**Analyzing global trends of water quality using GLOBE hydrosphere data**

Koich Kameyama, Kazuki Tane, Teruma Shinagawa, Kazuki Shiroto,

Kaishi Takahashi, Kose Oka, Ayumu Oura

Ikuta High School

Kanagawa, Japan

Teacher: Ms. Reiko Nezu

February 3, 2025

**ABSTRACT**

We belong to the Science Club at Ikuta High School in Kanagawa Prefecture, Japan. Since 2016, we have been participating in GLOBE activities, conducting water quality surveys of a small river called Hirase River, which is located very close to our school, using the GLOBE Hydrology Protocol once a week. In this report, we analyzed the global trends of water-related data (i.e., water temperature, pH, nitrates, conductivity, and dissolved oxygen) collected from all around the world through GLOBE using a Python program. We also collected some environmental variables (i.e., urban areas around each GLOBE observation site and river width) and evaluated the relationships among the data of water qualities and the environmental variables. We found that the average water temperature has been rising year by year as global trend, indicating the effects of global warming. We also found that the overall GLOBE observation sites were divided into three categories: (a) narrow rivers located in a rich natural setting, (b) wide rivers in urban areas, and (c) narrow rivers in urban areas by cluster analysis. We examined the factors affecting water qualities in these three types of rivers and clarified the characteristics of the Hirase River, which we are observing, in comparison to these three types of rivers. In addition to the weekly hydrosphere observations in our school, we are conducting two types of laboratory experiments: one is how to use phytoplankton and aquatic plants to improve water quality, and the other is testing whether Fulvic acid in water promotes photosynthesis. By examining global trends of water quality data around the world, we could find some basic insights into our future experiments for water purification and photosynthesis promotion in the river aiming at mitigation of global warming.

1. Introduction

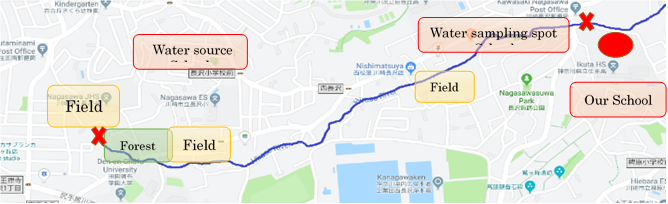
Our school, Ikuta High School, is located in the urban area of Kawasaki City in Kanagawa Prefecture, Japan. There is also a rich natural park called Ikuta Ryokuchi nearby. Close to Ikuta High School, there is a small river called Hirase River, which is about 3 meters wide (Figure 1). We have been conducting water quality surveys (i.e., water temperature, pH, COD, phosphoric acid, ammonium, and nitrates) using pack tests once a week since 2016 (Figure 2).

Since 2018, we have been conducting laboratory experiments under various conditions to determine how much nitrogen and phosphorus (i.e., referring to as nitrates) can be absorbed by the aquatic plant, Canadian waterweed (*Elodea canadensis*), in order to get insights for improving water quality in the Hirase river. Additionally, starting in 2024, we will also conduct experiments to verify whether divalent iron ions can promote photosynthesis, aiming at mitigating global warming by promoting photosynthesis in rivers. By using Euglena (genus of freshwater fish in the aquatic plant family Euglenaidae), we have already confirmed that iron fulvate containing divalent iron ions (i.e., ferric Fulvic acid) promotes photosynthesis. As the next issue, we are verifying whether the same trend is confirmed by typical water plants such as Canadian waterweed (Reference 1).

In this report, we aimed to detect global trends of the effects of global warming on the world's rivers by using hydrosphere data extracted from GLOBE Advanced Data Access Tool. We also aimed to identify the characteristics of the Hirase River when compared to trends in rivers around the world, and to utilize this information for our future research for water purification and mitigation of global warming. For the data analysis in this research, we used a Python program (Reference 2 - 3) and a tool called Vscode. We also created various graphs to visualize the analysis results.

