Article

Not peer-reviewed version

Estimation Method of Chlorophyll Concentration Distribution Based on UAV Aerial Images considering Turbid Water Distribution in a Reservoir

Mitsuteru Irie * , Yugen Manabe , Masafumi Yamashita

Posted Date: 5 April 2024

doi: 10.20944/preprints202404.0428.v1

Keywords: Algae; Chlorophyll-a; Turbidity; Reservoir; Machine learning; UAV; Insolation

Article

Estimation Method of Chlorophyll Concentration Distribution Based on UAV Aerial Images Considering Turbid Water Distribution in a Reservoir

Mitsuteru Irie [1,*], Yugen Manabe [2] and Masafumi Yamashita [3]

[1] Faculty of Engineering, University of Miyazaki, 889-2192, Miyazaki, Japan
[2] NIHON SHINKO, Ltd; yugen1225mnb@gmail.com
[3] Kagoshima prefectural office; masafumi521mh@gmail.com
* Correspondence: irie.mitsuteru.p2@cc.miyazaki-u.ac.jp Tel.: +81-2958-7341

Abstract: The causes of algal blooms in reservoirs are often complexly intertwined with chemical, physical, and biological factors such as the supply of nutrients. Observation of phytoplankton distribution with high spatiotemporal resolution is necessary to track the nutrient sources that cause algal blooms and to understand their behavior in response to wind and water temperature stratification. Observation from a UAV, which has excellent temporal and spatial resolution, is considered to be an effective method to obtain water quality information comprehensively. On the other hand, it is not only the growth of plankton that affects the color of the water surface but also turbidity. Furthermore, since the brightness value of passive sensors such as optical cameras changes depending on the amount of insolation, it is necessary to perform analysis after making corrections for this. In this study, we attempted to develop a method for estimating chlorophyll concentration from aerial images taken from UAVs using machine learning that takes into account brightness correction based on insolation and the spatial distribution of turbidity evaluated by satellite image analysis.

Keywords: Algae; Chlorophyll-a; Turbidity; Reservoir; Machine learning; UAV; Insolation

---

1. Introduction

In enclosed water bodies such as lakes, bays, and reservoirs, the water exchange is gradual, so that the nutrients of nitrogen and phosphorus flowing in from the catchment area tend to accumulate [1]. When particulate phosphorus is deposited deep down, it takes on complex dynamics due to the stratification of the water body. When nutrients eluted by internal flow are delivered to a field with suitable water temperature and illuminance conditions, the bloom of phytoplankton occurs in a field that is spatiotemporally different from the direct supply of nutrients ([2–4]). The predominance of certain species of phytoplankton can cause taste and odor problems that contribute to the degradation of drinking water supplies, inhibit recreational uses of surface waters [5], the reduction of suitable fish habitats, and some assemblages of cyanobacteria produce toxins that are harmful to humans and animals [6,7]. In eutrophic lakes, the species of the Microcystis, which have gas vesicles, often dominate and spread on the whole water surface. Harmful Algal Bloom (HAB) are raised as social problems that pose a major hindrance to the environment and human activities [8–11].

A radical treatment for this problem is to identify the toxic species and find the environmental variation patterns in which the species occurs [12,13], and to control the supply of the nutrients so that they do not exceed thresholds that could lead the algal bloom because it is difficult to control the other factors of the blooming; insolation and water temperature in water bodies. There are many attempts to predict the abnormal occurrence of phytoplankton blooming based on machine learning from various water quality parameters [14–16]. However, as mentioned above, the spatiotemporal behavior of the nutrients within a water body is complex. Even if the species identification and water quality analysis are performed from the water sampled at a limited number of locations in the water body, achieving the goal is difficult, and a correspondingly large-scale effort and cost-intensive investigation is required [17–21].

On the other hand, the sophistication of the simulations in hydrodynamics and that in the associated changes in water quality have progressed rapidly in recent years. The reproducibility and the spatiotemporal resolution of those distributions within a water body become reliable and high precision [22–24]. In proportion to this, the development of observation technology with high spatiotemporal resolution for HABs is expected. This is because the distribution of HABs can change rapidly due to variations in population dynamics and environmental conditions. Normal water quality analysis and microscopic observation require a large amount of time and labor even for just one sample, so it is difficult to understand the entire dynamics of HABs. Except for the above example, there are very few cases where a sufficient number of observations have been made to enable discussion about the transportation of nutrients by currents within the water body.

Remote sensing enables the spatiotemporally high-density observations required for HAB observation in a wide area of water at a relatively low cost. This is because phytoplankton generally proliferates only in shallow layers where sunlight reaches from the surface of the water. In addition, the species of Microcystis, which is one of the groups of HAB, is floating with gas vesicles. Therefore, it is highly compatible with remote sensing from the sky, which measures the intensity of reflection on the water surface, especially in large lakes and coastal areas. There are many examples of such observations. Li et.al. [27] used multiple satellite image analyses to observe the green tide, which occurs in the southern Yellow Sea and moves northward, severely impacting the coastal ecological environment. Hu et. al. [28] developed an objective method to estimate the biomass of *Ulva prolifera* using satellite image analysis and clarified the transition of its blooming area. In general, the satellite data such as MODIS that comes in every day has a coarse spatial resolution, while the satellites with high spatial resolution such as Sentinel come in once every 1-2 weeks, and the images can only be captured if the sky is clear on that day. In satellite image analysis, it is often necessary to choose either spatial or temporal resolution.

Chen et. al. [29] calculated the Floating Algae Index (FAI) based on the Advanced Himawari Imager (AHI) data acquired by the geostationary meteorological satellite for the detection of the diurnal algae dynamics in Lake Tai, China. The observations by the geostationary satellites guarantee exceptionally high resolution and imaging frequency. However, such satellite data is not obtained in every region. In analyses based on data from global observation satellites, which can be applied anywhere, it is necessary to select either spatial resolution or inbound frequency depending on the characteristics of the research objects. To supplement this, there are examples of combining various types of satellite images and discussing the short-term decline of HABs [30–33].

On the other hand, HAB observation based on the analysis of aerial images taken from Unmanned Aerial Vehicles (UAV) is superior to satellite images in terms of spatiotemporal resolution and in that it is not affected by clouds [34]. Using a small Unmanned Aerial Vehicle system equipped with a consumer-grade camera, Qu et al [35] determined surface-floating cyanobacteria at a maximum detection altitude of 80 m. The small UAV can cover up to 1 km^2 per flight mission, and the short time lag between sampling and flight allows for follow-up monitoring and treatment. Guimarães et al. [36] photographed a small reservoir with an NGB (near-infrared (N), green (G), blue (B)) camera connected to a UAV, and extracted normalized differential vegetation index (NDVImod) from orthorectified images. The results of multivariate analysis using each spectrum data as an explanatory variable and NDVImod showed some correlation with chlorophyll-a (Chl-a) concentration. Su et al. [37] tried the estimation of the concentration of Chl-a, Total phosphate and using the average method and pixel-by-pixel matching (MPP) method to search for the optimal regression model from the brightness values of multispectral images obtained by UAV. In each of these research cases, the study site is narrower than in the research based on satellite image analysis, and the observation could be completed with a limited number of images.

Cheng et al. [38] developed the estimation model for Chl-a concentration from the images taken by the digital camera mounted on the UAV and validated the estimated result against the observed data over a year. As the photography duration increases, the insolation conditions change in a day, and the insolation conditions also differ if the photography takes place on different days and in different seasons, so the brightness value of the photograph changes even for the same color on the water surface. In that study, the brightness values of each spectrum were corrected based on the amount of insolation. The need to correct insolation conditions is an issue not only for aerial images taken by UAVs but also when comparing images taken on different dates in satellite image analysis [32,39,40].

In addition to the insolation and Chl-a, turbidity is a factor that greatly affects the color of water surfaces. Kishino et. al [41] proposed using a neural network to estimate the concentration of turbidity and Chl-a at once from the images taken by the Aster satellite. Sakuno et.al [42] developed an integrated algorithm for the remote sensing of Chl-a and turbidity in eutrophic and hyper-turbid waters such as Lake Shinji and Nakaumi in Japan and Vaal Dam Reservoir in South Africa. Palmer et. Al. [43] investigated indicators for evaluating the distribution of Chl-a concentration in shallow lakes with extremely high turbidity.

Most of the studies on the wide-area observation of HABs using satellite images or UAVs have mainly focused on natural lakes, rivers, and seawater, and only a few studies have been conducted on reservoirs though the water quality management of a reservoir is extremely important for drinking water. The reason is due to the characteristics of each observation method mentioned above. The reservoirs in mountainous areas are smaller than those in the ocean, and satellites such as MODIS, which frequently take images, have insufficient resolution. Alternatively, if a satellite has a certain high spatial resolution, the imaging interval will be weekly, making it impossible to track the dynamics of phytoplankton. To compensate for this drawback, it is preferable to analyze aerial images taken from UAVs and collect spatiotemporally high-resolution information, but fluctuations in insolation strongly affect the analysis results. In addition, because reservoirs are generally narrow and laterally long, there is a spatial distribution of turbidity depending on the distance from the river inflow and the season, which affects the color tone. This tends to cause errors in determining the Chl-a concentration. In addition, in such a narrow and closed water body that has some inflow river, to determine the source of nutrients, it is necessary to be able to observe the spatial distribution of low-ranging chlorophyll in the pre-occurrence stage of HAB bloom.

