Geostatistical Analyses of Some Water Quality Parameters in Laguna Lake: Some Recent Post-Lockdown Findings

Pantolla HG[1*], Malay CA[2,3], Leonisa MEM[2,3], Reyes Jr. RM[2,3], Legaspi MJS[4]

[1]Mathematics and Statistics Unit, Institute of Applied Sciences,
Kolehiyo ng Lungsod ng Dasmariñas, City of Dasmariñas, Cavite, Philippines
Email: hgpantolla@kld.edu.ph
[2]Science Unit, Institute of Applied Sciences,
Kolehiyo ng Lungsod ng Dasmariñas, City of Dasmariñas, Cavite, Philippines
[3]Graduate Student, College of Science and Computer Studies Graduate Studies,
De La Salle University – Dasmariñas, City of Dasmariñas, Cavite, Philippines
[4]National Service Training Program Unit, Institute of Human Potential and Development,
Kolehiyo ng Lungsod ng Dasmariñas, City of Dasmariñas, Cavite, Philippines

[*]*Corresponding Author*

Abstract
This paper underscores the benefits of employing data visualization and geostatistical analyses in monitoring the water quality parameters of Laguna Lake in the Philippines, particularly emphasizing the influence of quarterly seasonality based on the Philippine climate. Preliminary post-COVID-19 lockdown results align with prior literature, highlighting the persistent risks posed by population growth and industrialization of Metro Manila to the ecosystem and economic benefits of the lake. While findings corroborate existing studies, this paper proposes novel avenues for future research, pinpointing locations with elevated water quality parameters and exploring the impact of different seasonal quarters in the Philippines.

The findings offer potential contributions to Sustainable Development Goal 6 and proactive climate change responses. Targeting stakeholders, economic managers, and health officials overseeing Laguna Lake, the study provides a foundation for informed decision-making and anticipatory measures, crucial for addressing potential fish population declines especially during warmer seasons. Proximity to the NCR raises concerns about water quality, necessitating decisive interventions to mitigate long-term human and fish health risks stemming from contaminant accumulation. The integration of the Geographical Information System enhances interpretation by revealing clusters of high or low water quality parameter values, empowering effective resource management. This research may equip authorities with essential insights for safeguarding the sustainability, economic vitality, and well-being of the dependent populations of the lake.

Keywords: Geographical Information System, Laguna Lake, Spatial Analysis, Water Quality Parameters

1. Introduction

1.1 Water quality and human health

Even before the onset of the COVID-19 pandemic, the world was already off track on the path to achieving Sustainable Development Goal 6 (SDG 6), which aims to guarantee universal access to clean water and sanitation by the year 2030. A significant number of individuals across the globe continue to lack access to safe drinking water and sanitation, although both services have been recognized as fundamental human rights for a considerable period. Numerous water sources are experiencing depletion, increased contamination, or a combination of both. Water-intensive sectors such as industry, agriculture, and energy production are experiencing growth to fulfill the demands of a growing population [1].

According to the World Health Organization (WHO), the health consequences of water quality are substantial, regardless of its use for drinking, household purposes, food production, or recreational pursuits. Poor water quality can result in the occurrence of disease outbreaks and contribute to the prevalence of diseases occurring at different time intervals. Efforts aimed at ensuring water safety not only enhance public health but also often promote socioeconomic advancement and overall welfare [2].

Moreover, water quality plays a crucial role in the growth and maturation of fish. Farmers will suffer significant financial losses as a result of the deterioration in water quality [3][4][5][6]. Furthermore, pollutants eventually settle in aquatic ecosystems. Urbanization, industrialization, and agricultural practices are all instances of anthropogenic water pollution. In due course, aquatic environments become polluted due to the excessive utilization of pesticides, fertilizers, and waste from residential and industrial areas. The degradation of water quality can lead to the spread of infectious diseases such as dysentery, diarrhea, and jaundice [7].

Aquatic contamination is a prominent type of pollution that leads to significant negative health impacts and loss of life. The inherent capacity of water to counteract contamination can be compromised if contamination is not regulated [7]. Fluctuations in water quality have diverse economic repercussions, affecting human health, ecosystem health, agricultural and fisheries productivity, as well as recreational and amenity uses. While certain effects can be measured, many others cannot and require non-market valuation techniques to be quantified in monetary terms. Valuation necessitates prior scientific monitoring and comprehension of intricate biophysical relationships. For instance, to distinguish between pollution originating from agricultural activities and pollution from non-agricultural sources, or to track the movement of dispersed pollution through intricate hydrological systems. The latter aspect is significant because the combination of physical distance and time delays in cause-and-effect relationships introduces intricacy in the assessment and comparison of monetary values [8]. Therefore, it is crucial to conduct regular monitoring and control the discharge of pollutants into the nearby aquatic ecosystems [7]. This study would attempt to contribute to monitoring some of the most critical water quality parameters that are being monitored are the biochemical oxygen demand (BOD), dissolved oxygen (DO), and fecal coliform (FC) in the largest lake in the Philippines, specifically in the early periods after the consecutive COVID-19 lockdowns.

1.2 Water quality parameters

The BOD refers to the quantity of oxygen that organisms utilize to metabolize oxidizable organic substances within a specified timeframe [9][10]. BOD serves as a measure of organic pollution in freshwater bodies [11][12][10], and it is closely associated with microbiological contamination. Elevated levels of BOD decrease the amount of oxygen present, leading to the deterioration of aquatic habitats and the loss of biodiversity [13][10]. Additionally, it hampers the usability of water. The primary contributors to high BOD loadings in freshwater systems are human activities, including the discharge of domestic and livestock waste, industrial emissions, and combined sewer overflows. During transportation through the stream network, the levels of BOD are diminished through microbial degradation (river self-purification) and dilution before reaching the seas.

The concentration of dissolved oxygen (DO) is crucial for maintaining water quality in aquaculture [14][5][15]. Sufficient levels of oxygen are essential for intensive fish farming [16][15]. Estimating fish metabolism can be done indirectly by measuring oxygen consumption [17][15]. Dissolved oxygen (DO), in addition to factors such as feed and temperature, plays a crucial role in controlling growth. Maintaining a consistent DO level above a critical threshold is essential for optimal feed consumption, growth, and feed conversion efficiency [18][15]. In addition, changes in hematological, hormonal, biochemical, and osmoregulatory processes in the blood and plasma can be influenced by the availability of DO [19].

Coliforms are a type of bacteria that can survive with or without oxygen, have a negative reaction to the Gram stain, ferment lactose, and do not form spores. They are commonly found in the environment and the intestines of humans and other animals in significant quantities [20][21] The coliform group of bacteria, which includes both pathogenic and non-pathogenic forms, is commonly used as an indicator of microbial contamination [22][21]. Total coliforms, which include bacteria commonly found in soil, plants, and animals, including both fecal coliforms and Escherichia coli, respond to the natural environment and treatment processes like pathogens. An examination of the coliform bacteria provides an approximation of the quantity or density of harmful bacteria present in the sample [23][21]. In 2012, the United States Environmental Protection Agency (EPA) advised the use of thermo-tolerant coliforms, specifically fecal coliforms, as an indicator of water quality standards due to their association with the presence of fecal matter. Although fecal coliforms themselves may not directly result in illness, their abundance indicates an increased likelihood of encountering disease-causing bacteria and/or viruses, such as pathogenic Escherichia coli commonly known as *E. coli* [21]. The EPA advises the assessment of Escherichia coli or Enterococci to further ascertain potential health hazards. Escherichia coli, also known as E. coli, is a benign bacterium that naturally resides in the intestines of warm-blooded animals, including humans [21].