1. Research Question and Hypothesis

* What kinds of global trends can be identified from the hydrosphere data extracted by GLOBE?
* What are the characteristics of the rivers at the GLOBE observation site?
* As a global trend, what environmental factors are associated with river water quality?
* What kinds of insights can be obtained from the global analysis of GLOBE hydrosphere data to our future research aimed at mitigating global warming?



500ｍ

N

Figure 1. Location of Hirase River and our school



Figure 2. Observation activities

1. Materials and Methods
2. Local hydrosphere data observed at Hirase river

The GLOBE Hydrosphere protocols was utilized for water quality surveys. The measurement items for the water quality surveys conducted at Ikuta High School include water temperature, pH, COD, phosphate phosphorus, and nitrates. We set the two observation sites along the Hirase river (Figure 1). At the observation sites, we measured only water temperature, and other data were measured at the laboratory in our school after collecting water samples from the river. Among the measured data, the values for nitrates have been uploaded to the GLOBE website.

1. GLOBE hydrosphere data

GLOBE hydrosphere data was downloaded through the Advanced Data Access Tool on the GLOBE site. The downloaded data includes water temperature, pH, nitrates, conductivity, and dissolved oxygen. The downloaded data was combined using database software, and average values for each observation site, yearly averages, and averages for each observation point were calculated using spreadsheet software, and graphs were created. The yearly averages were calculated using data registered over the past 20 years, while other data used for analysis were from 2016 onwards to compare with the data from Ikuta High School.

The downloaded data included the latitude and longitude of the measurement sites. We printed out each observation site by showing Google Maps based on the location data (Figure 3). A circle with a radius of 2 km was drawn on the printed map, and the proportion of urban areas within that circle was expressed as a number from 0 to 10 (10 = 100%). The approximate width of the river was represented in three stages: 1 for within 5 m, 2 for 5 to 50 m, and 3 for over 50 m. These values were added to the downloaded data. Ultimately, the necessary data was extracted and saved as a CSV file.

The number of data points is as follows: there are a total of 1,152 GLOBE data entries, 17 entries representing the average values for each measurement location, and 147 entries for the measurement data from Ikuta High School. The CSV file was analyzed using Python, starting with simple linear regression. The tool used for Python was Vscode (Figure 4).

1. Data analyses

For evaluating global and local trends in water qualities over time, we presented simple linear regressions between the variables and year. We also evaluated correlations among the variables of water qualities and surrounding environmental variables (i.e., percentage of urban areas and river width) in the form of heat maps at both global and local levels.