This research aims to conduct observations using photographs taken from a UAV and to be able to conduct spatially high-resolution analysis freely without restrictions due to photographing date or cloud cover. After calibrating the reflectance of the images based on insolation, we investigated a method for estimating Chl-a concentration using machine learning based on the calibrated reflectance values and turbidity. Not only was the presence or absence of the bloom of HAB identified, but the lower concentrations of Chlorophyll were also assessed. The information on the distribution is expected to support reservoir management, such as identifying sources of nutrients.

2. Materials and Methods

2.1. Study Site and Photography from UAV

The study site is the Hitotsuse reservoir (Total storage capacity: 264,315,000m^3, Catchment area: 445 km^2), upstream of the Hitotsuse River in Miyazaki Prefecture, Japan (Figure 1). UAV aerial photography surveys were conducted 10 times from September 25, 2020, to October 23, 2022. The UAV aerial photography area is indicated by the red hatched area in Figure 1, and the locations where the water quality surveys were conducted are indicated by the yellow reverse triangles (A to E). The water sampling and the vertical profiling of water quality were conducted from the bridges spanning these points.

The bloom of Microcystis has been frequently confirmed in the reservoir in the past. Among the above observations, the bloom was visually confirmed at the survey on July 16, 2022, while it was not found in the other observations.

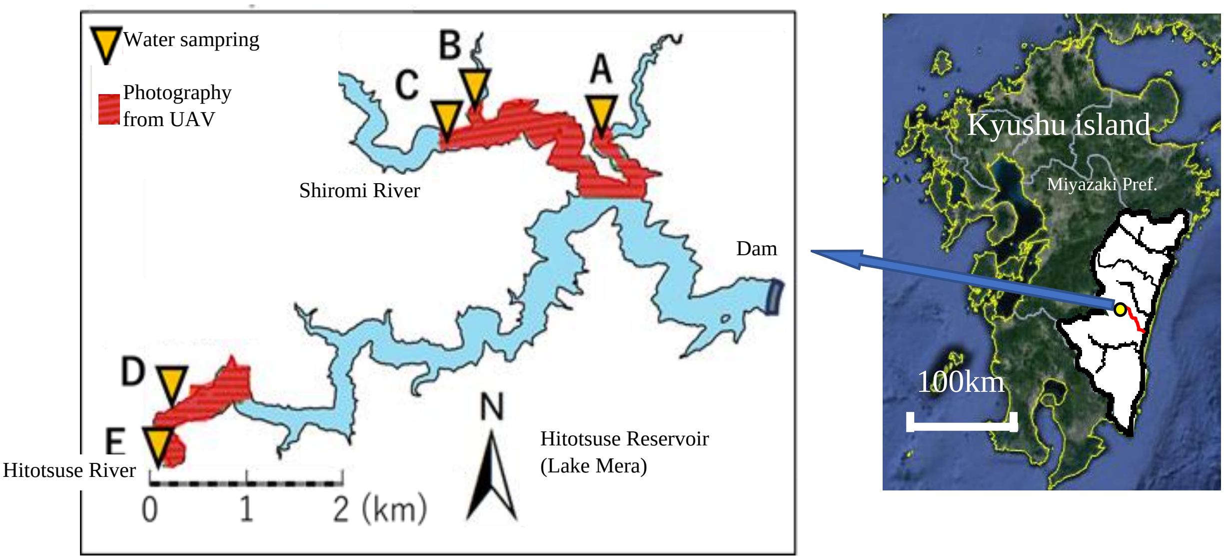

Figure 1. Location of Hitotsuse Reservoir (Lake Mera).

Phantom 4 pro-V2.0 manufactured by DJI was used to take aerial photographs near the inlet of Shiromi River and Hitotsuse River, shown as the red-colored area in Figure 1. Figure 2 shows the shooting setting using the UAV-equipped digital camera. The images were taken at a flight altitude of 140 m, with an overlap rate of 90% on the route and 50-60% between the routes respectively. The number of pixels (pix) of the digital camera is 5472 x 3648 pixels. The size of the image sensor that converts red, blue, and green visible lights into electrical signals is 13.2 x 8.8 mm. The lens focal length is 9.0 mm.

During a flight for 15 to 30 minutes, the brightness of the images fluctuated depending on the change of cloud cover and the solar angle varying in the shooting date and time. To calibrate the reflectance in the images, the insolation was measured with the pyranometer (ML-01, Eiko Seiki Co., Ltd.) near each water quality survey point at 5-second intervals in parallel with the flight.

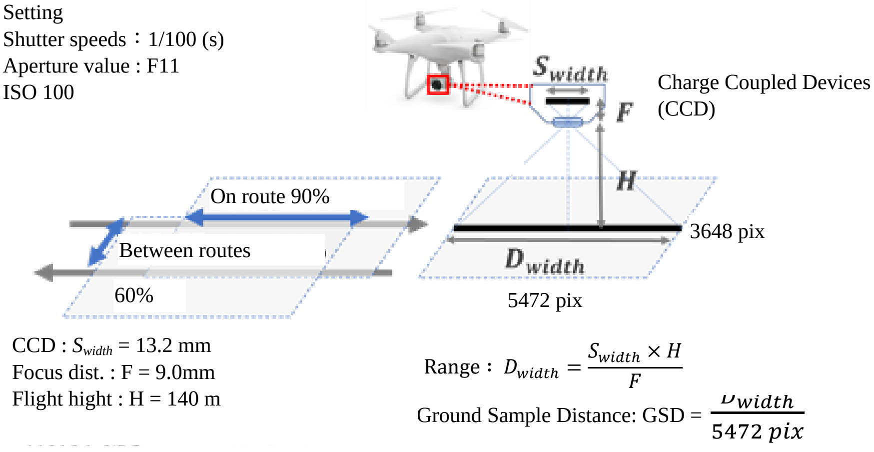

Figure 2. Shooting setting.

2.2. Water Sampling and Vertical Profiling

Fluorescence measurements of chlorophyll and turbidity were carried out by lowering a water quality profiler (AAQ-RINKO, JFE Advantech) from the bridges over the five points in Figure 1. At the same time, the water was sampled from a depth of 2 m and the concentration of Chl-a was measured by the SCOR-UNESCO method. Five liters of the water sample were filtered through a glass fiber filter (Whatman, GF/F, Diameter 47 mm). The residue was extracted with 30 mL of 99.5% ethanol, placed in a 50 mL centrifuge tube protected from light by aluminum foil, and refrigerated for 24 hours. Afterward, it was centrifuged at 3000 rpm for 5 minutes. The supernatant liquid was put into a glass cell (5 cm), and the absorbance was measured at wavelengths of 750, 663, 645, and 630 nm. The concentration of Chl-a was calculated as the below equation.

$$\mathrm{Chl-a} = 11.64 \times \mathrm{e663} - 2.16 \times \mathrm{e645} + 0.10 \times \mathrm{e630} \tag{1}$$

Where e663, e645, and e630 are the absorbance for each wavelength. Since the Chl-a concentration obtained by the fluorescence measurement with the profiler is the Uranine reference value, the calibration formula between the Chl-a concentration of the water sample measured by the SCOR-UNESCO method and the Uranine reference value measured at the same depth by the profiler was identified. The value measured with the profiler was converted to the Chl-a concentration with this calibration formula. Figure 3 is the calibration formula between the Chl-a concentration by the SCOR-UNESCO method and the uranium reference values by the profiler. There is a good correlation between the two values except for the day when the bloom of Microcystis occurred (July 16, 2022).

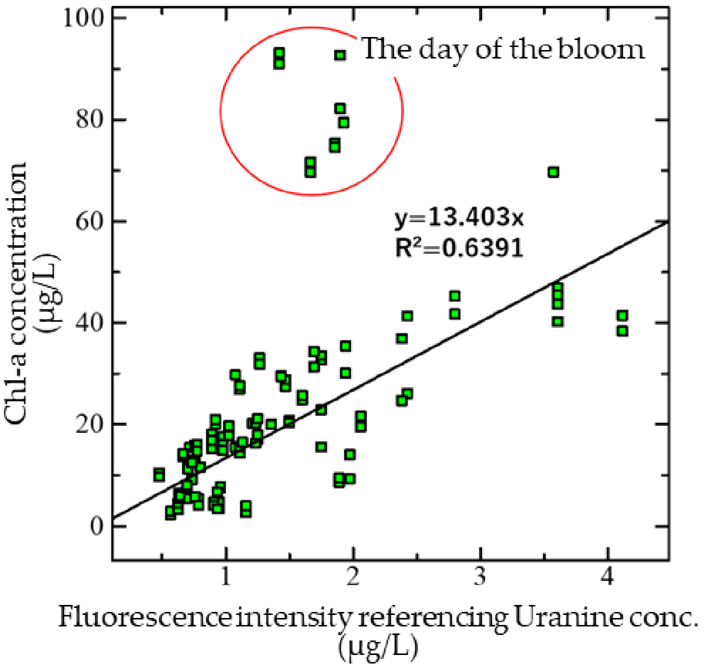

Figure 3. Correlation between Fluorescence intensity referencing Uranine concentration and the concentration of Chl-a by COR-UNESCO method.