1.3 Objectives of the study

The main objective of the study is to create a geographic information system (GIS) that visually displays the potential detection of spatial clusters with statistically significant high values (hot spots) and low values (cold spots) of BOD, DO, and FO. The results aim to offer insights into the distribution pattern of these common water quality parameters obtained from samples collected at the selected lake as the study setting as well as its tributaries. Should spatial clusters of hot spots or cold spots exist, the maps will be overlaid onto maps depicting spatial interpolations of the BOD, DO, and FO. Such maps may be used by policymakers and experts in providing policies and recommendations, especially in (a) conservation efforts to the lake and (b) valuable advice to the fisherfolk that is economically reliant on the selected study setting. The applied geostatistical analyses applied herein may also contribute to the scientific literature about monitoring water quality such as in chemical properties in an aquatic environment and ultimately, in identifying clusters of water quality parameters [24][25].

2. Methods

2.1 Study Setting

The selected study setting for this study is Laguna Lake which is also known as *Laguna de Bay* the Philippines. The lake is located at 14°22’59.99” N and 121°14’60.00” E [26]. As stated in the 2013 Laguna de Bay Ecosystems Health Report Card led by the Laguna Lake Development Authority (LLDA), the lake is the most expansive body of water located within the borders of the Philippines and ranks as the third largest in the Southeast Asian region [27]. It spans across a surface area of 900 square kilometers, with an average depth of 2.5 meters and an elevation approximately 1 meter higher than sea level. The lake is surrounded by the provinces of Laguna to the east, west, and southwest, Rizal to the north to the northeast, and Metro Manila also known as the National Capital Region (NCR) to the northwest. The lake comprises three separate bays: the West Bay, Central Bay, and East Bay, which meet at the South Bay. The West Bay watershed is characterized by its high population density and extensive urbanization, primarily due to its inclusion of a portion of NCR. In contrast, the East Bay watershed exhibits the lowest population and development levels. Talim Island serves as the geographical barrier that separates the West and Central Bays, making it the largest and most densely inhabited island in the Lake. The sole discharge point of the Lake is the Napindan Channel, which is linked to Manila Bay through the Pasig River [27].

The lake serves multiple purposes including flood control, power generation, recreation, industrial cooling, irrigation, waste disposal, and as a source of potable water. Its primary use is for fisheries and aquaculture, which form the fishing industry of the Locality. Approximately 13,000 fishermen are purportedly financially dependent on the lake. According to the LLDA, further, the lake produces an annual yield of approximately 80,000 to 90% metric tons of fish [28].

The western region is most profitable for capturing fisheries and two types of aquacultures. The constant influx of saline water creates a favorable environment for milkfish to thrive in brackish conditions. Furthermore, the lake sustains fisheries. However, pollution contamination presents a significant risk. Moreover, invasive species pose a threat not only to the ecological stability of the lake but also to the well-being of native species and their economic productivity [28].

2.2 Data to be Used

Overall, 52 monitoring stations are covering the lake. There are 15 Laguna Lake monitoring stations and 37 monitoring stations from its tributaries. In the analysis, however, Station 22b (Tanay River – Midstream) and Station 30 (Binangonan River) were dropped due to multiple missing values within the covered period. The data of BOD in mg/L, DO in mg/L, and FO in MPN/100ml which are openly available were collected from the website of the LLDA E-Library [29]. Figure 1 shows the map of the study setting with the locations of the monitoring stations. Panel A shows the location of Laguna Lake in the Philippines, Panel B shows Laguna Lake with its tributaries in blue lines, and Panel C shows the location of each monitoring station where the data of every water quality parameter was collected.

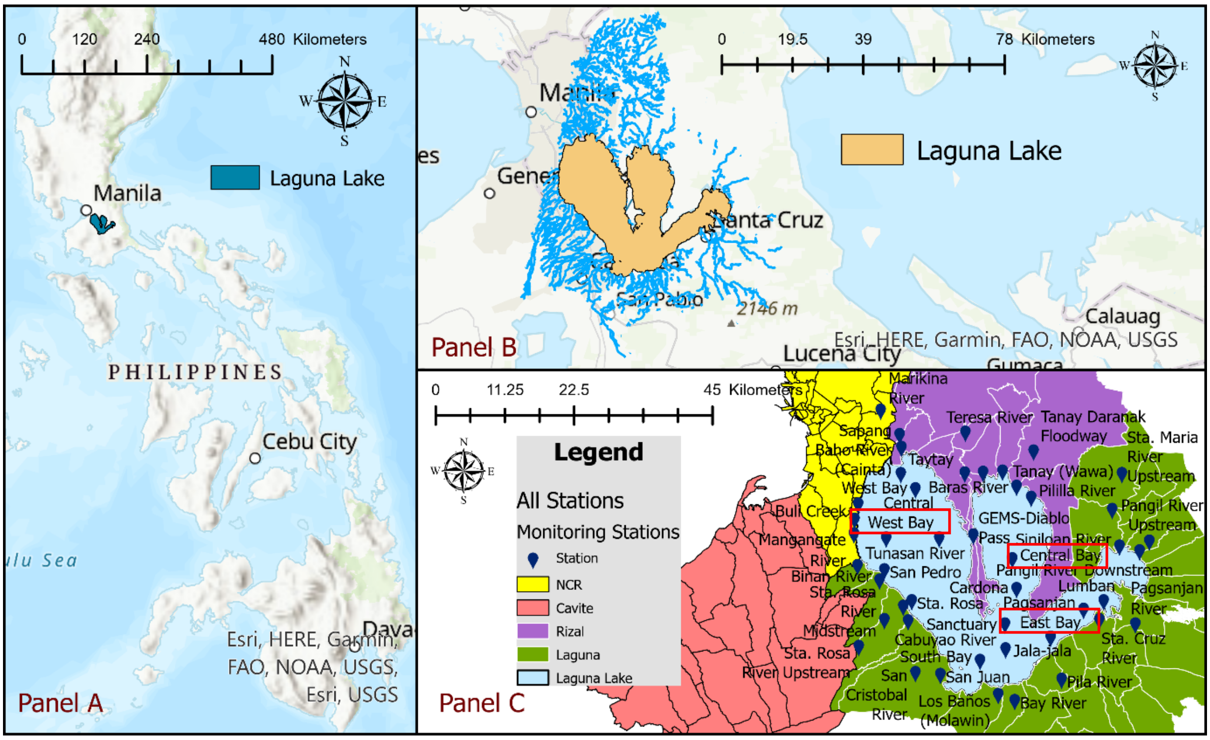


Figure 1: Laguna Lake with Monitoring Stations

The averaged data of December-January-February (DJF) from December 2021 to February 2022, as well as the quarters of March-April-May (MAM), June-July-August (JJA), and September-October-November (SON) in 2022, will be utilized. This was done to potentially accommodate variations in the environment caused by seasonal changes relative to the climatic system of the Philippines. Necessary imputations will be applied. The *mean imputation* of the two months covering missing data, or the *last observation carried forward* imputations were done accordingly.

