Modeling Continuous Water Quality in Real-Time with IoT and an Arduino Kit

Sania Rehman[1] and Denise Vinton[#]

[1]Central High School, USA
[#]Advisor

ABSTRACT

Quality water is essential to many aspects of life including agriculture, industry, and the economy. According to the World Health Organization (WHO), poor water quality is the root cause of 80 percent of the diseases in the world. In addition to supporting the health of people, the quality of water is one of the most important factors in a healthy ecosystem that supports a diversity of plants and wildlife. It is therefore imperative for us to continue to monitor the quality of our water supply. Some of the current methods for tracking water quality involve labor-intensive processes that can be expensive, time-consuming, and prone to data errors. As a result, this project aims to see if modern-day open-source technology can be deployed to make the process of water quality monitoring more effective and efficient by designing a proof-of-concept that visually depicts a water quality index (WQI) in real-time. The method used was to collect sample data from Fellows Lake in Missouri using Arduino sensors, IoT and then to use Python to compute an arithmetic WQI based on various water quality parameters. The results of the data indicated a WQI score of 70.14, which can be categorized as poor water quality. This proof-of-concept project demonstrates that water quality monitoring can be done by using modern-day open-source technology. While more research is needed, this new method for tracking water quality could be a game changer in terms of how and where we can track water quality.

Introduction

The number of people living on this planet is steadily increasing, which is driving up the need for water. Water is essential to many aspects of life, including agriculture, industry, and the economy. The rate of urbanization and industrialization has increased at an alarming rate, which has had a significant impact on both the quality and quantity of water. Drinking water is becoming increasingly problematic in several countries. According to the World Health Organization (WHO), poor water quality is the root cause of 80 percent of the world's diseases (Ramakrishnaiah et al., 2009). In addition to supporting the health of people, the quality of water is one of the most important factors in a healthy ecosystem that supports a diversity of plants and wildlife (Michigan Sea Grant).

According to the Centers for Disease Control and Prevention (CDC), some of the many possible sources of water contamination include sewage release, naturally occurring chemicals and minerals, local land use practices, manufacturing processes, and malfunctioning on-site wastewater treatment systems (Centers for Disease Control and Prevention). The presence of these water contaminants found in our water supply can lead to health issues for plants, wildlife, and people. It is therefore imperative for us to continue to monitor the quality of our water supply.

Some of the current methods of fresh water quality monitoring rely on observing macroinvertebrates (Kenney et al, 2009). Macroinvertebrates are organisms without skeletons that can be seen with the naked eye. Some examples of macroinvertebrates are the mayfly, stonefly, dragonfly-adult, dragonfly-nymph, caddisfly-larva, snails, crayfish, and so on. Pollution reduces the variety of these organisms surviving in the water, so healthy waters usually have many different kinds of these critters. Having a large variety of macroinvertebrates indicates good water quality and or good habitat conditions. The more types of these species you see in a body of water, the better the quality of water.

Other methods of tracking water quality include using established water quality indices or creating a new water quality index (Brown et al, 1972). A water quality index (WQI) is a means of summarizing complex water quality data that is based on a rigorous scientific method into easily understood or simple terms (e.g., poor, fair, good). These terms can then be communicated to the public. These indices are also helpful in allowing for management decisions to be made by various state, regional, and local agencies. Some of the more commonly used and existing water quality indices that have been established include the National Sanitation Foundation Water Quality Index (Brown et al, 1970) and the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI). The CCME WQI was developed by the British Columbia Ministry of Environment, Lands and Parks department (Rocchini and Swain 1995). Many other indices have been created by various state, regional, and local agencies as well. The limitations of water quality indices are that they are generalized, they do not give the whole story, they are often limited in terms of time and space, and there is no single index that is used to rule them all. For this reason, there is still some controversy, among water quality scientists, in terms of how to best mathematically quantify water quality using a WQI.

The methods described above for tracking water quality both require labor-intensive processes that can be expensive, time-consuming, and prone to data errors (Chen et al, 2022). Tracking water quality via macroinvertebrates or via a WQI requires that trained personnel be involved in collecting and processing the data.

To offset some of these costs, various state and county departments have workshops in place to get citizen scientists, or the public, involved with helping monitor water quality via macroinvertebrates. One such program is offered through the Missouri Stream team. Once individuals have gone through the necessary training, they are then certified to become water quality monitors (Water Quality Volunteers).