2.3. Distribution of Turbidity

In the later chapters, the observed concentration of Chl-a at a depth of 1 m will be estimated by the machine learning regression with the calibrated RGB reflectance values and other additional explanatory parameters. Hereafter, turbidity was incorporated as the explanatory parameter because it was considered to be a factor other than Chl-a concentration that influences the hue of the lake water surface. Figure 4 shows the turbidity (marks) and water temperature (bars) measured at a depth of 1 m on each observation day. There are cases in which the turbidity differs significantly between the Shiromi River (points A, B, and C) and the Hitotsuse River (points D and E) even on the same day. The spatial distribution of the turbidity occurs due to the differences in the turbidity loaded from each inflowing river and the behavior of turbid water after the inflows.

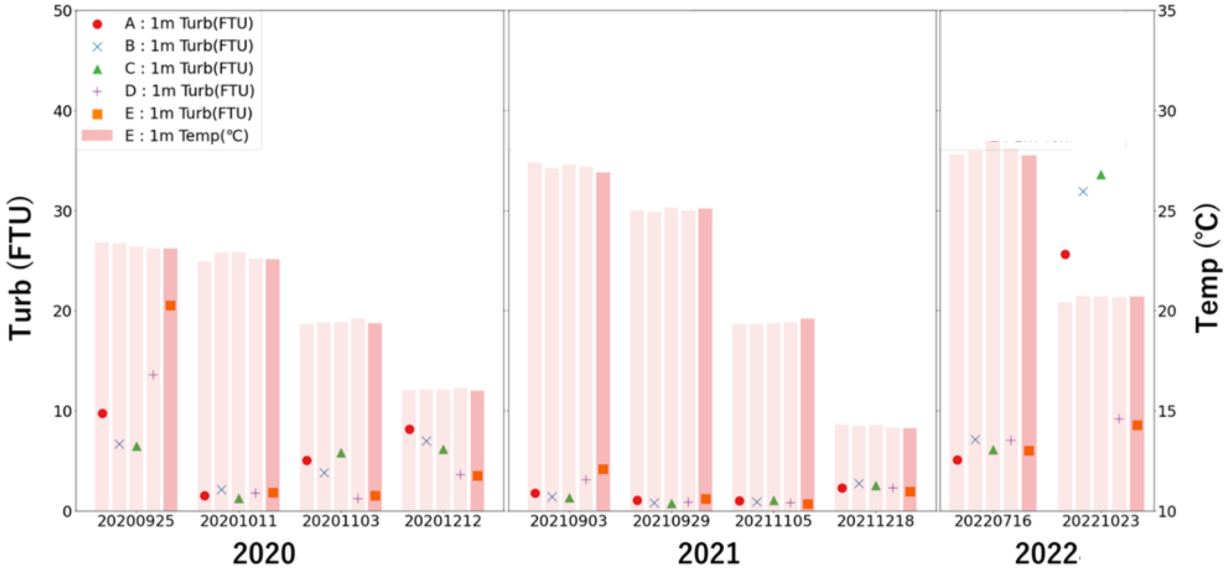

Figure 4. Water temperature (Bars) and Turbidity (Marks) of the water at the 1m depth of the profiling points.

It is necessary to measure the turbidity of the entire lake water surface directly for taking it into account of the factor of the hue, which is putting the cart before the horse. In this study, the spatial distribution of turbidity was estimated using the Water Turbidity Index (WTI) [44]. The WTI was calculated using the red and near-infrared bands of the Sentinel-2 satellite (spatial resolution: 10 m) and Landsat satellite data (30 m). The dates of these satellite images do not match those of the UAV aerial photography. However, it has been reported that the Hitotsuse River basin has a geological structure that is prone to prolonged turbidity [45,46], and its spatial distribution changes are gentler than those of Chl-a. If the temporal gap between the two images is less than one week, and there is no flooding during that time, it is considered that the satellite images are almost synchronized with the UAV aerial images. The WTI was calculated using Equation (3) to (8) shown below.

First, subtract the minimum value (Ref_{min}) from the maximum value (Ref_{max}) of the red band (RED) and near-infrared (NIR) reflectance.

$$b_{RED} = (Ref_{max} - Ref_{min})_{\mathrm{RED}} \tag{3}$$

$$b_{NIR} = (Ref_{max} - Ref_{min})_{\mathrm{NIR}} \tag{4}$$

In each band, b_{RED} and b_{NIR} are normalized. The normalized coefficients A_{RED} and A_{NIR} are obtained, divided by the normalized coefficient B. Finally, WTI is calculated by multiplying the reflectance of the red band and near-infrared image by the normalized coefficients.

$$B = \sqrt{b_{RED}^2 + b_{NIR}^2} \tag{5}$$

$$A_{RED} = \frac{b_{RED}}{B} \tag{6}$$

$$A_{NiR} = \frac{b_{NIR}}{B} \tag{7}$$

$$WTI_i = A_{RED}Ref_{RED} + A_{NIR}Ref_{NIR} \tag{8}$$

2.4. The Flow Chart of Overall Processes for Machine Learning

The Gradient Boosting was applied for the classification of the present of bloom and the regression of the Chl-a concentration at 1m depth, with the calibrated RGB reflectance, temperature, and WTI. In addition, the weak classifier used in this is a decision tree. Decision trees are well known for their use in Classification, but they can also be used in Regression. Gradient Boosting that utilizes this decision tree is called Gradient Boosting Decision Tree (GBDT), and a faster version of that

doi:10.20944/preprints202404.0428.v1

algorithm is called eXtreme Gradient Boosting (XGBoost) [47]. In this research, we will apply XGBoost to improve the accuracy of classifying the presence or absence of Microcystis and predicting the concentration of Chl-a in the stage before that bloom by the regression.

Figure 5 shows the flow of machine learning. The RGB reflectance calibrated by insolation, WTI values obtained from satellite images, and water temperature were used as input parameters. Since the water temperature was uniform over the entire lake surface on each survey day, the observed value at point E was used as the representative value. First, we use XGBoost and these parameters to try to determine the presence or absence of the bloom. This is because when a bloom occurred, the fluorescence characteristics of chlorophyll were different from those under normal conditions, as shown in Figure 3, so it was thought that the reflectance of the visible lights on the water surface would show a unique state. The purpose of this research is not only to determine the presence or absence of blooms but also to investigate the causes of their occurrence through normal-time observations that show the distribution of Chl-a concentrations at lower concentrations. Therefore, we attempted to reproduce the Chl-a concentration using the XGBoost regression model for the images that were determined to be free of the bloom.

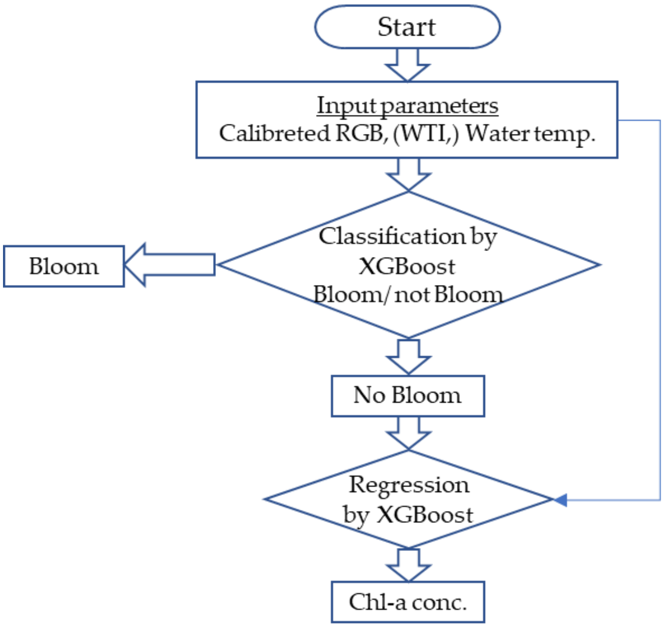

Figure 5. Flow chart of the classification of present/absence of bloom and the regression of the concentration of Chl-a.