2.3 Spatial Autocorrelation

The application of statistical techniques in a spatial context presents numerous obstacles. Waldo Tobler, a geographer, and statistician succinctly expressed a fundamental factor that influences the analysis of spatially referenced data in his well-known and often paraphrased *first law of geography*: “Everything is related to everything else, but near things are more related than far things” [30]. This law precisely defines the statistical concept of positive *spatial autocorrelation*, wherein observations taken nearby exhibit greater similarity compared to those taken at greater distances [31]. For this paper, the researchers would like to determine if the water quality parameters per quarter in the monitoring stations nearby with each other are spatially autocorrelated.

The process of determining the presence of significant clusters through spatial data analysis begins by selecting and assigning spatial weights based on contiguity. The queen contiguity criterion was chosen since it specifically deals with the possible errors in the coordinate pairs that define a polygon in the context of GIS [32]. The queen weighting procedure was chosen to measure the relationship between the centroids of the monitoring stations. Herein, a binary weight matrix called $\underline{\Upsilon}$ is established. This matrix defines a specific range of distances between the centroids of the monitoring stations. Certain coverable areas of Laguna Lake may have proximity to one another, but they are not geographically contiguous. The function is defined in Eq. 1, which calculates the level of connectedness between the monitoring stations.

$$c_{u,v} = \begin{cases} 0, \text{ if } \delta_{u,v} > d \\ 1, \text{ otherwise} \end{cases} . \quad (1)$$

It considers the distance $c_{u,v}$ between the centroids of monitoring stations u and v, as well as the minimum distance d required for each monitoring station to have at least one neighbor. When u and v are non-neighboring, $c_{u,v}$ is equal to 0. Else, $c_{u,v}$ is equal to 1 [33].

By utilizing $c_{u,v}$'s, the spatial analysis of the water quality parameters in Laguna Lake can be conducted, allowing for the application of a Global Indicator of Spatial Autocorrelation to ascertain the presence of clustering of outliers in the lake. The widely utilized Global Moran's *I* statistic will be employed to determine if the clustering of the monitoring stations in terms of specific water quality parameters by quarter are random events or not (ESRI 2021), i.e. statistically significant. The null hypothesis for this statistic posits that the observed clustering is a result of random chance. Conversely, the alternative hypothesis posits that the clustering that may have taken place is not random. The Global Moran's *I* exhibit a distinct attribute that deviates from the conventional correlation coefficient. It quantifies the resemblance between adjacent regions. The formula for the Global Moran's *I* [31][33], plugging in the concepts of Eq. 1 is given in Eq. 2.

$$I = \left[\sum_{u=1}^{A}\sum_{v=1}^{A} c_{u,v}\left(G_u - \bar{G}\right)\left(G_v - \bar{G}\right)\right] \Big/ s^2 \sum_{u=1}^{A}\sum_{v=1}^{A} c_{u,v}, \quad (2)$$

where $\bar{G} = \frac{1}{A}\sum_{u=1}^{A} G_u$ and $s^2 = \frac{1}{A}\sum_{u=1}^{A}(G_u - \bar{G})^2$. The total number of monitoring stations is denoted by A in eq. 2 and thus, $A = 50$. An explicitly defined water quality parameter, given a quarter, on the other hand, is represented by G statistic for every monitoring station u and v whereas $\bar{G}$ and s^2, in succession, are the mean and variance of each water quality parameter per quarter.

The Local Indicator of Spatial Autocorrelation (LISA) was employed to identify potential clusters. The statistic specifically employed was the standardized Local Moran's *I* [31][33]. The indicator, represented as I_u^z with s as the standard deviation of poverty incidence across the region, is calculated as follows:

$$I_u^z = \tfrac{1}{s}\left(G_u - \bar{G}\right)\left[\sum_{v}^{A} \tfrac{c_{u,v}}{s}\left(G_v - \bar{G}\right)\right].$$

2.4 Hot Spots and Cold Spots

The study will also apply Optimized Hot Spot Analysis. The implementation will specifically focus on the Getis-Ord G_i^* (G-I star) Analysis [34]. The G_i^* statistic is a means of quantifying the level of clustering of an attribute in different locations [35]. This statistic enables the identification of monitoring stations as either hot spots or cold spots with varying levels of confidence. In the context of this study, a *hot spot* of a water quality parameter is a monitoring station with extremely high values given a quarter that is clustered, while a *cold spot* of a water quality parameter is defined as a monitoring station with extremely low values. Furthermore, a hot spot is empirically identified when the observed values surpass the expected values, while a cold spot is statistically detected when the observed values fall below the expected values [36]. A statistically significant hot spot or cold spot is dependent and relative to the values of the neighboring monitoring stations. Eq. 3 provides a standardized value for the statistic [34][33].

$$G_i^* = \left(\sum_{v=1}^{A} \omega_{u,v} q_u - \bar{Q}\sum_{v=1}^{A} \omega_{u,v}\right) \Big/ \left\{\sigma_q \sqrt{\frac{1}{A-1}\left[\left(A\sum_{v=1}^{A} \omega_{u,v}^2\right) - \left(\sum_{v=1}^{A} \omega_{u,v}^2\right)\right]}\right\}. \quad (3)$$

In this paper, q_v is the value of the water quality parameter of interest in a monitoring station v with its corresponding spatial weight given by $\omega_{u,v}$ between monitoring stations u and v. A denotes the total number of monitoring stations. The mean value of a water quality parameter is given by $\bar{Q}$ and takes the value of $\bar{Q} = \frac{1}{A}\sum_{v=1}^{A} q_v$. On the other hand, σ_q denotes the standard deviation which is given by $\sigma_q = \sqrt{\left(\frac{1}{A}\sum_{v=1}^{A} q_v^2\right) - \bar{Q}^2}$.

2.4 Kriging Method to be Applied
To estimate the spatial distribution of BOD, DO, and FO per quarter, this study will specifically focus on the utilization of Empirical Bayesian Kriging (EBK). EBK is a spatial interpolation technique that necessitates minimal interactive modeling and is distinguished by its robustness and simplicity. This geostatistical interpolation method allows for automated kriging processing. The EBK automatically calculates the parameters using simulation and sub-setting, instead of manually adjusting the settings, to produce *accurate* results. EBK has a distinct advantage over other kriging techniques in its capacity to calibrate errors resulting from the estimation of the underlying semi-variogram [37]. Furthermore, EBK can predict the errors linked to the generated predicted values as well as the values of regions that have not been sampled. The variogram parameter is simulated numerous times. Subsequently, the variogram models are computed based on the simulated data. The utilization of this procedure enhances the accuracy of EBK prediction compared to traditional kriging methods [38][39]. Furthermore, empirical evidence has shown that EBK can produce precise forecasts for data that are nonstationary and non-Gaussian. This includes scenarios where data variations across different locations are not uniform, thereby establishing EBK as a dependable automated interpolation technique [40][39]. Other kriging methods may struggle to provide accurate forecasts due to the limited number of weather stations and the vast geographical area. The EBK will be implemented accordingly, therefore. The EBK framework [39] is presented in Eq. 4.

$$\sum_{a=1}^{A} \pi_a^{\kappa} \Theta(\lambda_a) + \sum_{a=1}^{A} \omega_a \mathrm{I}_m(\lambda_a) = P_m(\lambda_0) < \mathrm{M}_m \qquad (4)$$

In Eq. 4, M is the interpolated value, m is the parameter, and λ_0 is the critical value of the specified water quality parameter given a target quarter. A monitoring station in this study is represented by λ_a and the total number of locations is denoted by A. The weight provided to a known monitoring station is ω_a. On the other hand, the I_m is a multiplier could be 0 if it is higher than, or 1 if lower than P_m. The π_a^{κ} component is the kriging weight represented which is estimated based on the cross-variogram between components in Eq. 2 and Eq. 3 given below.

$$\Theta(\lambda) = \phi / a \text{ , and} \qquad (5)$$

$$\mathrm{I}_m(\lambda, m) = \begin{cases} 0 \text{ if } \lambda(\lambda) \geq P_m \\ 1 \text{ otherwise} \end{cases}. \qquad (6)$$

In Eq. 5, ϕ denotes the ϕ^{th} order statistics for a water quality parameter observed at the location λ [38][39]. This makes the estimation process for the regions without values simple [41][38]. The process begins with an initial estimation of a semi-variogram using existing data. This is then followed by simulating a new value based on the calculated variogram. Subsequently, a new semi-variogram is recalculated using the recently generated data, and the updated weight of the variogram is determined based on Bayes' Rule [41][38].