Collecting water samples that can be used to calculate a water quality index also requires trained personnel to go out in the field to collect the water samples. Some of the bodies of water are in remote rural areas that are difficult and costly to travel to as well. These water samples then have to either be transported to the lab for processing or evaluated via a field-testing kit. Due to this, community science programs such as the ROSS (Reservoir Observer Student Scientist) program have been created. These types of programs are mutually beneficial to the public and to scientists because they help make science attainable to students and allow scientists to collect a high volume of data that would otherwise be unattainable (Kinzinger and North, 2021).

As a result of the inefficiencies associated with current methods of water quality monitoring, this project aims to see if modern-day open-source technology can be deployed to make the process of water quality monitoring more effective and efficient. Specifically, and inspired by previous studies (Kinar and Brinkman, 2022), this project is to define a proof-of-concept method using data that is retrieved from Arduino sensors and IoT to then create a math model that visually depicts a WQI in real-time.

The sensors and the parameters used for the WQI are pH, turbidity, conductivity, and dissolved oxygen (DO).

Objective

This project aims to see if modern-day open-source technology can be deployed to make the process of water quality monitoring more effective and efficient. Specifically, and inspired by previous studies, this project is to develop a proof-of-concept method using data that is retrieved from Arduino sensors and IoT (Internet of Things) to then create a math model that visually depicts a water quality index (WQI) for real-time water quality monitoring.

Using Python and Jupyter Notebook, this project's goal is to create a water quality index (WQI) model and depict the relationships between various water quality parameters leveraging data from an Arduino and IoT (Internet of Things) application. The end result will be to have a proof of concept that demonstrates a more effective and efficient way to track water quality using modern-day open-source technology.

Methods

Materials

Hardware

- ARDUINO
- Power Supply
- ESP8266 Wi-Fi module
- pH sensor
- Turbidity sensor
- Conductivity sensor
- Temperature sensor
- DO sensor
- LCD display
- LED indicator

Software

- Python
- Jupyter Notebook
- ThingSpeak App (IoT)

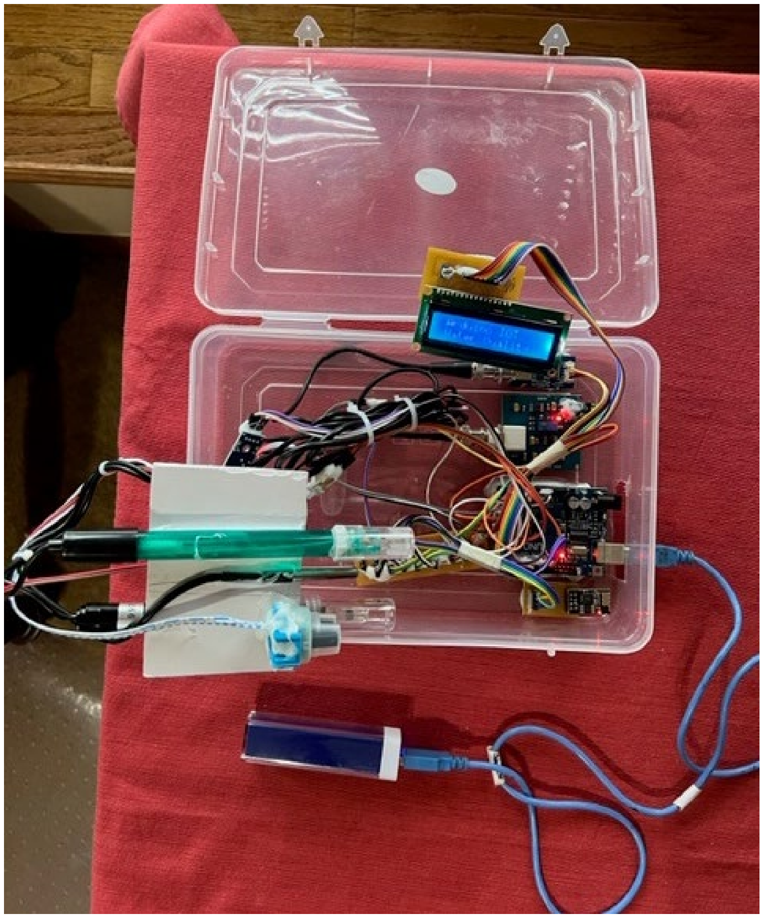


Figure 1. Picture of assembled Arduino with 5V power supply

Procedures

Researchers have proposed different approaches to compute a WQI. Sener et al. (2017) proposed the following first 4 steps noted below to calculate a weighted arithmetic WQI, which is one of the most commonly used methods to assess the quality of water (Srinivas et al., 2011), (Abuzaid, 2018).

