information regarding the lake's water quality (Totsuka et al., 2019). However, a comparison of classification results with respect to differences in analysis methods has not yet been conducted. Therefore, this paper describes the differences in classification results appeared in UAV NIR data by both FRA and FCM using an NLM filter.

2. STUDY AREA AND MATERIALS

2.1 Study Area

Lake Sakurako is a reservoir of Miharu Dam and has a complicated shape; both domestic and industrial wastewater flow into Lake Sakurako. The water quality becomes eutrophic, and the occurrence of blue-green algae becomes pronounced in summer. To solve this water problem, various attempts have been made to conduct water quality preservation measures, e.g., pre-reservoir, influent bypass pipe, shallow layer circulation, and deep layer aeration (Miharu Dam official site). However, these attempts have not yet achieved a reliable solution.

2.2 Water Quality Data and UAV Data

Figure 1 shows the target area from the pre-reservoir in Lake Sakurako to the Fudodakibashi Bridge to estimate the presence of blue-green algae. Water quality data sampled at 20 points synchronized with UAV data were acquired. Five water quality parameters were directly measured: suspended solids (SS), total nitrogen (T-N), total phosphorus (T-P), chlorophyll a, and phycocyanin. We used UAV data acquired on August 12 (August data A) and September 4 (September data A), 2015, and August 4 (August data B) and August 9 (August data C), 2016. These water quality parameters were used both individually and in combination with the UAV data, and were applied to an FRA.

3. DATA ANALYSIS

The analysis method adopted in this study is comprised of preprocessing, NLM filtering, FRA, fuzzy level slices, and FCM. Figure 2 shows a flowchart of the analysis for the proposed methods A and B, and the previous method (simple FRA).

3.1 Preprocessing

Although the resolution of the employed UAV data is approximately 3 cm, it cannot accurately maintain its characteristics in estimation maps when it considers water sampling. Therefore, coarse-graining was performed to speed up the classification and reduce noise. A filter size of 5 × 5 pixels was set, and the median value among the 25 pixels was selected as the value of the target pixel. By performing coarse-graining, the data resolution was changed from ~3 to ~15 cm. Then, mask processing was applied to the land area in order to extract the water area.

3.2 NLM Filter

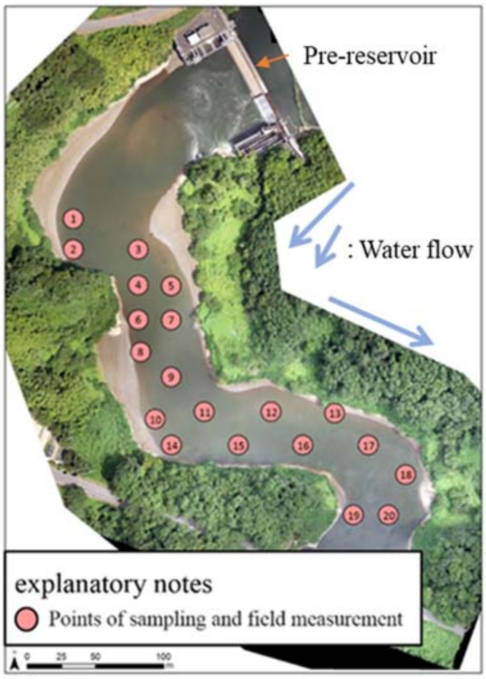


Figure 1 Water sampling points in the target area.