2.5 Procedures
All maps and analyses shall be generated in ArcGIS 3.1.2. The (a) Global Moran's *I*, (b) Getis-Ord G_i^*, and EBK shall be generated separately for each quality parameter by quarter. Any Global Moran's *I* that showed significant results shall entail running Local Moran's *I* analyses with maps that will be overlaid to the corresponding EBK maps. On the other hand, significant Getis-Ord G_i^* results shall be compared to the next quarter. i.e., comparisons of the DJF results with the MAM results, the MAM results with the JJA results, and the JJA results with the SON results shall be presented visually. All applicable analyses shall be replicated by 9,999 to achieve robustness of estimates.

3. Results and Discussion

3.1 Preliminaries for the Results
The results of both the Global Moran's *I*, Local Moran's *I*, Getis-Ord G_i^* all showed p-values of less than 0.05. Therefore, the intended procedures were done correspondingly. Table 1 is provided for ease of interpretation of the used abbreviations in the produced individual maps. Further, the three distinct bays, namely West Bay, Central Bay, and East Bay, as seen also in Panel C of Figure 1, shall be used for easier interpretability.

Table 1: Abbreviations used in the maps

Abbreviation	Definition	Abbreviation	Definition
BODDJF	BOD for DJF	BODDJF EBK	EBK of BOD for DJF
BODMAM	BOD for MAM	BODMAM EBK	EBK of BOD for MAM
BODJJA	BOD for JJA	BODJJA EBK	EBK of BOD for JJA
BODSON	BOD for SON	BODSON EBK	EBK of BOD for SON
DODJF	DO for DJF	DODJF EBK	EBK of DO for DJF
DOMAM	DO for MAM	DOMAM EBK	EBK of DO for MAM
DODJJA	DO for JJA	DOJJA EBK	EBK of DO for JJA
DODSON	DO for SON	DOSON EBK	EBK of DO for SON
FCDJF	FC for DJF	FCDJF EBK	EBK of FC for DJF
FCMAM	FC for MAM	FCMAM EBK	EBK of FC for MAM
FCDJJA	FC for JJA	FCJJA EBK	EBK of FC for JJA
FCDSON	FC for SON	FCSON EBK	EBK of FC for SON

3.2 Spatial Autocorrelation and Interpolation

Figure 2, Figure 3, and Figure 4 shows the GIS of (a) the values in a 10-level white-colored graduated circle, where the largest circles denote the location of a monitoring station with the largest range of values; (b) the nearby areas of Laguna Lake also known as *Laguna de Bay*, (c) contours of the estimated values based on the EBK, and (d) results of the Local Moran's *I*. In Figure 2, Figure 3, and Figure 4, a red circle denotes a monitoring station that exhibits a high value of a specific water quality parameter that is nearby to at least one other monitoring station that has a high value of water quality parameter as well for such quarter, and thus giving rise to the term High-High cluster. Low-Low in dark blue circles, Low-High in light blue circles, and High-Low in pink circles are all monitoring stations characterized by the same context, whereas circles with × marks are stations that have no significant clustering. Given that one of the objectives of this study is to identify clusters of monitoring stations with significant clustering of water quality parameters, monitoring stations that are categorized as High-High and Low-High were prioritized.

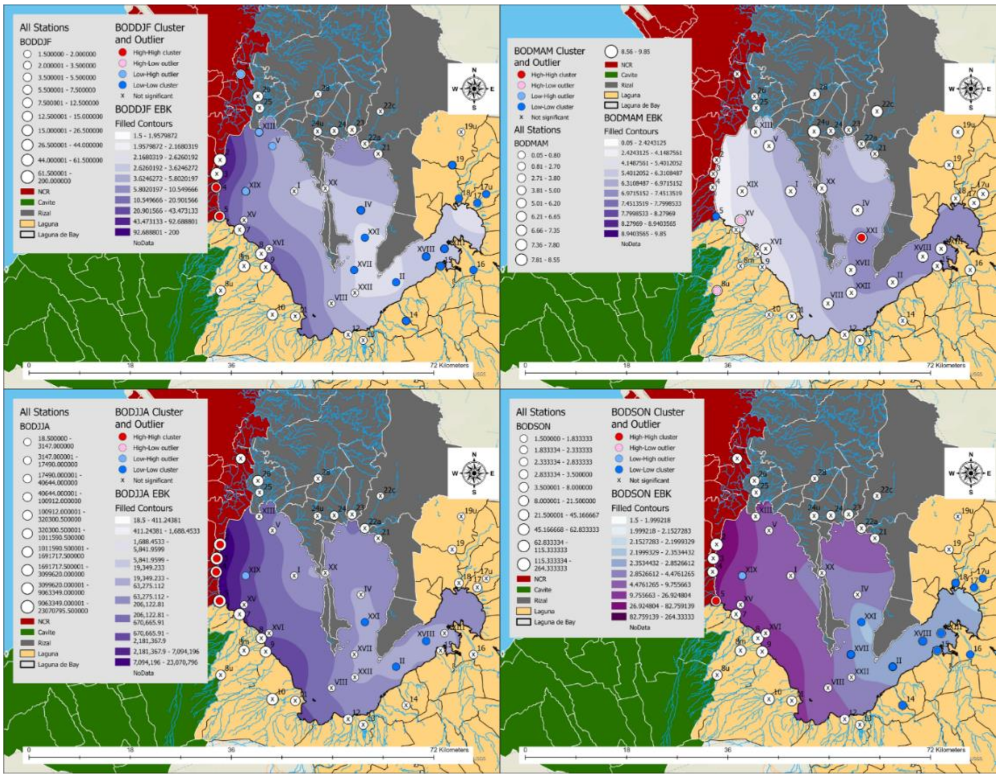


Figure 2: Map of BOD by quarter

Figure 2 illustrates the presence of clustered groups of high values of BOD in the West Bay, which is adjacent to the NCR, in DJF and JJA. The highest top EBK range was in JJA which was from 7.09 million mg/L to 23.07 million mg/L and lowest top EBK range was on MAM with 8.94 mg/L to 9.85 mg/L. Higher BOD levels are linked to the presence of organic pollution, specifically sewage [27]. The West Bay watershed is characterized by its high population density and extensive urbanization, primarily due to its inclusion of a portion of the NCR [27]. In contrast, the East Bay watershed has the lowest population and development levels [27]. The high BOD in the monitoring stations near the NCR is a glaring problem. Figure 2 also shows that except for the MAM quarter, the EBK in the maps shows the concentration of high BOD in the West Bay except for the warmer months. These consistently elevated levels of BOD situation may further deteriorate since ultimately, all sewage effluent from the city's waterways eventually flows into the Laguna Lake or Manila Bay [42].