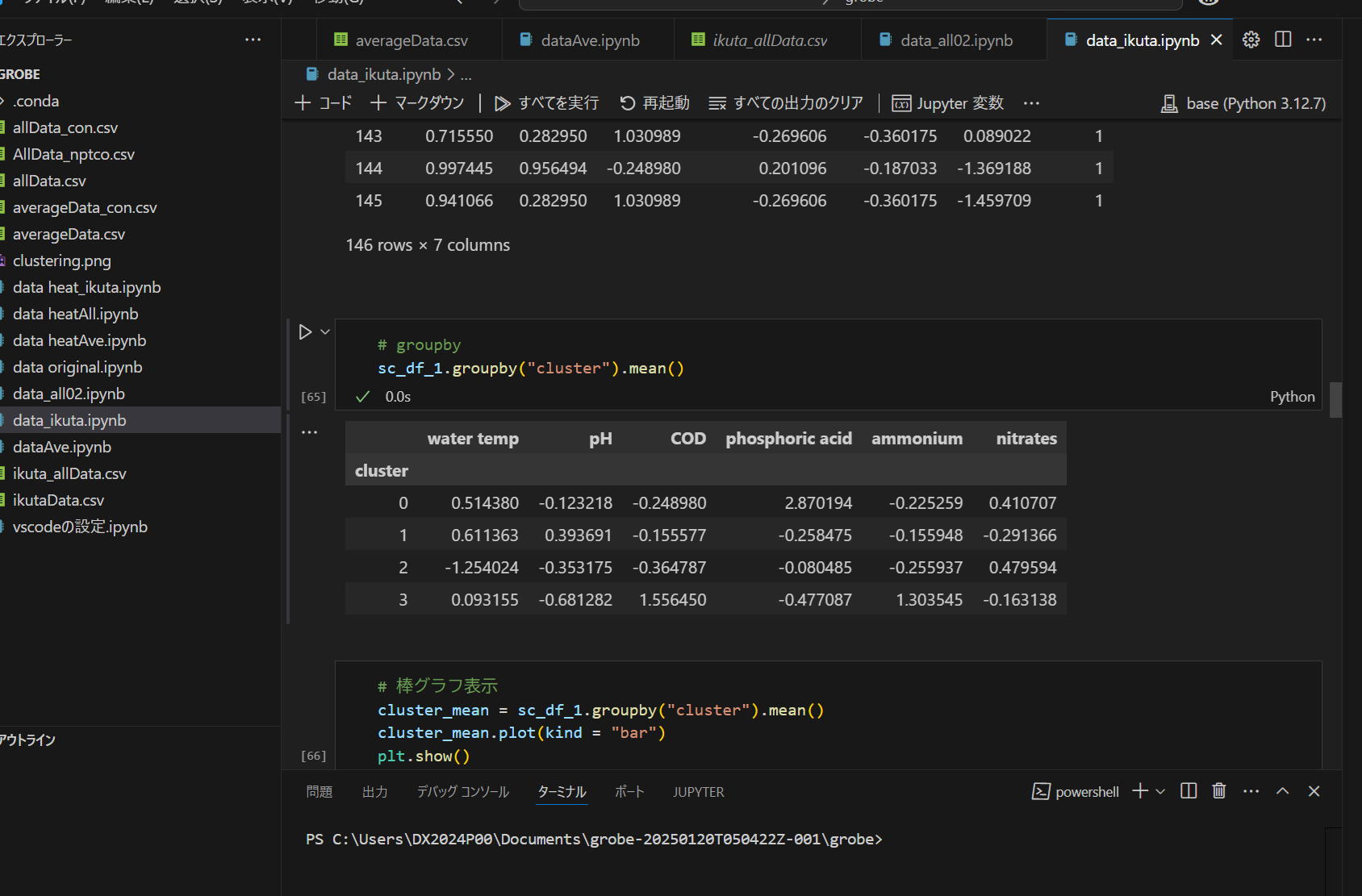
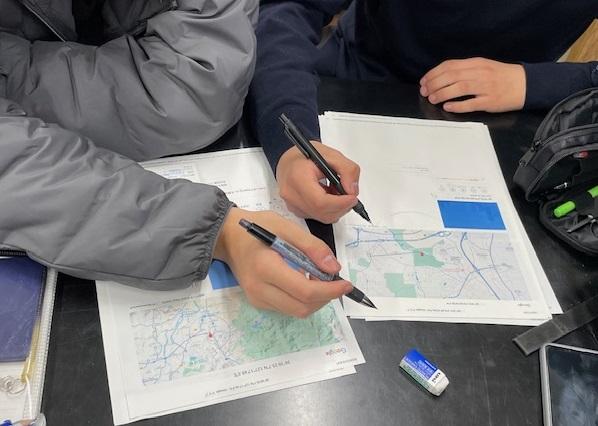
To characterize the rivers that are GLOBE observation sites, we performed the cluster analysis using the k-means method with the data matrix consisting of water qualities (i.e., water temperature, pH, nitrates, conductivity and dissolved oxygen) and the other environmental variables obtained from the maps (i.e., percentage of urban areas and river width) per observation points of GLOBE sites. Cluster analysis involves dividing data into groups with high similarity. The optimal number of groups was calculated using the elbow method.

Figure 4. Python Screen

Figure 3. Assessing the degree of urbanization in the surrounding area of the observation points of GLOBE sites and the river widths

1. Results and Discussion
2. Global and local trends of hydrosphere data

From the global water temperature data, we identified that the average water temperature has been rising year by year, indicating the effects of global warming (Figure 5). Although there is some variability in the data from Ikuta High School, an increase in water temperature is also observed (Figure 6). The nitrogen concentration measured in both the global GLOBE data and at Ikuta has shown a slight increase (Figures 7 and 8).