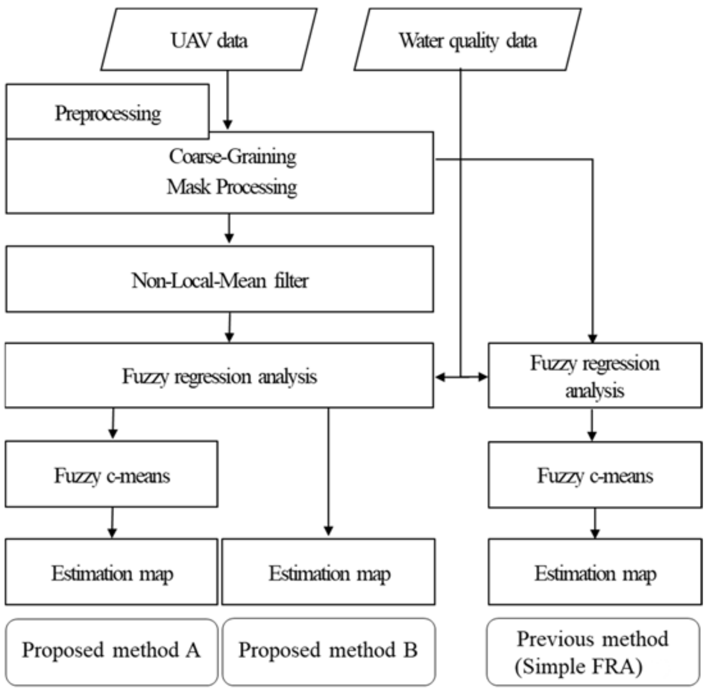


Figure 2 Flowchart of data analysis.

To reduce the noise in the UAV data, an NLM filter that performs averaging by using weights according to the similarity of pixel value patterns in addition to distance weights was used for small regions around the pixel of interest (Totsuka et al., 2019). The filter size of 3 × 3 pixels was set.

$$v(\vec{r}) = \sum_{\vec{s} \in I} w(\vec{r}, \vec{s}) u(\vec{s}) \tag{1}$$

Here, $\vec{r}$ and $\vec{s}$ are position vectors, $u(\vec{r})$ is the data before filtering, $v(\vec{r})$ is the data after filtering, $w(\vec{r}, \vec{s})$ is a weight based on similarity, and I is the entire image. The filtered data ($v(\vec{r})$) is a weighted sum of the point of interest ($\vec{r}$) and the entire image ($\vec{s} \in I$).

The weighting based on similarity is performed as follows:

$$w(\vec{r}, \vec{s}) = \frac{1}{Z(\vec{r})} \exp\left(-\frac{\left\|\vec{N}(\vec{r}) - \vec{N}(\vec{s})\right\|^2}{\sigma^2}\right), \tag{2}$$

$$Z(\vec{r}) = \sum_{\vec{s} \in I} \exp\left(-\frac{\left\|\vec{N}(\vec{r}) - \vec{N}(\vec{s})\right\|^2}{\sigma^2}\right) \tag{3}$$

In the equations, $\vec{N}(\vec{r})$ is a vector in which the peripheral pixels of $\vec{r}$ are arranged. The difference between $\vec{N}(\vec{r})$ and $\vec{N}(\vec{s})$ is($\boldsymbol{D}$), which is the Euclidean distance of the surrounding pixel vector. By using the Gaussian function $\exp(-(\boldsymbol{D}/\boldsymbol{\sigma}))$, the distance 0 is the heaviest, and weighting is performed to reduce the weights by moving away from 0. $\boldsymbol{\sigma}$ is a parameter. $Z(\vec{r})$ is the sum of the weights and is used for normalization.

3.3 **Fuzzy Regression Analysis**

Fuzzy set theory provides useful concepts and tools for addressing uncertainties. The fuzzy regression model assumes that the differences between observation data and the model prediction indicate the system fuzziness, thus revealing the relationship between the input and the output (Ishibuchi, 1992). In our previous studies, FRA was also applied to UAV data and measurements of water quality parameters (Kageyama et al., 2016a).

The fuzzy regression model was computed using the following equations:

$$Y(X_p) = A_0 + A_1 X_{p1} + \cdots + A_n X_{pn} \tag{4}$$
$$= (a(X_p), e(X_p))_L, \tag{5}$$

where

$$a(X_p) = a_0 + a_1 X_{p1} + \ldots + a_n X_{pn}, \tag{6}$$
$$e(X_p) = e_0 + e_1 \mid X_{p1} \mid + \ldots + e_n \mid X_{pn} \mid, \tag{7}$$

A_i is a fuzzy coefficient, $Y(X_p)$ is an estimate (fuzzy number), $a(X_p)$ is the center (mean) of $Y(X_p)$, and $e(X_p)$ is the width of $Y(X_p)$. Equation (5) is the fuzzy regression model expressed by the center and width of the fuzzy number. Because the fuzzy regression model has one input and one output, four variables (a_0, e_0, a_1, e_1) are obtained from the following equations:

$$Y(X_p) = A_0 + A_1 X_p \tag{8}$$
$$= (a_0, e_0)_L + (a_1, e_1)_L X_p, \tag{9}$$

where X_p is a measure of water quality, A_1 is a triangular fuzzy number, a_1 is the center of the fuzzy number, e_1 is the width of the fuzzy number, and $Y(X_p)$ is the fuzzy output interval estimation.