In this study, the number of field surveys was limited. To obtain a sufficient number of datasets for the training and assessment of the machine learning, we extracted 25 pixels from an area of 3.3 x 3.3 m around the water quality meter point of the image at each point and used their RGB values as the parameters. The mesh size of WTI is 10 × 10 m. On the other hand, as shown in Figure 6, it has been confirmed that there is some variation in the RGB dataset. Based on this, we increased the data used for machine learning by preparing 25 types of RGB datasets for one WTI value, water temperature, and Chl-a concentration. Therefore, 1000 sets (25 x 5 points x 8 observation days) were used for classification, and 875 sets(25 x 5 locations x 7 days) were used for regression.

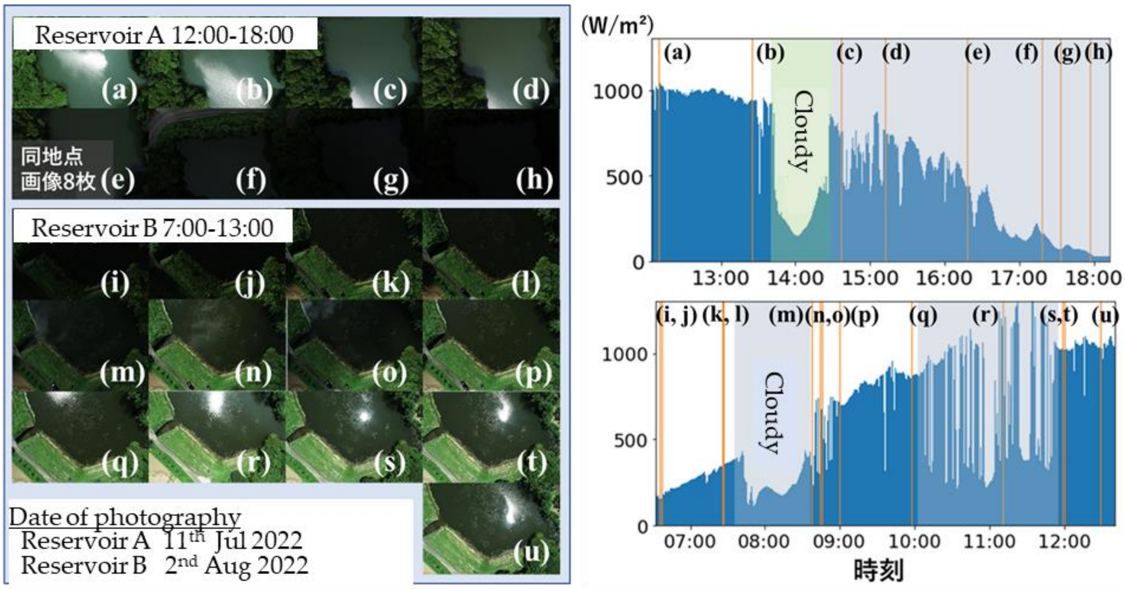

Figure 6. Photographs of the same area at different times within the same day and fluctuation of irradiance.

The values of calibrated RGB reflectances, WTI, and water temperature were normalized by the Min-Max method scaling the data from 0 to 1 as Eq. (9). Normalization is expected to improve classification and prediction because it reduces the values of data features and equalizes the numerical range.

$$x_n^{'} = \frac{x_n - \min(x_i)}{\max(x_i) - \min(x_i)} \ (i = 1 \cdots N) \qquad (9)$$

Where x_n is the original value of the input parameters, x_n' is the normalized value and N is the number of the dataset.

2.5. Rectification of the Coordinates (Georeferencing)

The images obtained by UAV aerial photography covered an area of 205 x 137 m in one image, and the resolution was 3.75 cm/pix. When creating the distribution map of Chl-a, the RGB reflectance values given for each pixel are used as an explanatory variable in the machine learning model, and the surface Chl-a concentration is regressed. The distribution of the estimated values can be expressed by projecting them onto the lat/long global coordinates.

Cheng et.al [38] aligned the raw images by matching feature points between images and projected them onto the global coordinates, by the same method as the intermediate process of Structure from Motion (SfM). With this reprojection method, the positional information is usually corrected by installing Ground Control Points (GCP), making it possible to create highly accurate composite images of several centimeters. However, it is difficult to find some feature points on the image that capture only the almost uniform water surface in the center of the narrow lake in this study. The trial of Cheng et.al shows good alignment for the water surface that has the floating houses while the bare water surface required the rectification software of Image Composite Editor [48] which distorts the images.

Therefore, first, each image was analyzed one by one without performing the image aligning method in SfM analysis. The images were divided into small meshes of 5m x 5m. In the areas where the water surface was rippled, diffuse reflections caused the pixels to exceed the maximum detection limit of the CCD (showing a white spot in an image). These white spots serve as blanks, and the average value of the pixels within each mesh without the brank is used as reflectance values of each mesh. The Chl-a concentration of each mesh is estimated with the method mentioned above from the averaged values of the reflectance. The global coordinates were given to each mesh based on the shooting coordination and the Yaw angle, recorded in each image, as equation 10 [49].

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \tag{10}$$

Where x and y are the local coordinates of each mesh on the image, X_1 and Y_1 are the shooting global coordinates of the image, θ is the Yaw angle.

3. Results and Discussion

3.1. Calibration of the Reflectance

The RGB values of each pixel on the photograph change depending on the insolation. The water surfaces at the two small reservoirs near our university campus were photographed at different times during the same observation day when the water surface conditions were considered constant. While the water quality of the reservoir does not change significantly through the day, the insolation conditions differ depending on the time of day when the image was taken, so that the changes in the reflectance due to the insolation could be evaluated. The photographing settings were the same as shown in Figure 2. The aerial photographs of the water surfaces were taken at each reservoir every 1 to 2 hours.

The observations of insolation intensity were recorded using the same apparatus as the field survey at Hitotsuse reservoir. To obtain data on a wide range of insolation intensity, we conducted the survey in summer, when the solar altitude changes significantly depending on the time of day. Figure 6 shows images of the reservoirs taken on July 11 and August 2, 2020, and the insolation intensity observed at each reservoir. The reflectance of the image changes as the insolation changes. When direct insolation is blocked by physical factors such as clouds, the insolation intensity is greatly attenuated. The relationship between lake surface reflectance and insolation intensity was formulated and the reflectance was calibrated by the insolation.

A high correlation was found between the insolation and the lake surface reflectance. From these correlations, the correction factor (α) was determined as Equation (11) to convert the observed value to the reflectance when the insolation was 600 (W/m²).

$$\alpha = \frac{DN_{(R,G,B)600}}{DN_{(R,G,B)i}} \tag{11}$$

Figure 7 shows the relationship between insolation and calibration coefficient (α). The results for the two reservoirs with different water quality and color were shown in the one plot but it could be uniformly corrected using the above coefficients.

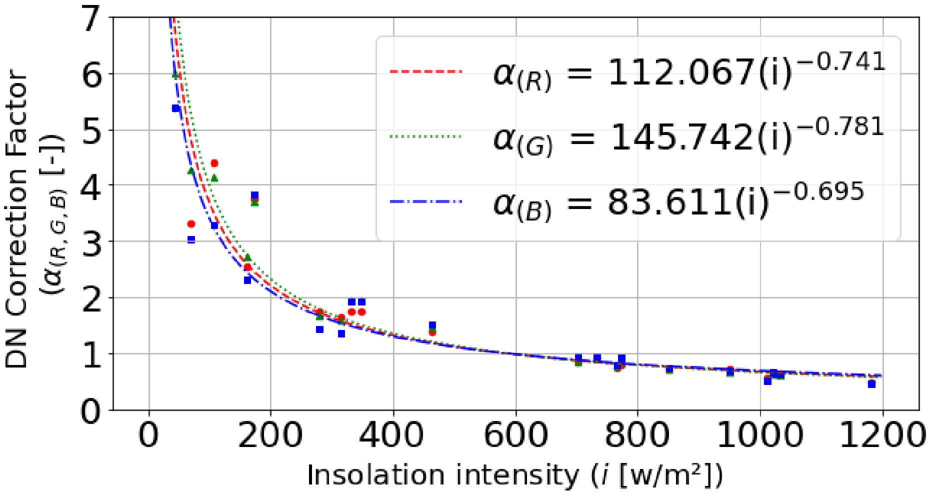

Figure 7. Relationship between Insolation intensity and calibration coefficients.

In addition, to verify the applicability of the calibration coefficient mentioned above method, the photography was conducted in the same location of Hitotsuse reservoir under different cloud conditions during the day. For the images used for verification, we selected two images taken at different times and with significantly different insolation intensities at the same location on each shooting day. Using the calibration coefficient, the time of the same day at Hitotsuse reservoir is calculated. Table 1 shows the results of the correction. In all cases, the difference in the values after the correction decreased.

Table 1. Comparison of the reflectance values before and after calibration for two images with the rapid changes in Insolation.