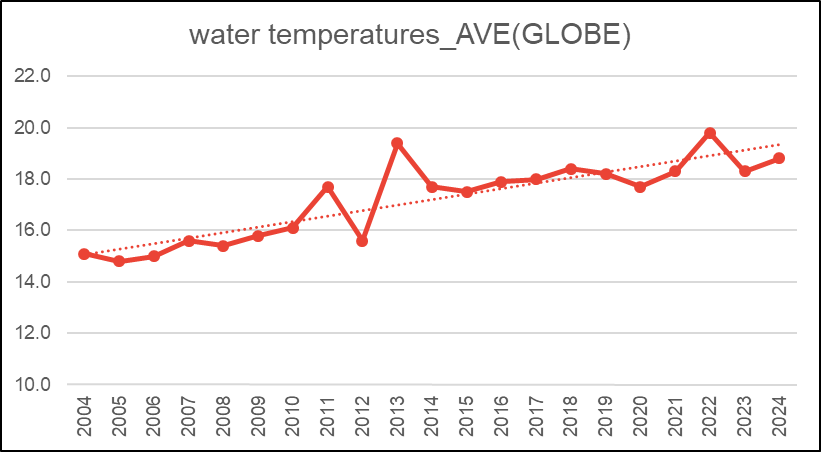
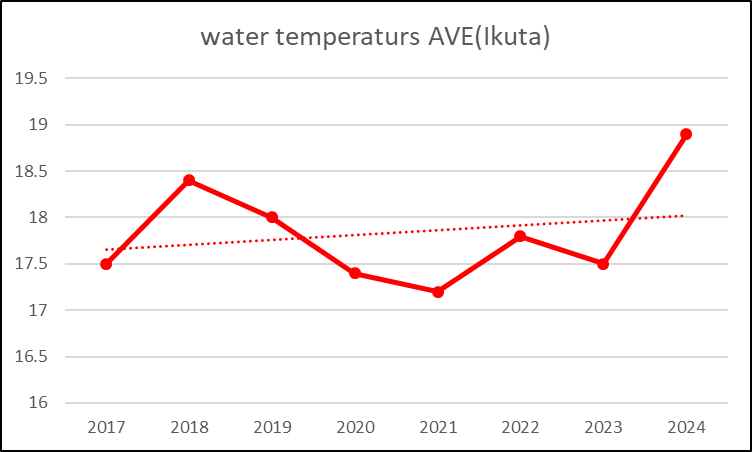


Figure 5. Global trend of water temperature over time

Figure 6. Local trend of water temperature over time (Ikuta High School)

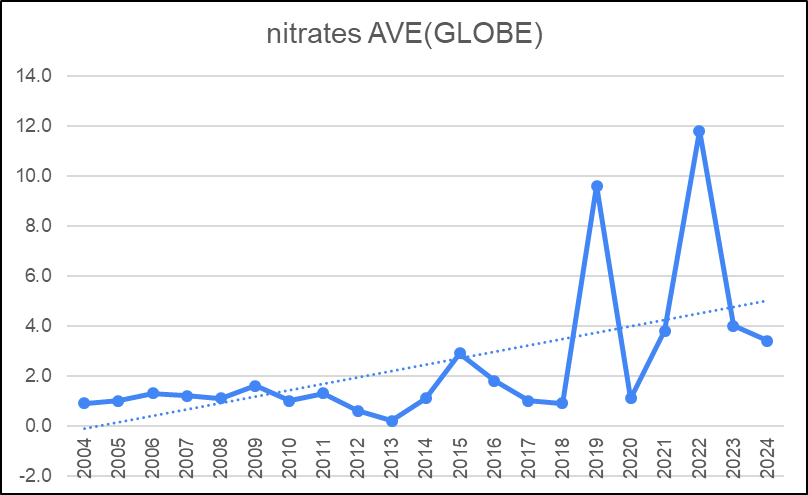
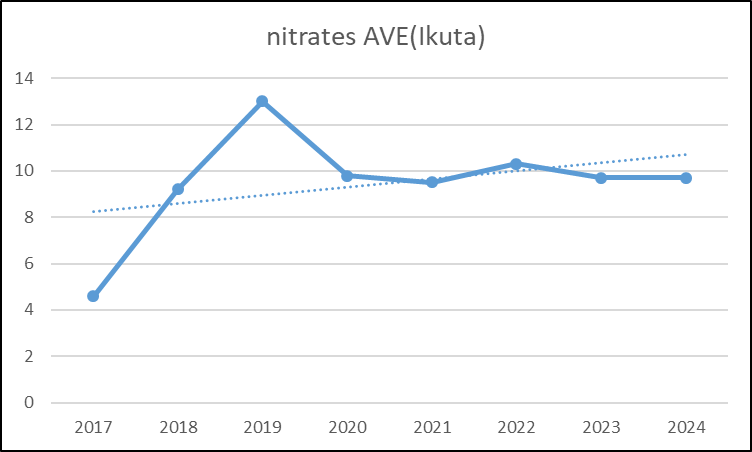