In this work, the triangular fuzzy number produced from the vertex (a_1) and the width (e_1), which was the mean of the DN and twice the standard deviation $(ó)$ of 25 pixels around each water quality measurement site, was used for the analysis. In FRA, linear programming problems referred to as minimum and maximum problems can be formulated using interval data (Ishibuchi, 1992). Here, the minimum problem is a process that requires a linear regression model with a width minimum that includes all the interval data. Conversely, the maximum problem is a process that requires a linear regression model with a width maximum that is included in all the interval data. We analyzed the minimum problem because it is effective in estimating the water surface conditions of the lake (Kageyama et al., 2016a).

3.4 **Fuzzy Level Slice**

The fuzzy output interval obtained using the fuzzy regression model shows that the DN corresponds to the forecast range obtained by measuring the water quality parameters and the UAV data. It has been demonstrated that water quality estimate maps generated using fuzzy level slice processing can yield intermediate levels of water quality that are comparable to those obtained by conventional level slice processing. To understand the water quality in detail, estimate maps were created by fuzzy level slice processing. We assumed that the DN could correspond to a specific level of water quality conditions set over an optional range. The production rule for estimating the water quality for a given pixel is as follows:

$$\begin{array}{l} \textit{Rule} \quad 1: Y_1 \rightarrow Z_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \textit{Rule} \quad n: Y_n \rightarrow Z_n, \\ \underline{\textit{Input} \quad : S \qquad\qquad} \\ \textit{Output} \quad : \qquad Z_0 \end{array} \tag{10}$$

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 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, as determined by water quality experts. S is the input for the DN. Z_0 is the final output, and is given as follows:

$$Z_0 = \frac{\sum_{i=1}^{n} h_i Z_i}{\sum_{i=1}^{n} h_i} \tag{11}$$

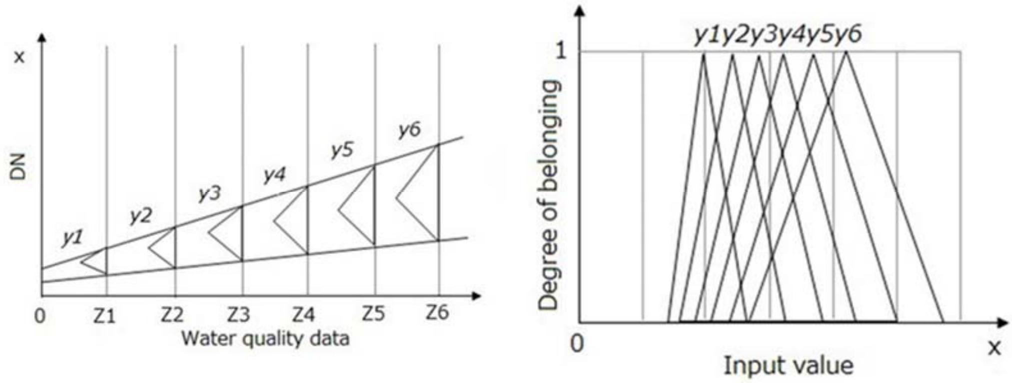
When the input S is known, h_i is the ratio for obtaining Z_i. The rule number corresponds to the slice number. In this study, we used seven slices. Fig. 3 shows an example of a seven-slice fuzzy regression model and a fuzzy set. For example, we assume seven rules of DN corresponding to each water quality measurement. In Eq. (10), Y_i is the DN and Z_i corresponds to the water quality measurements. The relationship of the rules is shown in Fig. 3(a). When the input S is known, we can calculate the ratio h_i for each rule n, as shown in Fig. 3(b). We substitute the values of h_i and Z_i in Eq. (11), and calculate the output Z_0 for input S.

3.5 **Grayscale Image Creation**

For the values computed by the above process, the water estimation map was produced by outputting it with seven gradation levels. We assume that a more detailed water quality analysis becomes possible by increasing the number of gradations. The definite value of Y, as calculated by performing fuzzy level slice processing on the UAV data, was converted to 256 gradation levels using Eq. (12).

$$Y' = \frac{Y \quad - Y_{min}}{WQ_{max} - Y_{min}} \times 255 \tag{12}$$

Here, Y' is the DN value after the gradation conversion, WQ_{max} is the maximum possible value of each water quality item, and Y_{min} is the minimum value of Y.