		Insolation	Original DN			Equivalent reflectance under insolation of 600 W/m^2		
Date	Time	(W/m^2)	b	g	r	b	g	r
11 Oct 2020	11:31	352	37	39	11	53	58	16
	11:37	760	79	95	31	66	78	25
3 Sep 2021	13:39	889	95	121	58	71	88	42
	14:57	166	31	35	17	74	94	43
29 Sep 2021	13:35	905	61	65	19	45	46	14
	14:40	184	18	16	6	40	40	14
16 Jul 2022	11:02	846	96	139	72	74	105	55
	11:54	449	77	102	53	92	126	64

Figure 8 shows the image combining some shooting for a certain range for the comparison between before and after the calibration. Due to the fluctuations in insolation during the automatic navigation flight with a lot of shoots, the northern part was photographed during the cloudy period, surrounded by the dotted red line in the figure, and the image before the calibration had low brightness. The proposed calibration results of the combined image look continuous without any discomfort.

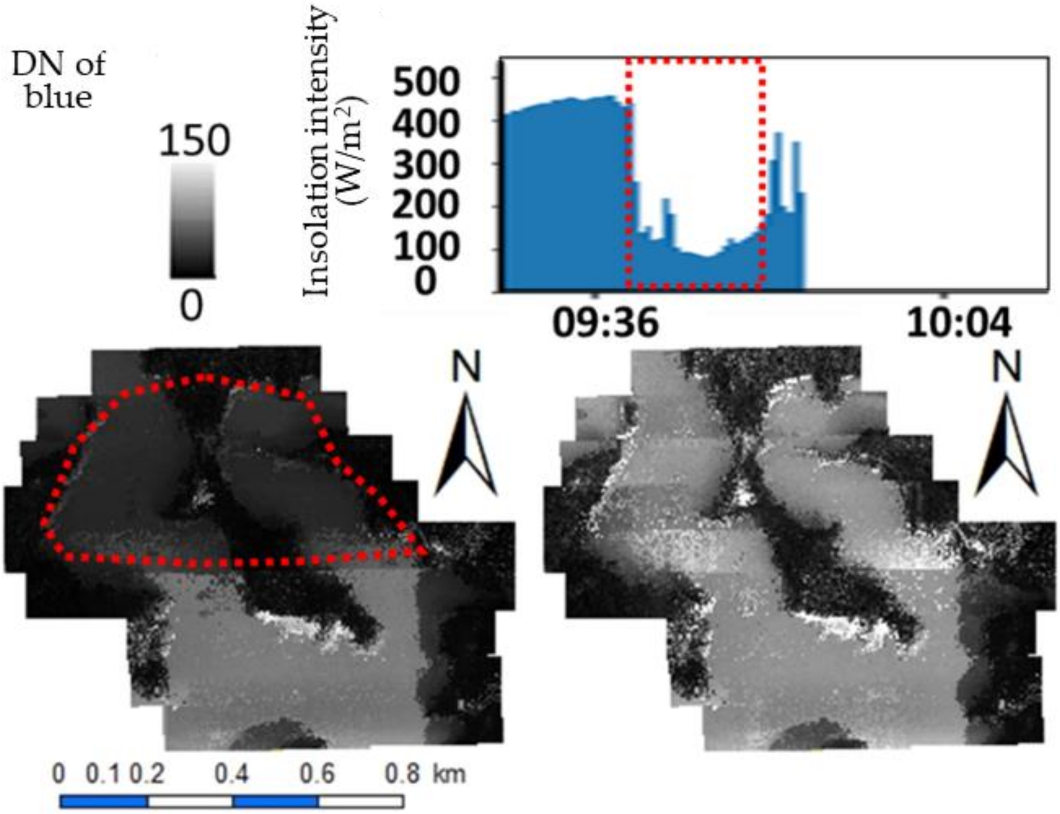

Figure 8. Reflectance calibration for the images taken in a flight based on insolation.

3.2. Turbidity Estimation from Satellite Images

Satellite images have lower spatial resolution than UAV aerial images. In addition, there is little flexibility in the shooting date, so it does not necessarily coincide with the UAV aerial shooting date. However, since the flow in the lake is slow during fine days, the turbidity changes gentler than in Chl-a. Considering this difference in the speed of time progression, the WTI distribution obtained from the satellite images with a temporal gap of less than one week from the date of the aerial photography and no precipitation during that time can be used as a substitute for the turbidity distribution on the date of the aerial photography.

Figure 9 shows a correlation between WTI and measured turbidity at a depth of 1 m. Some points do not show good correspondence. These are the results for the day when the bloom happened, and the point near the river inlet one month after the large-scale flood. As mentioned previously, when the bloom occurred, the dominant species near the surface layer were different from those in normal times, and it is assumed that the reflection characteristics were different from the normal condition, which also affects the turbidity evaluation results. In addition, near the river inlet after the flooding, dense turbid water was distributed in the deep layer, while the low-density clear water from the inflow river flowed into the surface layer. As a result, it is supposed that the color of the deep turbid water was transmitted through the shallow layer and appeared on the lake surface. Regarding these points, the relationship between turbidity at a depth of 14 m and WTI obtained from profiler observation results is shown in the figure. These are like the correlations of the other results.

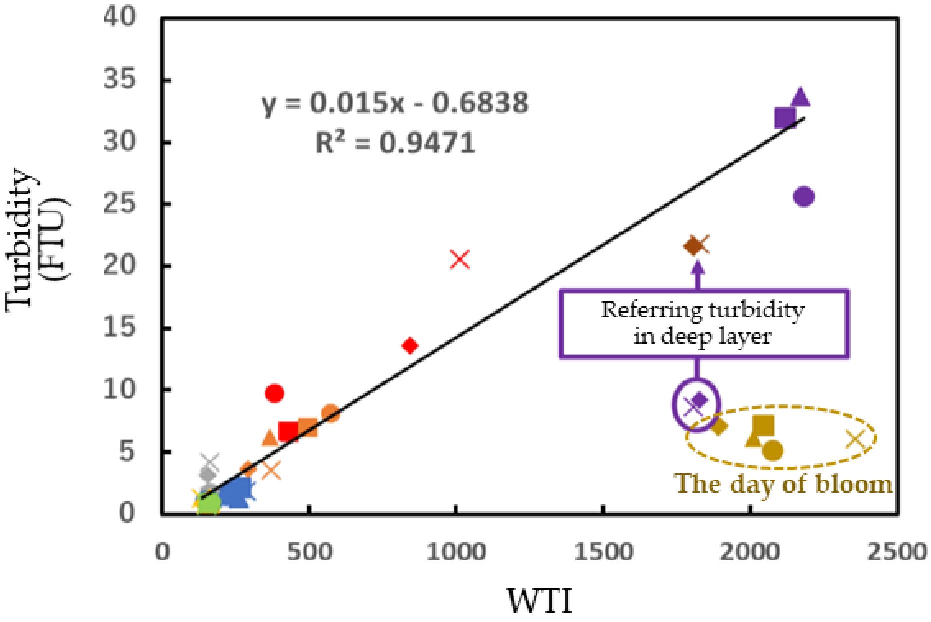

Figure 9. Correlation between turbidity and WTI.

3.3. Classification of Presence or Absence of Bloom and Regression of Chl-a Concentration

The datasets without the bloom (7 days) and with the bloom (1 day: July 16, 2022) were used for training and validation tests of the classification of the presence/absence of bloom. The bloom was visually confirmed on the aerial images as shown in Figure 10. The test data were prepared from the datasets of 2 sampling points (50 sets) out of the 5 sampling points on the day of the bloom, and the other 50 sets were randomly selected from the data when the bloom did not occur. The other data (900 sets) were used as training data.

Figure 10. Aerial photograph when the bloom.

Table 2 is the confusion matrix of the classification results. The total accuracy was 82%, the recall rate was 64%, and the precision rate was 100%. The false negative rate is high, but this is likely due to the use of WTI as a parameter. WTI can increase not only by turbidity but also by bloom, as mentioned above. For this reason, a lot of "false negatives" might be found because it was not possible to classify cases when the bloom occurred and cases when the bloom did not occur but the turbidity was high.

Table 2. Confusion matrix of the classification of Presence/Absence of Bloom by XGBoost using the calibrated RGB, WTI, and Water temperature as input.

		Predicted	
		Presence	Absence
True class	Presence	32	18
	Absence	0	50

Therefore, WTI was omitted from the explanatory parameters for the classification whether the presence or absence of a bloom. The result using only the calibrated RGB reflectance and water temperature as the explanatory parameters is shown in Table 3. By omitting WTI, the presence/absence of bloom could be classified with high accuracy.

Table 3. Confusion matrix of the classification of Presence/Absence of Bloom by XGBoost using the calibrated RGB and Water temperature as input.

		Predicted	
		Presence	Absence
True class	Presence	50	0
	Absence	0	50

For the detection of the source of the nutrients contributing blooms but before blooms, the distribution of the lower concentration of Chl-a is important information. The datasets of the days when the bloom did not occur were used for the regression of the Chl-a concentration. The explanatory parameters were the calibrated RGB, WTI, and temperature. All these parameters were normalized before the training and the validation. In machine learning, success or failure depends on how the data is divided into test data and training data. Especially when the total amount of data is limited as in this case, it affects the accuracy of evaluation results. We randomly selected two points (50 sets) from points a to e on each day as test data, and three points (75 sets) as training data. The coefficient of determination was 0.84, and reliable estimation accuracy was obtained in both cases. Figure 11 shows the correlation between the actual Chl-a concentrations and the estimated concentrations. Estimations can be made over a wide range from low to high concentrations, suggesting that it is possible to evaluate the Chl-a concentrations in the early stages of algae with lower concentrations from aerial images.