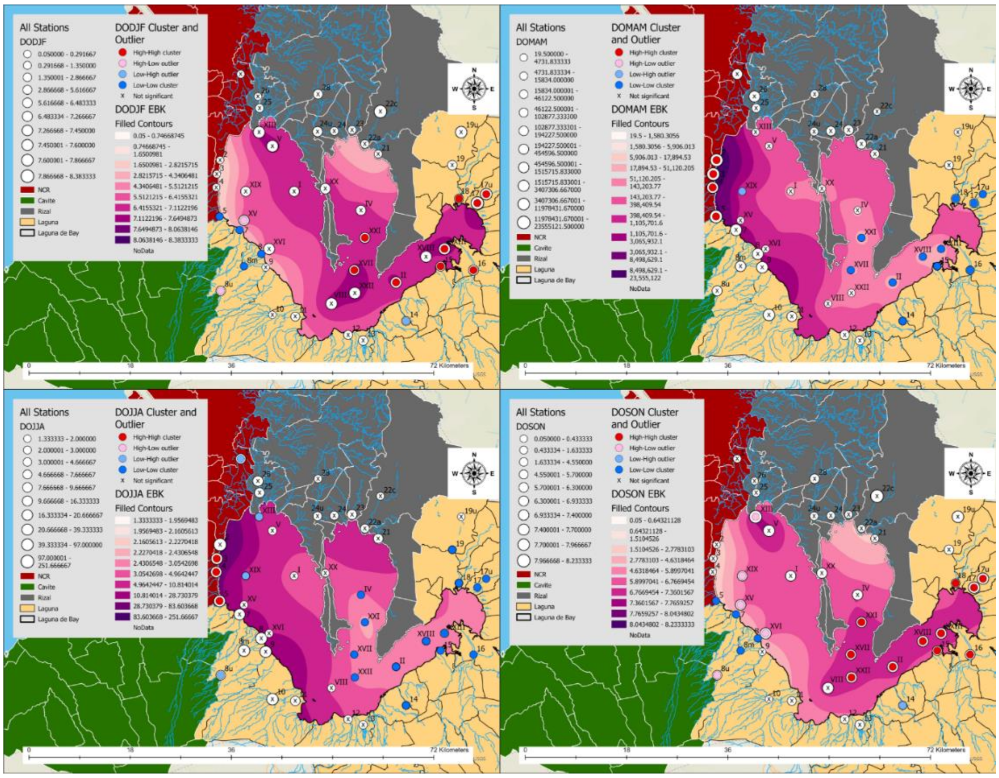


Figure 3: Map of DO by quarter

Figure 3 shows an interesting pattern for the DO, at least in the duration of the time frame of the used data. The DJF and SON, the colder seasons primarily, have clusters of monitoring stations with high levels of dissolved oxygen in the Central Bay and East Bay. This aligns with the assertion that warm water has a lower capacity to retain DO compared to cold water [43], relative to the climatic seasonality of the NCR. The NCR is classified under the Köppen climate classification as having a combination of tropical wet and dry climate and tropical monsoon climate. The dry season in Metro Manila spans from January to May, while the wet season extends from June to December [44]. Conversely, the monitoring stations in the West Bay near the NCR have high DO levels in the MAM which are generally the warmer months, and JJA where seasonal rains are prevalent. The highest range of EBK for DO was in MAM which is from around 8.50 million to 23.56 million mg/L. The other quarters have ranges that peaked only at 83.60 to 251.67 mg/L in JJA based on the generated EBK.

The East Bay region has a greater density of fishermen operating within a relatively small fishing area, with a fishing ground allocation of only 1 fisherman per 28 hectares. Additionally, it is characterized by the highest concentration of the invasive clown knife fish. This species was introduced into the Lake via the East Bay and likely proliferated more rapidly due to the superior water quality of the East Bay [27]. On the other hand, Central Bay exhibits the highest proportion of indigenous fish in its catch composition and zooplankton ratio. The fishing ground allocation, moreover, is approximately 1 fisher per 110 hectares. [27].

Previous studies have consistently reported a correlation between fecal coliform concentrations and levels of phosphorus and nitrogen [45][46][47] [48]. A study revealed the correlation as well between elevated levels of phosphates and reduced concentrations of fecal coliform. This implies that fecal coliform bacteria may have a greater affinity for feeding on phosphates compared to nitrogen. Consequently, high levels of phosphates could indicate absence or low concentration of fecal coliform bacteria [48]. It seems that the same characteristics apply for the fecal coliform concentrations in Figure 4. The province of Laguna is a top rice-producing province and thus requires a huge volume of fertilizers. Such an economic feature is observable in JJA. This may have been apparent in the generated results. The upper limit of the EBK range for DO was observed in MAM, ranging from approximately 8.50 million to 23.56 million mg/L. In the other quarters, the highest recorded values ranged from 83.60 to 251.67 mg/L during the JJA quarter, as determined by the EBK.

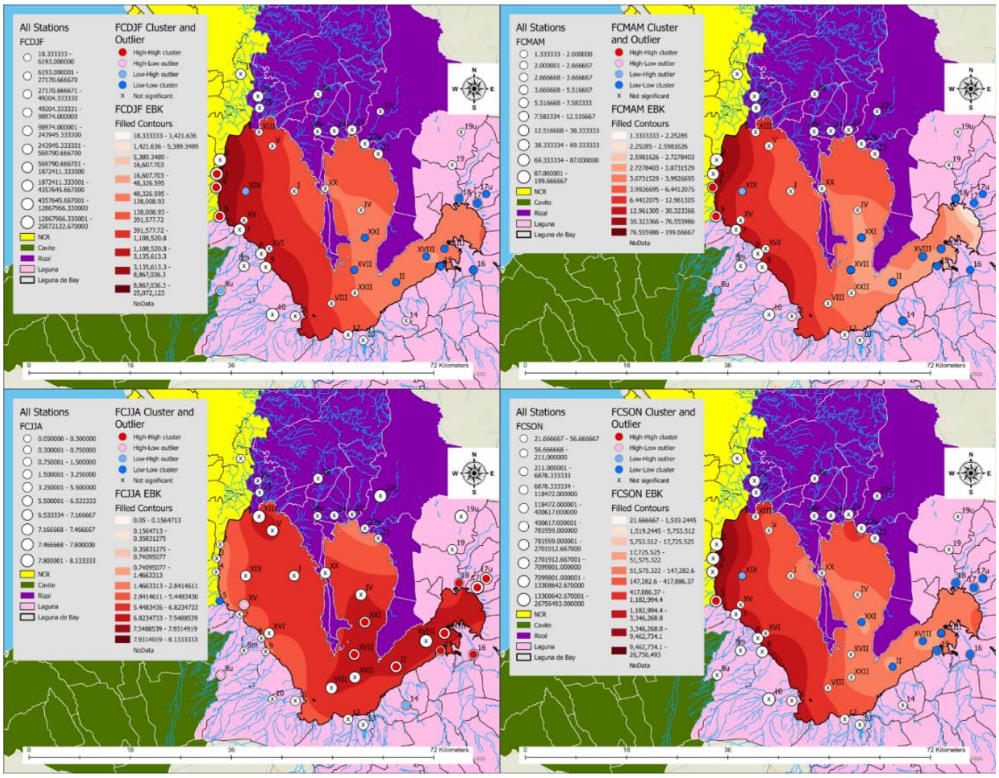


Figure 4: Map of FC by quarter

Fecal coliform contamination may originate from sewage treatment facilities or certain industrial sources. Frequently, it originates from numerous minor sources, with each one making a small contribution to the overall problem. Every individual residing within a watershed plays a role in the creation of these minor origins through their management of on-site septic systems, livestock waste, pet waste, and sanitary practices [49]. Unsurprisingly, the West Bay, which is highly populated and industrialized, has at least one monitoring station with High-High fecal coliform concentration. Regardless of the origin, an increase in the concentration of fecal coliform bacteria in water directly correlates with an elevated risk to public health [49].