Figure 8. Local trend of nitrates over time (Ikuta High School)

Figure 7. Global trend of nitrates over time

1. Factors related to water quality

We presented a heatmap showing the results of simple regression analysis for all data from GLOBE observation sites (Figure 9), the average values for each observation site (Figure 10), the data extracted only from the GLOBE data at Ikuta High School (Figure 11), and all data measured at Ikuta High School (Figure 12). In the heatmap showing the correlation of data from all GLOBE sites, a positive correlation was observed between the proportion of urban areas and river width. This indicates that, among the observation sites, many urban rivers are located in the lower reaches of rivers.

In the heatmap of average data, this relationship is somewhat weakened. Particularly in the average data for each observation point, there was a strong negative correlation between nitrogen and dissolved oxygen. The observational data from the GLOBE showed a weak negative correlation between nitrates and pH. In the data measured at Ikuta High School, no correlation was observed for either dataset.

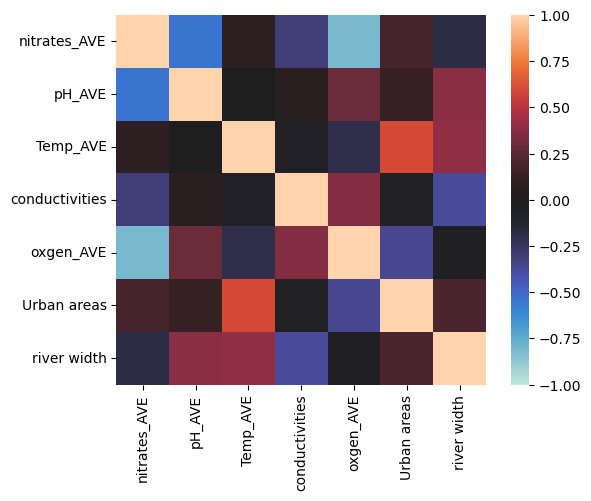
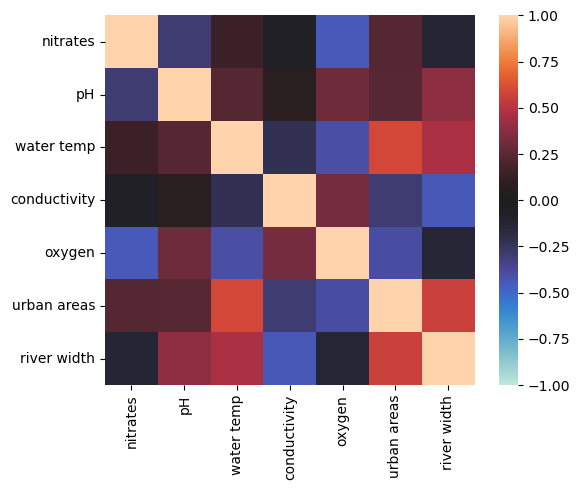


Figure 10. Correlations among the average values for each observation site of GLOBE