(a) Fuzzy regression model (b) Fuzzy set

Figure 3 Example of a seven-slice fuzzy regression model and a fuzzy set.

3.6 **Fuzzy C-Means**

The reflection caused by the terrain was large, owing to the complicated and narrow shape of the study area. Therefore, water estimation maps were drawn using FCM used in Reference (Kageyama et al., 2016b), which uses the analysis results of FRA as input values. The input values were fuzzy numbers and FCM is a clustering algorithm that allows one pixel to belong to two or more classes. FCM was adopted to show the relative differences in water quality. Figure 4 shows a flowchart of FCM. First, the target area was divided into two classes: C1 and C2. The DNs of C1 were low, and those in C2 were high. Second, the clustering process was completed when the number of moving pixels between the two classes reached 1% or fewer of the total number of pixels. The initial point of C1 was selected from an average value of 2% from the minimum value of the study area, and the initial point of C2 was selected from an average value of 2% from the maximum value, excluding noise from the histogram information computed in the study area. Finally, the degree of belonging to C2 was divided into preset levels.

4. RESULTS AND DISCUSSION

4.1 **Evaluation Criteria for Estimation Maps**

To evaluate the water quality conditions, we visually compared the estimation maps with the water quality conditions. The results were evaluated as follows:

◎: These results reflect the water quality data, and are consistent with the ground survey results.

○: Most parts of the results reflect the water quality data, and are broadly consistent with the ground survey results.

△: In the fuzzy regression analysis, a solution is obtained, but the output image does not reflect the water quality condition.

×: In the fuzzy regression analysis, a solution is obtained; however, it cannot be used for analysis.

4.2 **Comparison Between Proposed Method A and the Previous Method**

Tables 1 and 2 show a comparison between the analysis results of the proposed method A and the previous method, simple FRA (Kageyama et al., 2016a), using four scenes of UAV data. They were acquired August 12 (August data A) and September 4 (September data A), 2015, and August 4 (August data B) and August 9 (August data C), 2016. In the results using the data acquired in 2016, there were few items that were evaluated as ○. Compared with the data acquired in 2015, the 2016 data (August data B and C) had lower water quality values throughout the water areas, making it difficult to correlate in the FRA. However, the proposed method A obtained results that better reflected the water quality conditions compared with those of the previous method.

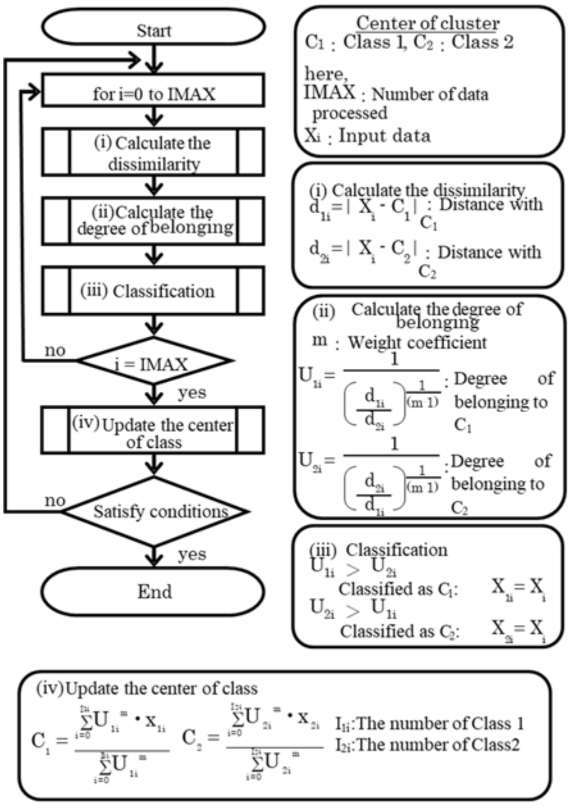


Figure 4 Flowchart of FCM.

Table 1 Comparison results with water quality distribution map by proposed method A.

	Water parameters				
	SS	T-N	T-P	chlorophyll a	phycocyanin
August data A	○	○	○	○	○
September data A	○	○	△	○	○
August data B	○	△	×	×	○
August data C	○	×	×	×	×

Table 2 Comparison results with water quality distribution map by previous method.