1. Selecting the set of variables to compute WQI
    a. The first step in any WQI method is the selection of parameters. Water has approximately 78 different physical, chemical, and biological parameters. Selecting from among these parameters is driven by what is being evaluated. This project will be using the following physical and chemical parameters: pH, Turbidity, Conductivity, and Dissolved Oxygen. These are commonly tested parameters, and our Arduino has these sensors plus a temperature sensor. The temperature was omitted from the WQI calculation due to the standard and ideal values (standard and ideal values will be explained in section 1b below) not being readily available within the literature. However, the temperature will be evaluated to understand the relationships between it and the other variables.
    b. Various Organizations such as the World health organization (WHO) and the Centers of Disease Control (CDC) have defined ideal and standard values of water quality parameters (Maheswaran et al., 2014), (Suganthi et al., 2019). Ideal values (V_i) and standard values (S_n) of the water parameters used in this project are listed in Table 1 below:

Table 1. Standard value (S_n) and Ideal value (V_i) of parameters

Parameters	S_n	V_i
Potential of Hydrogen (pH)	8.5	7
Turbidity	5	0
Dissolved Oxygen (DO)	5	14.6
Electrical Conductivity (EC)	300	0

2. Create sub-indices (q_i) of selected variables.
    a. The sub-index for each parameter can be created by using the following formula.

$$q_i = \frac{V_n - V_i}{S_n - V_i} \times 100$$

Here,
V_n = Observed actual value
V_i = Ideal Value
S_n = Standard Value

3. Compute the relative weight of each variable.
    a. The Relative weight (Rw_i) is calculated as $w_i/\sum w_i$, where w_i is the weight of each parameter. The weight of each parameter is assigned a weight according to its standard value as follows: $w_i = 1/S_n$

4. Aggregate the products of the sub-indices and associated weights to compute the final WQI score.
    a. The WQI can be calculated as follows.

$$WQI = \frac{\sum(q_i.Rw_i)}{\sum(Rw_i)} = \frac{q_1.Rw_1 + q_2.Rw_2 + \ \dots q_n.Rw_n}{Rw_1 + Rw_2 + \ \dots + Rw_n}$$

    b. Below is an example of the completed 4 steps above using hypothetical, but realistic, observed values.

Table 2. Example of WQI Computation using Hypothetical Observed Values

Parameters	Unit	(Standard Value) S_n	(Ideal Value) V_i	(Observed Value) V_n	Weight $w_i = 1/S_n$	Relative weight $Rw_i = w_i/\sum w_i$	$q_i = \frac{V_n - V_i}{S_n - V_i} \times 100$	$Q_i(Rw_i)$
Potential of Hydrogen (pH)	---	8.5	7	8.5	0.117647059	0.225818592	100	22.58185924
Turbidity	NTU	5	0	2.75	0.2	0.383891607	55	21.11403839
Dissolved Oxygen (DO)	mg/l	5	14.6	1.535	0.2	0.383891607	136.09	52.24380881
Electrical Conductivity (EC)	µmhos/cm	300	0	78.25	0.003333333	0.006398193	26.08	0.166864885
					$\sum w_i$ = .520980392	$\sum Rw_i = 1$		WQI = 96.11

    c. Based on the index, water is classified into the following categories:

Table 3. WQI Value Categories

WQI Value	Rating

0 – 25	Excellent
26 – 50	Good
51 – 75	Poor
76 – 100	Very Poor
Above 100	Unsuitable

   d. So, the WQI score in our example (i.e., 96.11) would be classified as very poor quality.