3.3 Comparison of Spots

With different levels of confidence, the results of the Getis-Ord G_i^* were all significant. Appendix 1, Appendix 2, and Appendix 3 show the changes in quarters of each of the spots by water quality parameter. In general, the West Bay, particularly the monitoring stations near the NCR, are alternating into a Hot Spot, or Not Significant. Since this comes anticipatedly, the consistency of EBK, Moran's *I*, and Getis-Ord G_i^* entails immediate actions to conserve Laguna Lake.

4. Conclusion and Recommendations

This paper demonstrates the advantages of utilizing data visualization and geostatistical analyses for monitoring the water quality parameters of Laguna Lake. The study demonstrated that employing quarterly seasonality based on the climate in the Philippines could yield more comprehension of the association between environmental changes and the fluctuations in the common water quality parameters obtained from the lake. Preliminary findings following the lifting of COVID-19 lockdowns show consistency of previous results from the literature. The burgeoning population and industrialization of the NCR present significant risks to the ecosystem and economic benefits of the lake. More studies can be conducted to establish more conclusive scientific generalizations especially with the inclusion of socioeconomic, climatic, chemical, and physical factors, when applicable.

Although the findings align with previous studies, this paper introduces two new areas for future research: (a) identifying the specific locations with high concentrations of water quality parameters, and (b) examining the impact of different seasonal quarters in the Philippine climate. This could potentially contribute to achieving the targets outlined in Sustainable Development Goal 6, as well as anticipatory actions in response to the impacts of climate change.

The stakeholders, economic managers, and health officials responsible for maintaining or regulating Laguna Lake can utilize this study. With additional conclusive evidence, proactive measures can be taken to prepare for the potential decline in the fish population during warmer seasons. This has a significant impact on the economic and food security of the population that relies on the lake. Areas near the NCR frequently exhibit poor levels of water quality parameters in general. The rapid and ongoing urbanization in the region may ultimately present significant risks to human health if no decisive intervention is undertaken, particularly due to the potential long-term accumulation of contaminants in the lake. The GIS also enhances the ability to interpret the clustering of high or low values of the water quality parameters.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] United Nations. (2021). *Summary progress update 2021: SDG 6 - water and sanitation for all*. UN. https://www.unwater.org/publications/summary-progress-update-2021-sdg-6-water-and-sanitation-all

[2] WHO. (2016). *Water Quality and Health Strategy 2013-2020*. World Health Organization. https://www.who.int/publications/m/item/water-quality-and-health-strategy-2013-2020

[3] Oppedal, F., Dempster, T., & Stien, L. H. (2011). Environmental drivers of Atlantic salmon behaviour in sea-cages: A Review. *Aquaculture*, *311*(1–4), 1–18. https://doi.org/10.1016/j.aquaculture.2010.11.020

[4] Seager, J., Milne, I., Mallett, M., & Sims, I. (2000). Effects of short-term oxygen depletion on fish. *Environmental Toxicology and Chemistry*, *19*(12), 2937–2942. https://doi.org/10.1002/etc.5620191214

[5] Cerezo Valverde, J., Martínez López, F.-J., & García García, B. (2006). Oxygen consumption and ventilatory frequency responses to gradual hypoxia in common dentex (Dentex Dentex): Basis for suitable oxygen level estimations. *Aquaculture*, *256*(1–4), 542–551. https://doi.org/10.1016/j.aquaculture.2006.02.030

[6] Sun, L., Wang, B., Yang, P., Wang, X., Li, D., & Wang, J. (2022b). Water quality parameter analysis model based on fish behavior. *Computers and Electronics in Agriculture*, *203*, 107500. https://doi.org/10.1016/j.compag.2022.107500

[7] Bashir, I., Lone, F. A., Bhat, R. A., Mir, S. A., Dar, Z. A., & Dar, S. A. (1970, January 1). *Concerns and threats of contamination on aquatic ecosystems*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-030-35691-0_1

[8] Moxey, A. (2012). *Agriculture and Water Quality: Monetary Costs and Benefits across OECD Countries*. OECD. https://www.oecd.org/greengrowth/sustainable-agriculture/49841343.pdf

[9] European Environment Agency. (2018). *The European Pollutant Release and Transfer Register (E-PRTR)*. https://environment.ec.europa.eu/topics/industrial-emissions-and-safety/european-pollutant-release-and-transfer-register-e-prtr_en

[10] Vigiak, O., Grizzetti, B., Udias-Moinelo, A., Zanni, M., Dorati, C., Bouraoui, F., & Pistocchi, A. (2019). Predicting biochemical oxygen demand in European freshwater bodies. *Science of The Total Environment*, *666*, 1089–1105. https://doi.org/10.1016/j.scitotenv.2019.02.252

[11] Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., & Davies, P. M. (2010). Global threats to human water security and River Biodiversity. *Nature*, *467*(7315), 555–561. https://doi.org/10.1038/nature09440

[12] Wen, Y., Schoups, G., & van de Giesen, N. (2018). Global impacts of the meat trade on in-stream organic river pollution: The importance of spatially distributed hydrological conditions. *Environmental Research Letters*, *13*(1), 014013. https://doi.org/10.1088/1748-9326/aa94f6

[13] Ferreira, A. R. L., Sanches Fernandes, L. F., Cortes, R. M. V., & Pacheco, F. A. L. (2017). Assessing anthropogenic impacts on riverine ecosystems using nested partial least squares regression. *Science of The Total Environment*, *583*, 466–477. https://doi.org/10.1016/j.scitotenv.2017.01.106

[14] Timmons, M. B. (2002). *Recirculating Aquaculture Systems*. Cayuga Aqua Ventures.

[15] Bagherzadeh Lakani, F., Sattari, M., & Falahatkar, B. (2012). *Effect of different oxygen levels on growth performance, stress response and oxygen consumption in two weight groups of Great Sturgeon Huso Huso*. Iranian Journal of Fisheries Sciences. https://jifro.areeo.ac.ir/article_114297.html

[16] Ritola, O., Tossavainen, K., Kiuru, T., Lindstrom-Seppa, P., & Molsa, H. (2002). Effects of continuous and episodic hyperoxia on stress and hepatic glutathione levels in one-Summer-old rainbow trout (Oncorhynchus mykiss). *Journal of Applied Ichthyology*, *18*(3), 159–164. https://doi.org/10.1046/j.1439-0426.2002.00324.x

[17] Pichavant, K., Person-Le-Ruyet, J., Bayon, N. L., Severe, A., Roux, A. L., & Boeuf, G. (2001). Comparative effects of long-term hypoxia on growth, feeding and oxygen consumption in juvenile Turbot and European sea bass. *Journal of Fish Biology*, *59*(4), 875–883. https://doi.org/10.1111/j.1095-8649.2001.tb00158.x

[18] Jobling, M. (1994). *Fish Bioenergetics*. SpringerLink. https://link.springer.com/book/9780412580901

[19] Samaras, A., Tsoukali, P., Katsika, L., Pavlidis, M., & Papadakis, I. E. (2023). Chronic impact of exposure to low dissolved oxygen on the physiology of Dicentrarchus labrax and Sparus aurata and its effects on the acute stress response. *Aquaculture*, *562*, 738830. https://doi.org/10.1016/j.aquaculture.2022.738830

[20] Madigan, M. T., Martinko, J. M., Stahl, D. A., & Clark, D. P. (2012). *Brock Biology of Microorganisms*. B. Cummings.