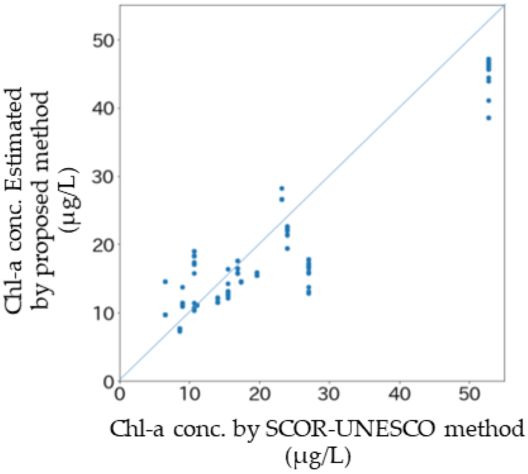

Figure 11. Comparison between True and Estimated Chl-a concentration.

3.5. Distribution of the Concentration of Chl-a Estimated by the Proposed Method

Figure 12 shows the image before and after resizing. In the figure, GSD 10 m and 30 m images, which correspond to the resolution of Sentinel and Landsat satellites, are also shown for comparison. Looking at Figure 12, it was thought that the Chl-a distribution could be sufficiently expressed with a GSD 5 m image. In addition, the hovering accuracy range of the UAV aircraft is ±0.5 m in the horizontal direction and ±1.5 m in the vertical direction25), and it was thought that the error would become even larger due to the influence of the wind in the upper atmosphere. Therefore, even if we estimate the Chl-a concentration of each pixel by obtaining luminance values at a resolution of 3.75 cm/pix, it is difficult to accurately project them onto a map. Therefore, in this study, we decided that a resolution of 5 m GSD was appropriate, considering the computational load also, and resized all images to be analyzed to a 5 m mesh using bilinear interpolation.

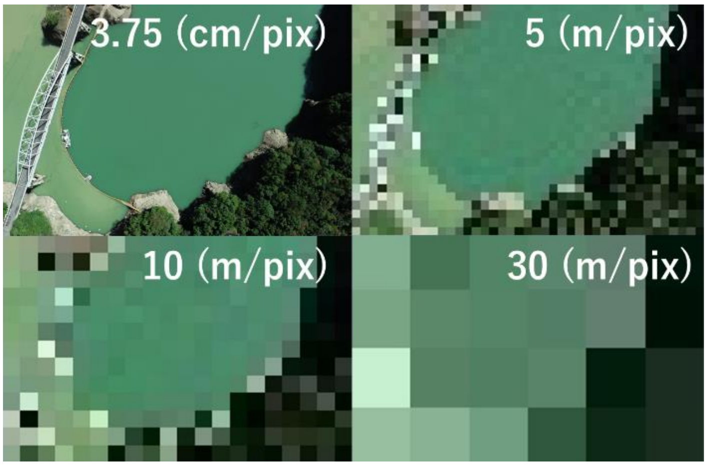

Figure 12. Images resized with different resolutions.

Figure 13 shows the distribution of the estimated Chl-a concentration on each observation date. No large spatial deviations were observed in the series of images from the same day, except in the case of 23 Oct. 2022, when the largest flood happened one month before. It is known that when such a large-scale outflow occurs in this reservoir, the turbid state differs depending on the inflowing river due to differences in the geological characteristics of the catchment area[45]. In this study, WTI is included as an explanatory variable for the machine learning. Therefore, the reason why the Chl-a concentration around the Shiromi River inflow area is evaluated to be high is due to the high WTI, and it can be said that the actual Chl-a concentration is not being reproduced. However, in this study, the training for the machine learning includes actual observation results with the high WTI, and the test data in that area was reproduced well, as shown in Figure 11.

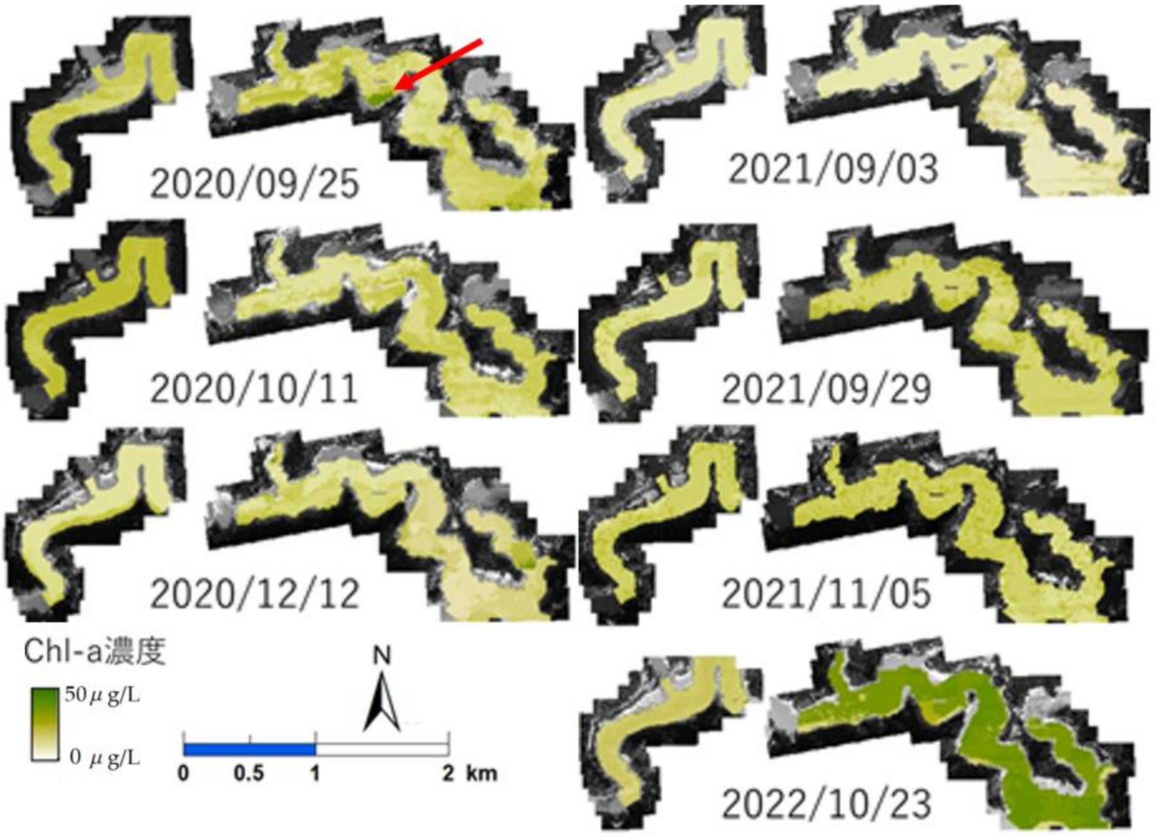

Figure 13. Distribution of the estimated Chl-a concentration on each observation date.

On the other hand, it was confirmed that the concentration of Chl-a differed depending on the survey date. During normal times when there is little rainfall, such as in September and October 2020, the Chl-a concentration is slightly higher at the Hitotsuse River inlet. It is known that the Hitotsuse River basin has a relatively larger population and a slightly higher nutrient load than the Shiromi River basin. The results might reflect this difference in nutrient supply.

In the image of 25 Sep 2020, the small area with high concentration, indicated with the red arrow was found. This irregular distribution is due to the shade which is created by the neighboring mountain on the south side of the lake. Currently, there is no solution for such error induced by local shades, while the influence of the cloud cover change generally covers the survey area on the prediction of Chl-a concentration.

5. Conclusions

In this study, we tried to estimate the distribution of chlorophyll-a by analyzing the aerial images taken from UAVs, which have a high spatial resolution and a high flexibility of shooting schedule. The bloom of harmful algae is the most important event to control in reservoir management. It is necessary to detect environmental factors such as oversupply of nutrients at an early stage and take countermeasures. However, in closed water bodies such as reservoirs with limited surface area, the lake water is often completely eutrophic during the bloom stage. Furthermore, since floating harmful algae are blown by wind, it is considered difficult to identify the source of nutrients from their spatial distribution.