5. Travel to Fellows Lake in Missouri
   a. For this project, we will determine the water quality index for Fellows Lake in Missouri because it is used for human consumption.
6. Power the Arduino on by connecting it to a 5V power supply.
7. Connect the Arduino to a hotspot via a cellular device.
   a. The Arduino used for this project is preconfigured to connect to a hotspot named *waterqualityproject*.
8. Insert temperature, pH, conductivity, dissolved oxygen, and turbidity sensors into multiple sample sites within Fellows Lake.
   a. The readings will be uploaded to the ThingSpeak app in real time based on the hotspot Wi-Fi connection.
9. Export the data from the ThingSpeak app
10. Use Python, Jupyter Notebook, and the data exported from ThingSpeak to create a model that displays the relationships between the various parameters, including temperature, and to calculate the WQI.
    a. This project used Python (i.e., programming language) to compute the WQI and to display the various relationships between the parameters. Jupyter Notebook is a web-based, interactive, and lightweight text editor that was used for writing the Python code and displaying the results. For additional functionalities like processing, visualization, numeric computation etc., various Python libraries such as pandas, matplotlib, seaborn, and NumPy were imported.

Results

Our dataset contains 25 samples with 6 attributes representing site name, temperature, pH, turbidity, DO, and EC. The first few samples of the dataset are displayed in Figure 2 below.

	Site	Temp (Celsius)	pH	Turbidity (NTU)	DO (mg/L)	Conductivity (µmhos/cm)
0	Fellows Lake	16.3	8.05	2.96	6.78	226
1	Fellows Lake	16.2	7.99	3.49	6.84	225
2	Fellows Lake	15.9	8.54	1.94	6.93	223
3	Fellows Lake	15.7	8.19	1.31	7.23	221
4	Fellows Lake	15.5	8.43	1.42	7.53	216

Figure 2: First few samples of the dataset

It is normal practice to take the average of all of the readings of the site to eliminate chances of error in measurements or outliers in data. We averaged 25 readings of the Fellow Lake site, and the average values of all five parameters are shown in Table 4 below.

Table 4. Average Values of Parameters

Volume 13 Issue 2 (2024)

Parameter	Average value	Units
Temp	15.568	Celsius
pH	8.412	
Turbidity	2.6856	NTU
DO	7.6468	Mg/L
EC	216.12	µmhos/cm

We then used these values to calculate the WQI score. Based on these values, the water quality index value is 70.14. As this value is between the 51 to 75 range, the quality of this water is categorized as poor. The following figure 3 below shows the output from Jupyter Notebook, and Table 5 below shows the detailed breakdown of the calculation.

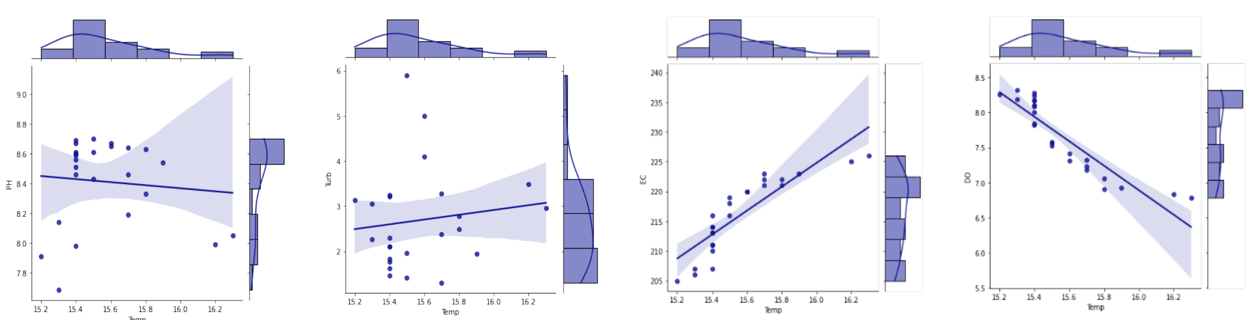
	Temp	PH	Turb	DO	EC	WQI_Val	WQI
Site							
Fellows Lake	15.568	8.412	2.6856	7.6468	216.12	70.142518	Poor