Figure 9. Correlations among all data from GLOBE observation sites

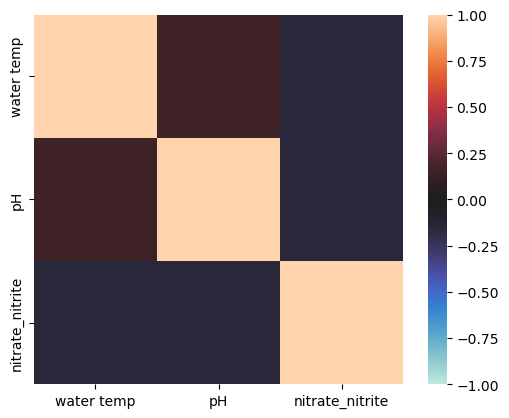
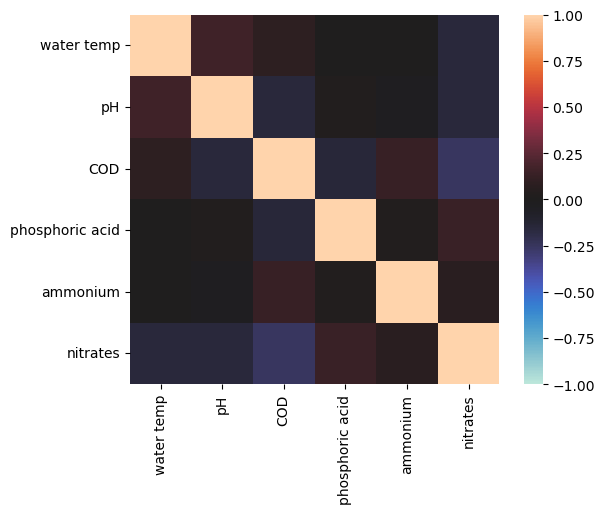


Figure 11. Correlations among the extracted data from the GLOBE database at Ikuta High School

Figure 12. Correlations among all data measured at Ikuta High School

1. Results of cluster analyses

グラフ

AI によって生成されたコンテンツは間違っている可能性があります。The results of the cluster analysis used the elbow method to determine the optimal class division. The elbow method calculates the distance from the centroid of each group to decide the number of classes. The shape of the graph is likened to an elbow, and the value indicated by the bent part of the elbow is said to represent the optimal number of classes, although it is not perfect. As shown in the elbow method graph (Figure 13), it indicates that dividing into 3 to 5 groups is appropriate, and for simplicity in analysis, we classified it into 3 classes. The number of data points when all GLOBE data was grouped (Figure 14) and the content of that grouping (Figure 15) are displayed. Clustering was also performed on the data calculated as the average for each observation point, but since the trends were the same, only the graph of all data is shown. Furthermore, clustering of the Ikuta High School data also resulted in 3 classes, as shown in Figure 16.

Figure 13. Results of the elbow method graph

グラフ, 棒グラフ

AI によって生成されたコンテンツは間違っている可能性があります。

Figure 14. GLOBE Number of data points in each cluster

グラフ

AI によって生成されたコンテンツは間違っている可能性があります。

Figure 15. Characteristics of each cluster by variables of water qualities and the other environmental variables

The GLOBE measurement sites were divided into three categories: narrow rivers in areas rich in nature (Fig. 15, cluster 0), wide rivers in urban areas (Fig. 15, cluster 1), and narrow rivers in urban areas (Fig. 15, cluster 2). The most common observation site was the river in a natural area (Figure 14). The group of rivers rich in nature shows high transparency and dissolved oxygen concentration, indicating that they are clean rivers with minimal human pollution. It is believed that diverse organisms inhabit the surrounding areas. In urban areas, there is a tendency for water temperatures to be higher, which may be attributed to the effects of industrial and domestic wastewater. In the same urban area, rivers with a wider width have lower nitrogen concentrations, while narrower rivers have higher nitrogen concentrations. The dissolved oxygen concentration is higher in wider rivers. Generally, narrower rivers are thought to have faster flows, making it easier for oxygen to dissolve; however, the results are contrary to this expectation. Two reasons can be considered for this: (1) Wider rivers tend to have slower flows and are often shallower, allowing sunlight to reach the riverbed more easily. This facilitates the proliferation of algae and aquatic plants, which absorb nitrogen through photosynthesis (Reference 4) and supply oxygen. (2) Wider rivers have a larger surface area, making it easier to absorb oxygen from the atmosphere. While we cannot dismiss the possibility of (2), we believe that the influence of (1) is more significant. The reason is that urban rivers with wider widths also show higher values in terms of pH. This may be because CO2 in the water is absorbed through photosynthesis, reducing acidic components and resulting in a relatively alkaline water quality.