It was expected that the hue of the lake surface was determined by the turbidity discharged after flooding, in addition to the chlorophyll concentration. We calculated WTI from satellite images, and used it as one of the explanatory variables, together with the RGB values of aerial photographs to regression the chlorophyll concentration. The proposed method obtained good reproducibility. Turbid water behaves more slowly than chlorophyll, so even if there is a slight temporal gap between the date of aerial photography by UAV and the day of shooting from satellites, the turbidity distribution does not change significantly. It will be possible to synchronize the evaluation of turbidity and spatial distribution of chlorophyll by equipping a UAV with a near-infrared camera [36], instead of using satellite images to evaluate WTI. Since, likely, appropriate satellite images cannot be obtained due to weather conditions, it is desirable to obtain near-infrared reflectance using a UAV in the future to determine WTI. In this study, turbidity was recognized as a factor that could affect the hue of the lake surface after a considerable number of observations had been completed, so the monitoring with the UAV mounting a near-infrared sensor has not been done.

The chlorophyll-a concentrations evaluated by the proposed method were plotted on maps, which visually represented the temporal fluctuations of chlorophyll-a, and it was also possible to see that the chlorophyll-a concentrations differed depending on the inflowing rivers, even on the same day. However, the reflectance characteristics of chlorophyll and turbid water depend on the type of plankton that dominates and the geological characteristics of the basin. So, to regress chlorophyll-a concentration generally for any reservoir using a machine learning model, it is necessary to prepare learning data for each reservoir and build a model or to build a single model using images obtained from multiple reservoirs as training data. The results presented in this study were only based on the observations from the reservoir. For further study, it is necessary to improve the accuracy of the model by photographing and observing water quality in a lot of reservoirs.

Author Contributions: Conceptualization, M.I.; methodology, M.I., Y.M. and M.Y.; software, Y.M. and M.Y.Y.M.; validation, M.I., Y.M. and M.Y.; formal analysis, M.I., Y.M. and M.Y.; investigation, M.I., Y.M. and M.Y.; resources, M.I.; data curation, Y.M. and M.Y.; writing—original draft preparation, M.I.; writing—review and editing, M.I.; visualization, Y.M. and M.Y.; supervision, M.I.; project administration, M.I.; funding acquisition, M.I.

Funding: This research was funded by Miyazaki Construction Technology Promotion Organization

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ellison, M. E.; Brett, M. T. Particulate phosphorus bioavailability as a function of stream flow and land cover. *Water Research*. **2006**, 40, 6, 1258. https://doi.org/10.1016/j.watres.2006.01.016
2. Azhikodan, G.; Yokoyama, K. Spatio-temporal variability of phytoplankton (Chlorophyll-a) in relation to salinity, suspended sediment concentration, and light intensity in a macrotidal estuary. *Continental Shelf Res.* **2016**, 126, pp. 15–26. doi:10.1016/j.csr.2016.07.006
3. Allison, A.O.; Randy, A.D.; Michae, L.D. The upside-down river: Reservoirs, algal blooms, and tributaries affect temporal and spatial patterns in nitrogen and phosphorus in the Klamath River, USA. *J. Hydrol.* **2014**, 519, pp. 164–176.
4. Dabrowski, T.; Lyons, K.; Nolan, G.; Berry, A.; Cusack C.; Silke, J. Harmful algal bloom forecast system for SW Ireland. Part I: Description and validation of an operational forecasting model. *Harmful Algae.* **2016**, 53, pp. 64-76. https://doi.org/10.1016/j.hal.2015.11.015
5. Welch, E.B. *Ecological Effects of Wastewater: Applied Limnology and Pollutant Effects.* Second ed. CRC Press, London, UK, 1992. https://doi.org/10.4324/9780203038499
6. Carmichael, W.W. The toxins of cyanobacteria. *Sci. Am.* **1994**, 270, pp. 78–84, https://doi.org/10.1038/scientificamerican0194-78

7. Downing, J.A.; Watson, S.B.; McCauley, E. Predicting cyanobacteria dominance in lakes. *Can. J. of Fish. Aquat. Sci.* **2001**, 58, pp. 1905–1908.
8. Jilbert, T.; Couture, R.M.; Huser, B.J. Preface: Restoration of eutrophic lakes: current practices and future challenges. *Hydrobiologia.* **2020,** 847, pp. 4343–4357. https://doi.org/10.1007/s10750-020-04457-x
9. Anderson, D. M.; Fensin, E.; Gobler, C. J.; Hoeglund, A. E.; Hubbard, K. A.; Kulis, D. M.; Landsberg, J. H.; Lefebvre, K. A.; Provoost, P.; Richlen, M.; Smith, J. L.; Solow, A. R.; Trainer, V. L. Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae.* **2021**, 102, 101975. https://doi.org/10.1016/j.hal.2021.101975
10. Anderson, D.M.; Cembella, A.D.; Hallegraeff, G.M. Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management. *Annu. Rev. Mar. Sci.* **2012**, 4, pp. 143-176. https://doi.org/10.1146/annurev-marine-120308-081121
11. Díaz, P.A.; Ruiz-Villarreal, M.; Pazos, Y.; Moita, T.; Reguera, B. Climate variability and Dinophysis acuta blooms in an upwelling system. *Harmful Algae.* **2016**, 53, pp. 145-159. https://doi.org/10.1016/j.hal.2015.11.007
12. Houliez, E.; Lizon, F.; Lefebvre, S.; Artigas, L.F; Schmitt, F. G. Phytoplankton photosynthetic activity dynamics in a temperate macrotidal ecosystem (the Strait of Dover, eastern English Channel): Time scales of variability and environmental control. *J. Mar. Sys.* **2015**, 147, pp. 61 – 75. https://doi.org/10.1016/j.jmarsys.2014.05.001
13. Azevedo, I. C.; Duarte, P. M.; Bordalo, A. A.; Understanding spatial and temporal dynamics of key environmental characteristics in a mesotidal Atlantic estuary (Douro, NW Portugal). *Estuar. Coast. Shelf Sci.* 2008, 76, 3, pp. 620 – 633. https://doi.org/10.1016/j.ecss.2007.07.034
14. Mori, M.; Gonzalez Flores, R.; Suzuki, Y.; Nukazawa, K.; Hiraoka, T.; Nonaka, H. Prediction of Microcystis Occurrences and Analysis Using Machine Learning in High-Dimension, Low-Sample-Size and Imbalanced Water Quality Data. *Harmful Algae.* **2022**. 117, 102273. DOI: 10.1016/j.hal.2022.102273
15. Shan, K.; Wang, X.; Yang, H.; Zhou, B.; Song, L.; Shang, M. Use statistical machine learning to detect nutrient thresholds in Microcystis blooms and microcystin management, *Harmful Algae*, **2020**, 94, 101807. https://doi.org/10.1016/j.hal.2020.101807
16. Wang, L.; Zhang, T.; Wang, X.; Jin, X.; Xu, J.; Yu, J.; Zhang, H.; Zhao, Z. An approach of improved Multivariate Timing-Random Deep Belief Net modeling for algal bloom prediction. *Biosyst. Eng.* **2019**, 177, pp. 130-138. https://doi.org/10.1016/j.biosystemseng.2018.09.005
17. Jiang, Y.J.; He, W.; Liu, W.X.; Qin, N. The seasonal and spatial variations of phytoplankton community and their correlation with environmental factors in a large eutrophic Chinese lake (Lake Chaohu). *Ecol. Indic.* **2014**, 40, pp. 58–67. https://doi.org/10.1016/j.ecolind.2014.01.006
18. Dale, B.; Murphy, M. A retrospective appraisal of the importance of high-resolution sampling for harmful algal blooms: Lessons from long-term phytoplankton monitoring at Sherkin Island, S.W. Ireland. *Harmful Algae.* **2014**, 40, pp. 23–33. https://doi.org/10.1016/j.hal.2014.09.007
19. Nwe, L. W.; Yokoyama, K.; Azhikodan, G.; Phytoplankton habitats and size distribution during a neap-spring transition in the highly turbid macrotidal Chikugo River estuary. *Sci. Total Environ.* **2022**, 850, 157810. https://doi.org/10.1016/j.scitotenv.2022.157810
20. Veerapaga, N.; Shintani, T.; Azhikodan, G.; Yokoyama, K. Study on Salinity Intrusion and Mixing Types in a Conceptual Estuary Using 3-D Hydrodynamic Simulation: Effects of Length, Width, Depth, and Bathymetry. In: Nguyen, K., Guillou, S., Gourbesville, P., Thiébot, J. (eds) *Estuaries and Coastal Zones in Times of Global Change*. Springer Water. Springer, Singapore, 2020. https://doi.org/10.1007/978-981-15-2081-5_2
21. Chaffin, J.D., Kane, D.D.; Stanislawczyk, K.; Parker, E.M. Accuracy of data buoys for measurement of cyanobacteria, chlorophyll, and turbidity in a large lake (Lake Erie, North America): implications for estimation of cyanobacterial bloom parameters from water quality sonde measurements. *Environ. Sci. Pollut. Res. Int.* **2018** Sep;25, 25, 25175-25189. doi: 10.1007/s11356-018-2612-z. Epub 2018 Jun 25. PMID: 29943249.
22. Nakayama, K.; Komai, K.; Amano, M., Horii, S.; Somiya, Y.; Kumamoto, E.; Oyama Y. Ideal water temperature environment for giant Marimo (Aegagropila linnaei) in Lake Akan, Japan. *Sci. Rep.* **2023**, 13, 16834. https://doi.org/10.1038/s41598-023-43792-6
23. Le, H.N.; Shintani, T.; Nakayama, K. A Detailed Analysis on Hydrodynamic Response of a Highly Stratified Lake to Spatio-Temporally Varying Wind Field. *Water.* **2023**, 15, 565. https://doi.org/10.3390/w15030565
24. Xu, X.; Ishikawa,T; Nakamura,T. Three-dimensional modeling of hydrodynamics and dissolved oxygen transport in tone river estuary. *J. JSCE.* **2013**, 1, 1, pp. 194 – 213. https://doi.org/10.2208/journalofjsce.1.1_194
25. Huang, J.; Zhang, Y.; Huang, Q.; Gao, J. When and where to reduce nutrients for controlling harmful algal blooms in large eutrophic lake Chaohu, China?. *Ecol. Indic.* **2018**, 89, pp. 808 – 817. https://doi.org/10.1016/j.ecolind.2018.01.056
26. Maguire, J.; Cusack, C.; Ruiz-Villarreal, M.; Silke, J.; McElligott, D.; Davidson, K. Applied simulations and integrated modeling for the understanding of toxic and harmful algal blooms (ASIMUTH): Integrated HAB