Figure 3. WQI Output from Jupyter Notebook

Table 5. Detailed Breakdown of WQI Calculation

Parameters	Unit	(Standard Value) S_n	(Ideal Value) V_i	(Observed Value) V_n	Weight $w_i = 1/S_n$	Relative weight $Rw_i = w_i/\sum w_i$	$q_i = \frac{V_n - V_i}{S_n - V_i} X 100$	$Q_i(Rw_i)$
Potential of Hydrogen (pH)	---	8.5	7	8.412	0.117647059	0.225818592	94.13333333	21.25705683
Turbidity	NTU	5	0	2.6856	0.2	0.383891607	53.712	20.619586
Dissolved Oxygen (DO)	mg/l	5	14.6	7.6468	0.2	0.383891607	72.42916667	27.80494919
Electrical Conductivity (EC)	µmhos/cm	300	0	216.12	0.003333333	0.006398193	72.04	0.460925856
					$\sum w_i = .520980392$	$\sum Rw_i = 1$		WQI = 70.14

The effect of temperate on the various other parameters was evaluated as depicted in the regression plots in Figure 3 below.



(a) Temp vs. pH (b). Temp vs. Turbidity (c). Temp vs. EC (d). Temp vs. DO

Figure 4. Effect of Temperature on pH, Turbidity, DO, and EC.

As the data points in the graphs of temp vs pH (Figure 3a) and Temp vs turb (Figure 3b) are more scattered, this implies that there isn't a strong correlation between these variables. Turbidity and pH appear to be independent of temperature.

In the temp vs DO (Figure 3d) graph, we do see that as temperature increases, DO decreases. This implies that temperature and DO are inversely related.

In the temp vs EC (Figure 3c) graph, we see that as temperature increases, EC also increases, which implies that there is a positive relationship between these two variables.

Next, we evaluate the relationship of one parameter to all of the other parameters by creating pair plots as shown in figure 4 below.

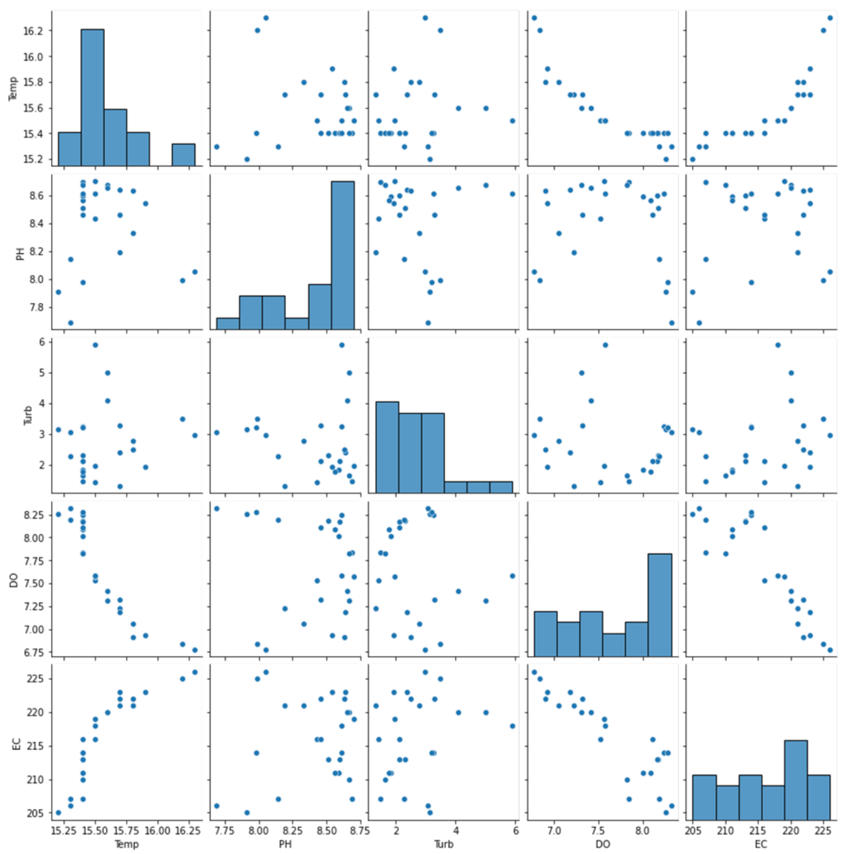


Figure 5. Pair Plots of all of the Variables

The pair plot is a great way of evaluating the dependencies between various attributes. If the data points are more scattered, it implies that there is less of a correlation between the pair of attributes. If the graph exhibits some linearity, then there is some relationship between those attributes.