	Water parameters				
	SS	T-N	T-P	chlorophyll a	phycocyanin
August data A	○	×	○	○	△
September data A	○	×	○	×	△
August data B	○	×	×	○	×
August data C	○	×	×	×	×

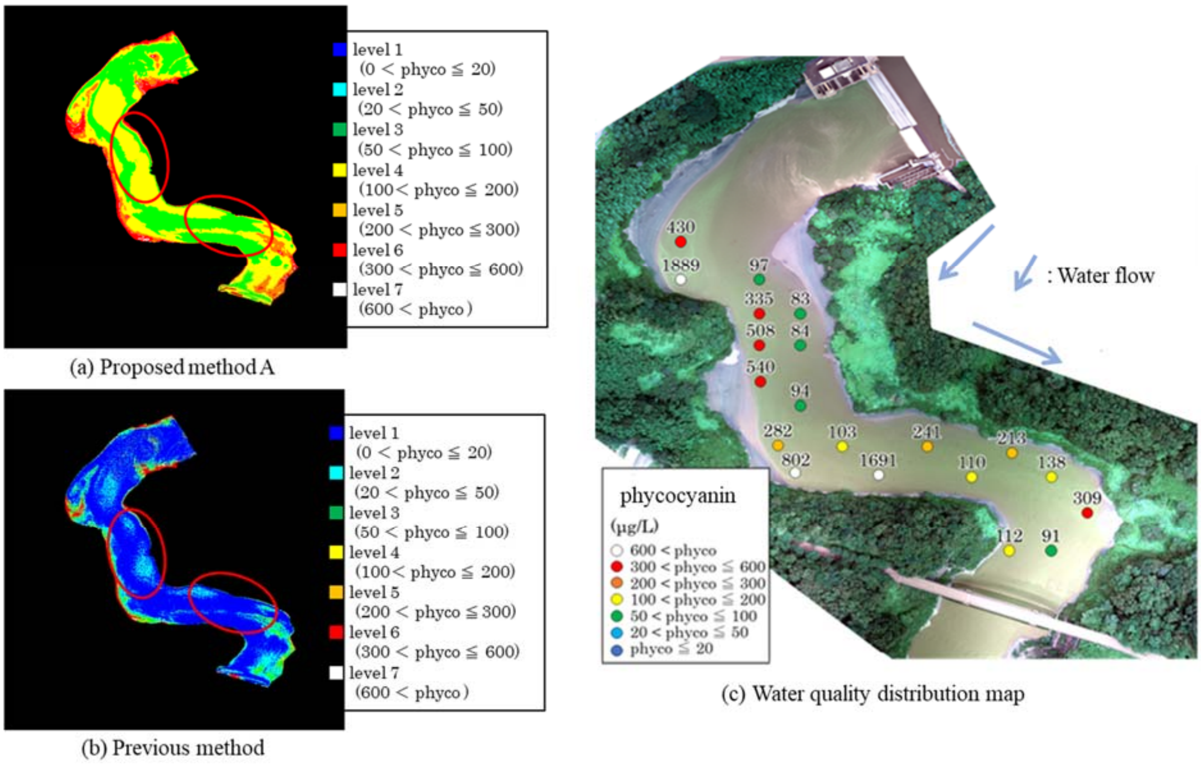


Figure 5 Water estimation maps (August data A; phycocyanin).

Figures 5 and 6 show the estimation maps. In Figure 5(b), the result by the previous method indicates a lower level (blue or light blue) than the regions with actual pollution levels, indicated by red circles. As shown in Figure 6 (b), the previous method estimates a higher level (white) than the regions with low actual pollution levels (blue), indicated by red circles. However, the proposed method A is useful in estimating water quality conditions, as listed in Table 1.

4.3 Detailed Estimation of Water Quality During the Occurrence of Blue-Green Algae

In order to estimate the detailed water quality at the time when blue-green algae occurred frequently, the proposed methods A and B draw the estimation maps using August data A. Table 3 lists the results of the comparison between the analysis results of the proposed methods A and B. In the result using the proposed method A, there is an item (T-N) that was evaluated as △ in the 256-level analysis, even though the evaluation result was ○ in the 7-level analysis result. When drawing maps of 256 gradations, the features that appeared in 7 gradations are integrated, such that the feature is offset and the difference is shown in the evaluation results. In the analysis using the proposed method B, all the water quality items were evaluated as ○.

Figure 7 shows the estimation maps in 256 gradations. When outputting 256 gradation levels, the proposed method B obtained good results in reflecting the water quality conditions compared with those of the proposed method A.