forecast systems for Europe's Atlantic Arc. *Harmful Algae.* **2016**, 53, pp. 160-166. https://doi.org/10.1016/j.hal.2015.11.006
27. Li, D.; Gao, Z.; Song, D.; Shang, W.; Jiang, X. Characteristics and influence of green tide drift and dissipation in Shandong Rongcheng coastal water based on remote sensing. *Estuar, Coast. Shelf Sci.* **2019**, 227, 106335. https://doi.org/10.1016/j.ecss.2019.106335
28. Hu, L.; Zeng, K.; Hu, C.; He, M.X. On the remote estimation of Ulva prolifera areal coverage and biomass. *Remote Sens. Environ.* 223, pp. 194 – 207, 2019. https://doi.org/10.1016/j.rse.2019.01.014
29. Chen, X.; Shang, S.; Lee, Z.; Qi, L.; Yan, J.; Li, Y. High-frequency observation of floating algae from AHI on Himawari-8. *Remote Sens. Environ.* 2019, 227, pp. 151 – 161. https://doi.org/10.1016/j.rse.2019.03.038
30. Ma, J.; Jin, S.; Li, J.; He, Y.; Shang, W. Spatio-Temporal Variations and Driving Forces of Harmful Algal Blooms in Chaohu Lake: A Multi-Source Remote Sensing Approach. *Remote Sens.* **2021**, 13, 427. https://doi.org/10.3390/rs13030427
31. Zhang, T.; Hu, H.; Ma, X.; Zhang, Y. Long-Term Spatiotemporal Variation and Environmental Driving Forces Analyses of Algal Blooms in Taihu Lake Based on Multi-Source Satellite and Land Observations. *Water* **2020**, 12, 1035. https://doi.org/10.3390/w12041035
32. Liu, M.; Ling, H.; Wu, D.; Su, X.; Cao, Z. Sentinel-2 and Landsat-8 Observations for Harmful Algae Blooms in a Small Eutrophic Lake. *Remote Sens*. **2021**, 13, 4479. https://doi.org/10.3390/rs13214479
33. Soomets, T.; Uudeberg, K.; Jakovels, D.; Brauns, A.; Zagars, M.; Kutser, T. Validation and Comparison of Water Quality Products in Baltic Lakes Using Sentinel-2 MSI and Sentinel-3 OLCI Data. *Sensors* **2020**, 20, 742. https://doi.org/10.3390/s20030742
34. Toth, C., Jóźków G., Remote sensing platforms and sensors: A survey. *ISPRS J. Photogramm. Remote Sens.* **2016**, 115, pp. 22–36. https://doi.org/10.1016/j.isprsjprs.2015.10.004
35. Qu, M.; Anderson, S.; Lyu, P., Malang; Y. Lai; J. Liu; J.; Jiang, B.; Xie, F.; Liu, H. H.T.; Lefebvre, D. D.; Wang, Y. S. Effective aerial monitoring of cyanobacterial harmful algal blooms is dependent on understanding cellular migration. *Harmful Algae*. **2019**, 87, 101620. https://doi.org/10.1016/j.hal.2019.101620
36. Guimarães, T.T.; Veronez, M.R.; Koste, E.C.; Gonzaga, L.; Bordin, F.; Inocencio, L.C.; Larocca, A.P.C.; De Oliveira, M.Z.; Vitti, D.C.; Mauad, F.F. An Alternative Method of Spatial Autocorrelation for Chlorophyll Detection in Water Bodies Using Remote Sensing. *Sustainability* **2017**, 9, 416. https://doi.org/10.3390/su9030416
37. Su, T.-C.; Chou, H.-T. Application of Multispectral Sensors Carried on Unmanned Aerial Vehicle (UAV) to Trophic State Mapping of Small Reservoirs: A Case Study of Tain-Pu Reservoir in Kinmen, Taiwan. *Remote Sens.* **2015**, 7, pp. 10078-10097. https://doi.org/10.3390/rs70810078
38. Cheng, K.H.; Chan, S.N.; Lee, J.H.W.; Remote sensing of coastal algal blooms using unmanned aerial vehicles (UAVs), *Mar. Pollut. Bull.* **2020**, 152, 110889. https://doi.org/10.1016/j.marpolbul.2020.110889
39. Cao, Z.; Duan H.; Feng, L.; Ma, R.; Xue, K. Climate- and human-induced changes in suspended particulate matter over Lake Hongze on short and long timescales. *Remote Sens. Environ.* **2017**, 192, pp. 98–113. https://doi.org/10.1016/j.rse.2017.02.007
40. Hu, C.; Chen, Z.; Clayton, T. D.; Swarzenski, P.; Brock, J. C.; Muller–Karger, F. E. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sens. Environ.* **2004**, 93, 3, pp. 423–441. https://doi.org/10.1016/j.rse.2004.08.007
41. Kishino, M.; Tanaka, A.; Ishizaka, J. Retrieval of Chlorophyll a, suspended solids, and colored dissolved organic matter in Tokyo Bay using ASTER data, *Remote Sens. Environ.* **2005**, 99, 1, pp. 66 – 74, https://doi.org/10.1016/j.rse.2005.05.016
42. Sakuno, Y.; Yajima, H.; Yoshioka, Y.; Sugahara, S.; Abd Elbasit, M.A.M.; Adam, E.; Chirima, J.G. Evaluation of Unified Algorithms for Remote Sensing of Chlorophyll-a and Turbidity in Lake Shinji and Lake Nakaumi of Japan and the Vaal Dam Reservoir of South Africa under Eutrophic and Ultra-Turbid Conditions. *Water* **2018**, 10, 618. https://doi.org/10.3390/w10050618
43. Palmer, S.C.J.; Hunter, P. D.; Lankester, T.; Hubbard, S.; Spyrakos, E.N.; Tyler A.; Présing, M.; Horváth, H.; Lamb, A.; Balzter, H.; Tóth, V. R. Validation of Envisat MERIS algorithms for chlorophyll retrieval in a large, turbid and optically-complex shallow lake. *Remote Sens. Environ*. **2015**, 157, pp. 158 – 169, https://doi.org/10.1016/j.rse.2014.07.024
44. Yamagata, Y.; Wiegand, C.; Akiyama, T.; Shibayama, M. Water turbidity and perpendicular vegetation indices for paddy rice flood damage analyses, *Remote Sens. of Environ.* **1988**, 26, pp. 241-251, https://doi.org/10.1016/0034-4257(88)90080-6.
45. Murakami, T.; Suzuki, Y.; Oishi, H.; Ito, K.; Nakao, T. Tracing the source of difficult to settle fine particles which cause turbidity in the Hitotsuse reservoir. Jpn. J. Environ. Manag. 2013, 120, 37–47. https://doi.org/10.1016/j.jenvman.2013.02.003. (In Japanese)
46. Murakami, T.; Nakayama, H.; Mizuguchi, S.; Sugio, S.; Ootawara, N. Effective Countermeasures to long-term Turbid Water Effluence from the Hitotsuse Dam Reservoir, In Proceedings of the 6th Int. Symp. Environ. Hyd., Athens, Greece, 23-25, June, 2010. https://doi.org/10.1201/b10553

47. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In KDD '16: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, California, USA, 13-17, August, 2016.
48. https://microsoft-image-composite-editor-64bits.en.softonic.com/?ex=RAMP-1768.3
49. Irie, M.; Arakaki, S.; Suto, T.; Umino, T. Classification of River Sediment Fractions in a River Segment including Shallow Water Areas Based on Aerial Images from Unmanned Aerial Vehicles with Convolution Neural Networks. Remote Sens. **2024**, 16, 173. https://doi.org/10.3390/rs16010173