According to data from Ikuta High School, the clustering results indicate that it is difficult to read the correlations in Figures 16 and 17, but there is a tendency for nitrogen concentrations to be higher in spring and autumn. Since there are fields around the observation points, the application of fertilizers during this period may be related.

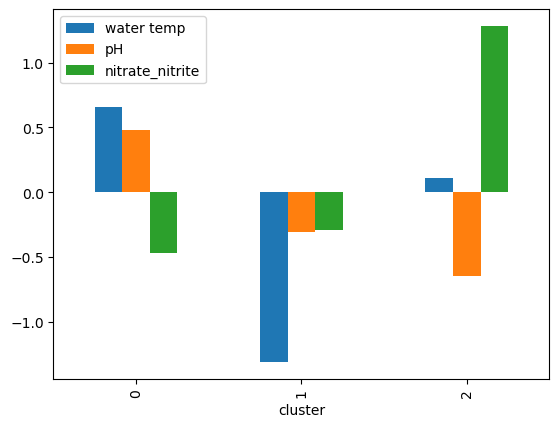


Figure 16. Characteristics of clustered data of Ikuta High School extracted GLOBE database

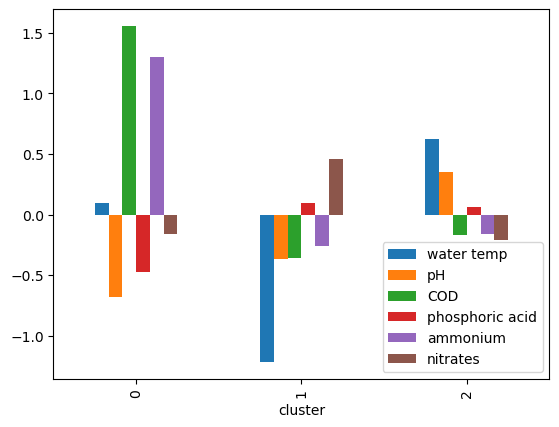


Figure 17. Characteristics of clustered data of Ikuta High School with all the observed data

1. Conclusion and Future Issues

In this study, we found that the more types of measurement items we include, the more diverse perspectives we can analyze the data from. Our analysis results are not complete, and if new observational data is added, it is possible that entirely different trends can be detected in the future. As a future development issue, we will conduct measurements of dissolved oxygen and transparency in addition to the items we have been measuring so far. If there are items related to river flow velocity in the data updated to GLOBE, there may be new discoveries.

As a result of the data analysis conducted this time, we could find the possibility that dissolved oxygen in the water is related to photosynthesis, which is supporting hypotheses in our experiments. As we mentioned in Introduction, we were conducting experiments to verify whether ferric Fulvic acid promotes photosynthesis by using the rate at which aquatic plants absorb nitrogen as an indicator of the strength of photosynthesis, but we have not yet obtained very clear results. We aim to achieve the goal of gaining insights into promoting photosynthesis by measuring the concentration of dissolved oxygen and conducting further verification experiments.

The GLOBE initiative, which collects data from observation sites around the world following the GLOBE protocol, is gaining importance. By continuing GLOBE's activities and gathering data from around the world, including from us, we may obtain hints to prevent global warming.

Bibliography/Citations:

Sugiyama, S., 2011: Tetsu ha Mahotsukai (Iron is a wizard). Tokyo, Shogakukan

Nakayama, K., 2019: Data Scientist Ikuseikouza (Data Scientist Training Course). Tokyo, Mainabi

Kato, K., 2019: Kikaigakusyu Zukan (Encyclopedia of Machine Learning).Tokyo, Shoeisha

Takahashi, E., 1976: Nogyo to Kagaku (Agriculture and Science). Tokyo, Jcamagri