[21] Raña, J., Domingo, J., Opinion, A. G., & Cambia, F. (2017). Contamination of coliform bacteria in water and fishery resources in Manila Bay Aquaculture Farms. *The Philippine Journal of Fisheries*, *24*(2), 98–126. https://doi.org/10.31398/tpjf/24.2.2016a0015

[22] Forsythe, S. J. (2010). *The Microbiology of Safe Food, 2nd edition*. John Wiley & Sons.

[23] Henze, M. (2008). *Biological wastewater treatment: Principles, modelling and Design*. IWA Pub.

[24] Yuan, F., Lu, L., Zhang, Y., Wang, S., & Cai, Y.-D. (2018). Data mining of the cancer-related lncrnas GO terms and KEGG pathways by using MRMR method. *Mathematical Biosciences*, *304*, 1–8. https://doi.org/10.1016/j.mbs.2018.08.001

[25] Sun, L., Wang, B., Yang, P., Wang, X., Li, D., & Wang, J. (2022a). Water quality parameter analysis model based on fish behavior. *Computers and Electronics in Agriculture*, *203*, 107500. https://doi.org/10.1016/j.compag.2022.107500

[26] Latitude.to. (n.d.). *GPS coordinates of Laguna de Bay, Philippines. latitude: 14.3833 longitude: 121.2500*. Latitude.to, maps, geolocated articles, latitude longitude coordinate conversion. https://latitude.to/articles-by-country/ph/philippines/14491/laguna-de-bay

[27] LLDA. (2016). *Laguna de Bay 2013 Ecosystem health report card*. PEMSEA. https://pemsea.org/publications/reports/laguna-de-bay-2013-ecosystem-health-report-card

[28] LLDA. (n.d.-a). *Existing lake uses | LLDA*. Laguna Lake Development Authority. https://llda.gov.ph/existing-lake-uses/

[29] LLDA. (n.d.-b). *Laguna de Bay and its tributaries | LLDA*. Laguna Lake Development Authority. https://llda.gov.ph/ldb-and-its-tributaries/

[30] Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, *46*, 234. https://doi.org/10.2307/143141

[31] Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons.

[32] Anselin, L. (2020). *Geoda*. Contiguity-Based Spatial Weights. https://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html

[33] Pantolla, H. G., & Nacion, N. A. (2023). Spatial Analysis of Poverty Incidence and Road Networks in Eastern Visayas Region, Philippines. *Philippine Journal of Science*, *152*(4). https://philjournalsci.dost.gov.ph/publication/regular-issues/past-issues/122-vol-152-no-4-august-

2023/1872-spatial-analysis-of-poverty-incidence-and-road-networks-in-eastern-visayas-region-philippines

[34] Getis, A., & Ord, J. K. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, *27*(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

[35] Ghodousi, M., Sadeghi-Niaraki, A., Rabiee, F., & Choi, S.-M. (2020). Spatial-temporal analysis of point distribution pattern of schools using spatial autocorrelation indices in Bojnourd City. *Sustainability*, *12*(18), 7755. https://doi.org/10.3390/su12187755

[36] Gorr, W. L., Kurland, K. S., & Dodson, Z. M. (2018). *GIS tutorial for crime analysis*. Esri Press.

[37] Ali, G., Sajjad, M., Kanwal, S., Xiao, T., Khalid, S., Shoaib, F., & Gul, H. N. (2021). Spatial–temporal characterization of rainfall in Pakistan during the past half-century (1961–2020). *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-86412-x

[38] Sahu, B., Ghosh, A. K., & Seema. (2021). Deterministic and geostatistical models for predicting soil organic carbon in a 60 ha farm on Inceptisol in Varanasi, India. *Geoderma Regional*, *26*. https://doi.org/10.1016/j.geodrs.2021.e00413

[39] Yang, R., & Xing, B. (2021). A comparison of the performance of different interpolation methods in replicating rainfall magnitudes under different climatic conditions in Chongqing Province (China). *Atmosphere*, *12*(10), 1318. https://doi.org/10.3390/atmos12101318

[40] Krivoruchko, K., & Gribov, A. (2013). *Pragmatic Bayesian Kriging for non-stationary and moderately Non-Gaussian Data*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-642-32408-6_15

[41] Rajesh, R., Elango, L., & Brindha, K. (2019). Methods for assessing the groundwater quality. *GIS and Geostatistical Techniques for Groundwater Science*, 57–78. https://doi.org/10.1016/b978-0-12-815413-7.00006-7

[42] Jalilov, S.-M. (2017). Value of Clean Water Resources: Estimating the Water Quality Improvement in Metro Manila, Philippines. *Resources*, *7*(1), 1. https://doi.org/10.3390/resources7010001

[43] United States Geological Survey. (2018). *Dissolved Oxygen and Water*. Dissolved Oxygen and Water | U.S. Geological Survey. https://www.usgs.gov/special-topics/water-science-school/science/dissolved-oxygen-and-water#:~:text=Cold%20water%20can%20hold%20more,oxygen%20concentration%20is%20often%20lower

[44] Environmental Management Bureau. (n.d.). *Regional profile | EMB - National Capital Region*. https://ncr.emb.gov.ph/regionalprofile/