While the pair plots are a good visual depiction of the relationship between various variables, they are subjective in that they do not give any quantitative measure of the relationship. To check how strong the relationship is, we computed the correlation between every pair of attributes. The correlation represents the strength of the dependency between two variables. The value of the correlation varies between -1 and 1. The value of -1 indicates a strong negative relationship (i.e., as one variable increases, the other decreases sharply). The value of 1 indicates a strong positive relationship (i.e., as one variable increases, the other decreases). The value of 0 indicates no correlation between the attributes. The correlation between all of the attributes in our dataset is depicted in Figure 5 below.

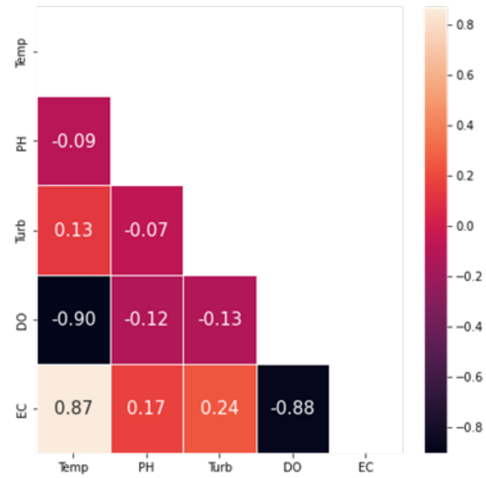


Figure 6. Correlation between all of the Parameters

From the above correlation plot (Figure 5), we can see that the correlation between temperature and DO is negative, and they are strongly correlated with a correlation value of -90. Similarly, the correlation between temperature and EC is positive, and they are strongly correlated with a correlation value of .87. The correlation values of all of the other attributes can be seen in the correlation plot.

Discussion

As demonstrated from this proof-of-concept project, it does appear that water quality monitoring can be done more efficiently and effectively by using modern-day open-source technology. This new method for tracking water quality is a very powerful idea that could transform the way that we track water quality. Furthermore, it will provide for the ability to monitor water quality in areas that would have otherwise been difficult, or impossible, to monitor. Such areas might include remote and rural areas where there are few people nearby (Nowicki et al., 2020).

The results section demonstrated that using an Arduino with IoT sensors will allow for the collection of water quality readings that can then be used to derive a water quality index in real time. This eliminates the need for trained personnel to spend their time and energy on making frequent trips to collect water samples.

The results from this project were compared with sample readings that were collected by the City Utilities organization during the same period. The results from the sensors used during this project were consistent with the results from the City Utilities sample readings.

The results section also demonstrated various relationships between the variables tested in our project. Some of these relationships are validated within the literature while others may not have been. For example, in our results, the relationship between temperature and DO show to be an inverse relationship. And, the relationships between temperature and EC show to be a positive relationship. The literature also validates these findings ("Dissolved Oxygen and Water" and "National Aquatic Resource Surveys"). However, in our results, pH and turbidity appear to be independent of temperature. This could be due to our limited sample size and the short period between the collection of our samples.

Understanding these relationships is helpful to those monitoring water quality as it provides additional insight into what interventions might be necessary given a certain set of conditions. For example, understanding that as the temperature of water increases, dissolved oxygen decreases might prompt water quality agents to assess the dissolved oxygen levels more frequently in the summer when the water temperature is increasing. The tools provided in this project would allow for this to happen.

Of course, more research in this area will be needed to better understand how these Arduinos and sensors can be deployed out in the field in a secure way for longer periods. For example, we may want to explore how mechanical devices can be used to remotely deploy the sensors into the water for real-time readings. We will want to look at how we could protect and shelter these devices from the elements or wildlife. We will want to explore further how we could use different power supplies that last longer and are more efficient. However, we believe that this project demonstrates a ground-breaking opportunity to further explore all of these things to make water quality monitoring more efficient and effective and possible for monitoring more bodies of water.