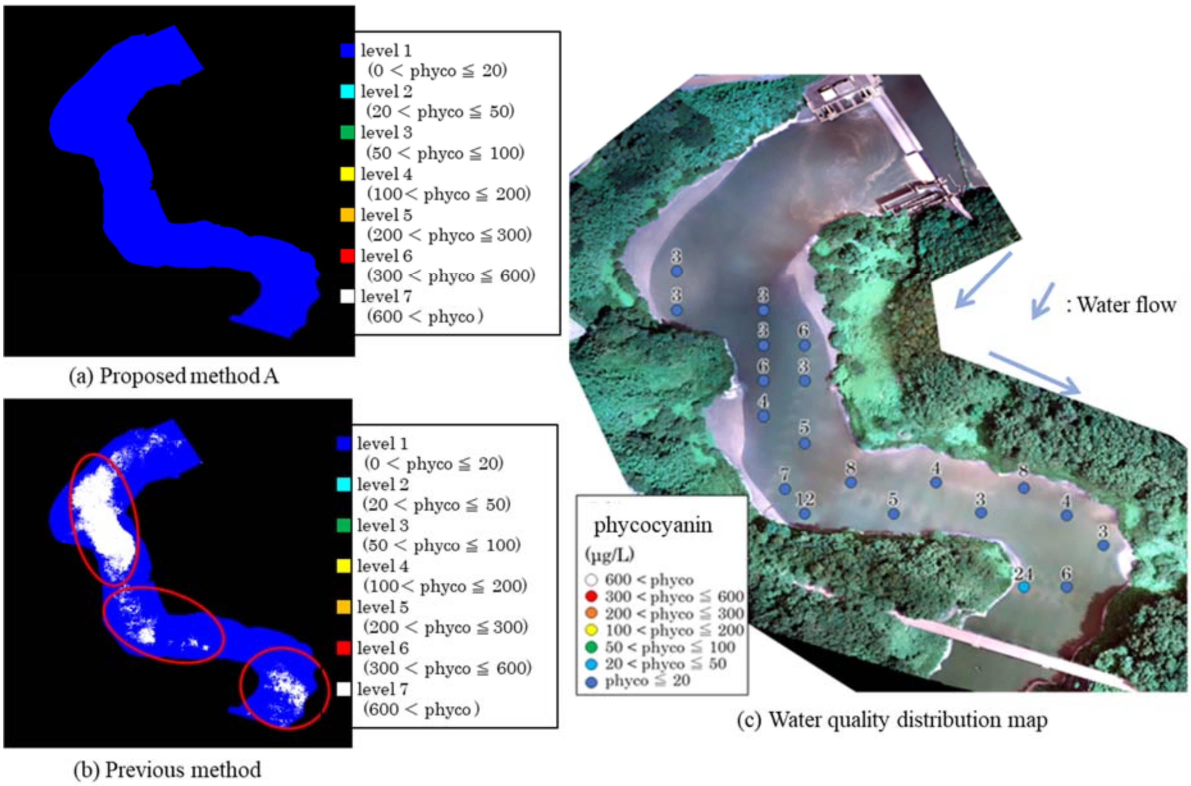


Figure 6 Water estimation maps (August data B; phycocyanin).

Table 3 Comparison results with water quality distribution map by proposed method A and B (August data A).

	Water parameters				
	SS	T-N	T-P	chlorophyll a	phycocyanin
Proposed method A	○	△	○	○	○
Proposed method B	○	○	○	○	○

5. CONCLUSIONS

In this study, we examined the differences in classification results that appeared in UAV NIR data by both FRA and FCM using an NLM filter. The following conclusions were obtained:

1. It was clarified that the estimation of water quality in consideration of noise obtained better results reflecting water quality data than the previous method (simple FRA), and was useful for understanding the water quality of Lake Sakurako.

2. When the blue-green algae occurred, the proposed method B helped to grasp the detailed water quality situation, by outputting the analysis results in 256 gray scales.

The authors thank the Miharu Dam Management Office for their help in conducting the ground survey.