Conclusion

The projects aim was to develop a proof-of-concept to see if modern-day open-source technology can be deployed to make the process of water quality monitoring more effective and efficient by leveraging Arduino sensors, IoT (Internet of Things), and Python to create a water quality index (WQI) for real-time monitoring. The results of the proof-of-concept clearly demonstrated that this is achievable but more research will be needed to fully deploy these devices on or near the bodies of water in a more permanent and secure manner. Once that has been determined, this method of water quality monitoring could be a game changer in terms of how and where we can monitor water quality. This method has the potential to significantly reduce the cost and time to collect the data.

Acknowledgments

I would like to thank the following individual for their guidance and mentorship during this research process.

- Mrs. Denise Vinton

References

Abuzaid, A. (2018). Evaluating surface water quality for irrigation in Dakahlia Governorate using water quality index and GIS. Journal of Soil Sciences and Agricultural Engineering, Mansoura University, 9(10), 481-490.

Brown, R. M., McClelland, N. I., Deininger, R. A., & O'Connor, M. F. (1971). A water quality index—crashing the psychological barrier. Proceedings of a symposium Indicators of Environmental Quality (pp. 173-182). Pennsylvania: Springer US.

Brown, R. M., McClelland, N. I., Deininger, R. A., & Tozer, R. G. (1970). A water quality index-do we dare. Water and sewage works, 117(10).

Chen, C.-H., Wu, Y.-C., Zhang, J.-X., & Chen, Y.-H. (2022). IoT-Based Fish Farm Water Quality Monitoring System. Sensors, 22(17), 6700.

Drinking Water. (2020, October 28). Retrieved January 4, 2023, from Ceters for Disaster Control and Prevention: https://www.cdc.gov/healthywater/drinking/public/water_quality.html

EPA. (2023, May 22). National Aquatic Resource Surveys. Retrieved May 26, 2023, from EPA: https://www.epa.gov/national-aquatic-resource-surveys

Kenney, M. A., Sutton-Grier, A. E., Smith, R. F., & Gresens, S. E. (2009). Benthic macroinvertebrates as indicators of water quality: The intersection of science and policy. Terrestrial Arthropod Reviews, 2(2), 99.

Kinar, N. J., & Brinkmann, M. (2022). Development of a sensor and measurement platform for water quality observations: Design, sensor integration, 3D printing, and open-source hardware. Environmental Monitoring and Assessment, 194(3), 207.
Kinzinger, E., & North, R. (2021). Reservoir Observer Student Scientists. North American Lake Management Society.
Maheswaran, G., & Elangovan, K. (2014). Groundwater Quality Evaluation in Salem District, Tamil Nadu, Based on Water Quality Index. Nature Environment and Pollution Technology, 13(3), 547.
Missouri Stream Team. (n.d.). Retrieved January 26, 2023, from Missouri Stream Team: https://www.mostreamteam.org/water-quality-monitoring.html
Michigan, S. G. (n.d.). Teaching Grate Lakes Science. Retrieved February 11, 2023, from https://www.michiganseagrant.org/: https://www.michiganseagrant.org/lessons/lessons/by-broad-concept/earth-science/water-quality/
Nowicki, S., Koehler, J., & Charles, K. J. (2020). Including water quality monitoring in rural water services: Why safe water requires challenging the quantity versus quality dichotomy. Npj Clean Water, 3(1), 14.
Ramakrishnaiah, C. R., Sadashivaiah, C., & Ranganna, G. (2009). Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. E-Journal of chemistry, 6(2), 523-530.
Rocchini, R., & Swain, L. G. (1995). The British Columbia Water Quality Index. Canada: EP Department, BC, Ministry of Environment, Land and Park.
School, W. S. (2018, June 5). Dissolved Oxygen and Water. Retrieved January 16, 2023, from USGS: https://www.usgs.gov/special-topics/water-science-school/science/dissolved-oxygen-and-water
Şener, Ş., Şener, E., & Davraz, A. (2017). Evaluation of water quality using water quality index (WQI) method and GIS in Aksu River (SW-Turkey). Science of the Total Environment, 584, 131-144.
Srinivas, P., Kumar, G. P., Prasad, A. S., & Hemalatha, T. (2011). Generation of groundwater quality index map-A case study. Civil and Environmental Research, 1(2), 9-21.
Suganthi, M., Ramesh, N., Sivakumar, C. T., & Vidhya, K. (2019). Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District. International Research Journal of Multidisciplinary Technovation, 1(6), 630-635.