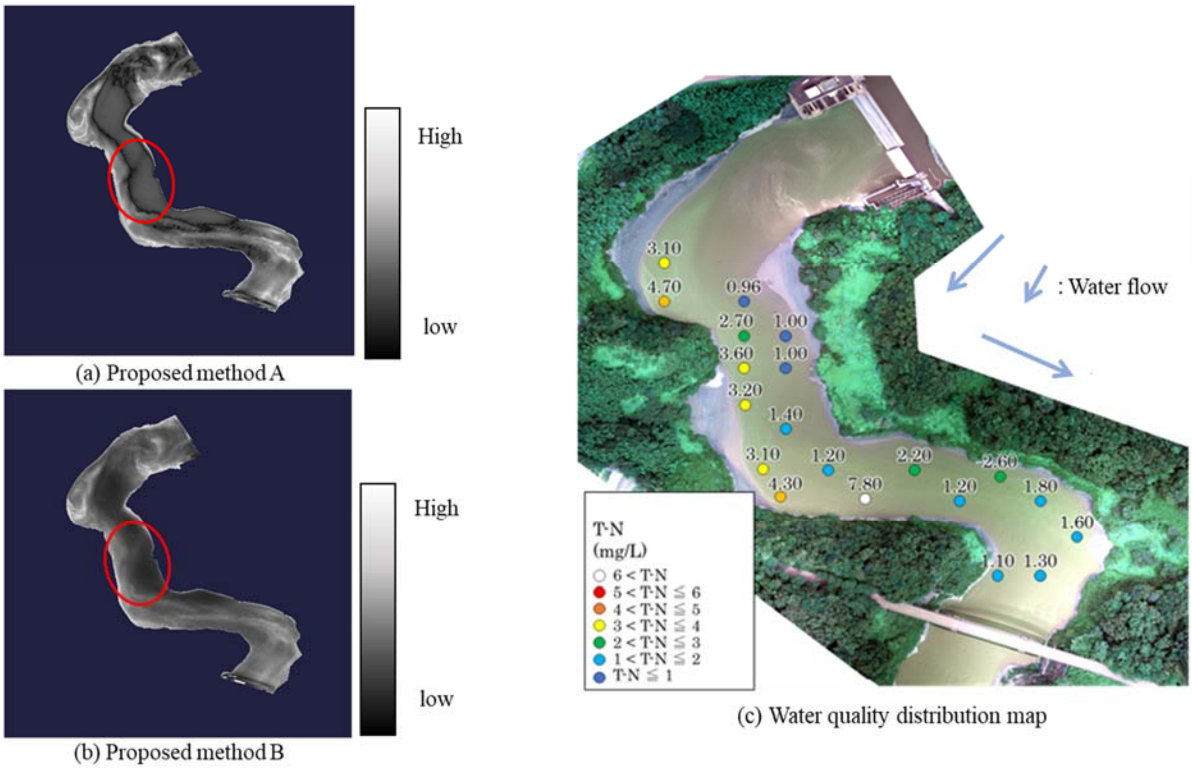


Figure 7 Water estimation maps in 256 gradations (August data A; T-N).

REFERENCES

Ishibuchi, H., 1992, Fuzzy regression analysis, Journal of Japan Society for Fuzzy Theory and Systems, 4(1), pp.52–60.

Miharu Dam official site: http://www.thr.mlit.go.jp/miharu/ , Retrieved August 18, 2019.

Kageyama, Y., Takahashi, J., Nishida., M., Kobori, B., Nagamoto, D., 2016a, Analysis of Water Quality in Miharu Dam Reservoir, Japan, Using UAV Data, IEEJ Transactions on Electrical and Electronic Engineering, 11(S1), pp.183–185.

Kageyama, Y., Izumi, A., Nishida, M., Yokoyama H., 2016b, Application of Fuzzy C-means for Understanding Water Quality in Lake Hachiroko, Japan, IEEJ Transactions on Electrical and Electronic Engineering, 11(6), pp.835-837.

Kageyama, Y., Wakatabe, K., Ishikawa, M., Kobori, B., Nagamoto, D., 2018, Application of Fuzzy Regression Analysis and Fuzzy C-means Technique Using UAV Data to Understand Water Quality in the Miharu Dam Reservoir, Japan, IEEJ Transactions on Electrical and Electronic Engineering, 13(12), pp.1831-1832.

Totsuka, S., Kageyama, Y., Ishikawa, M., Kobori, B., Nagamoto, D., 2019, Noise Removal Method for Unmanned Aerial Vehicle Data to Estimate Water Quality of Miharu Dam Reservoir, Japan, J. Adv. Comput. Intell. Intell. Inform., 23(1), pp. 34-41.