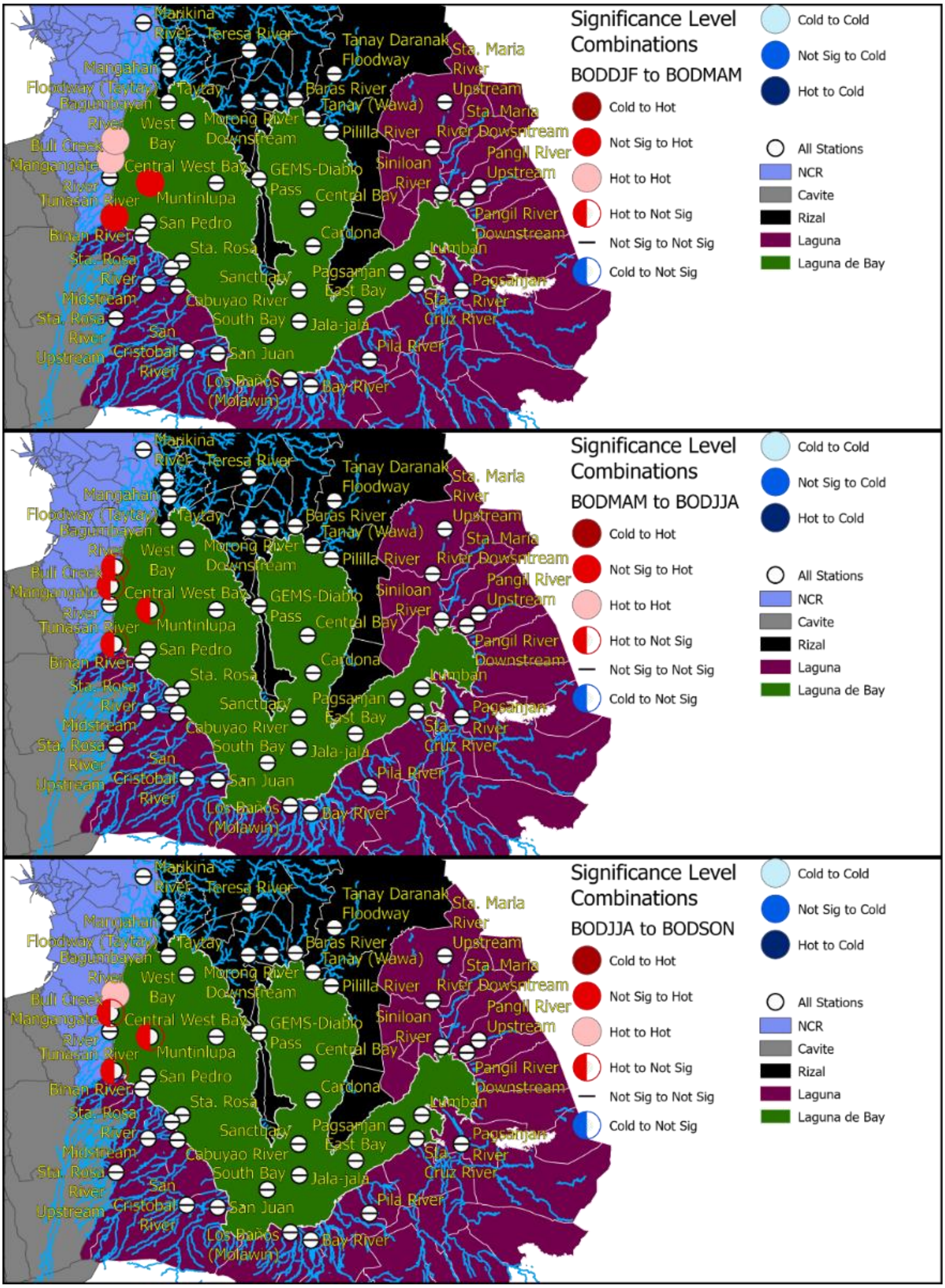
[45] Bolster, C. H., Bromley, J. M., & Jones, S. H. (2005). Recovery of chlorine-exposed *escherichia coli* in estuarine Microcosms. *Environmental Science & Technology*, *39*(9), 3083–3089. https://doi.org/10.1021/es048643s

[46] Cabral, J. P., & Marques, C. (2006). Faecal coliform bacteria in Febros River (northwest Portugal): Temporal variation, correlation with water parameters, and species identification. *Environmental Monitoring and Assessment*, *118*(1–3), 21–36. https://doi.org/10.1007/s10661-006-0771-8

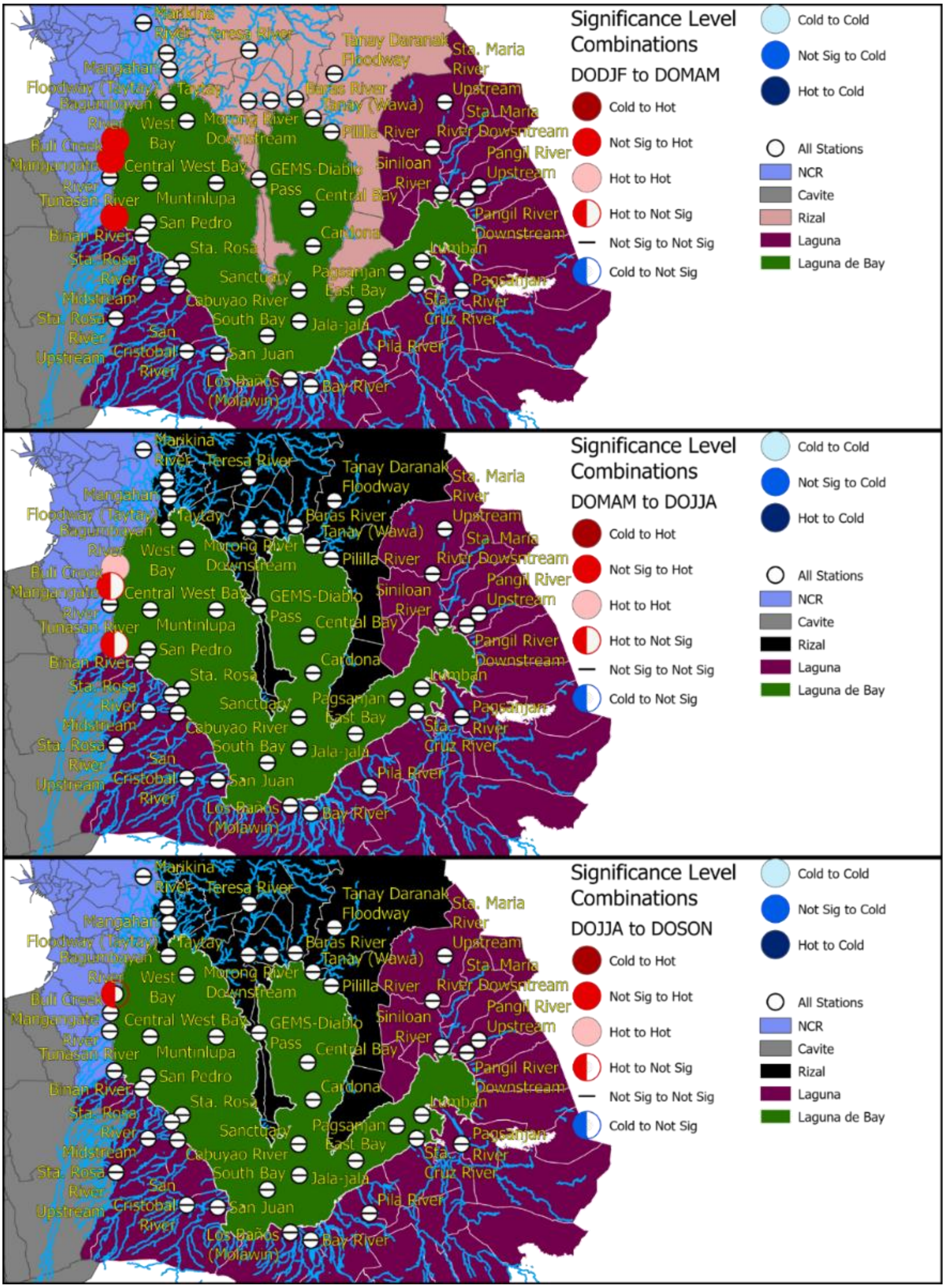
[47] Soo, C.-L., Ling, T.-Y., Lee, N., & Apun, K. (2014). Assessment of the characteristic of nutrients, total metals, and fecal coliform in Sibu Laut River, Sarawak, Malaysia. *Applied Water Science*, *6*(1), 77–96. https://doi.org/10.1007/s13201-014-0205-7

[48] Aram, S. A., Saalidong, B. M., & Osei Lartey, P. (2021). Comparative assessment of the relationship between coliform bacteria and water geochemistry in surface and ground water systems. *PLOS ONE*, *16*(9). https://doi.org/10.1371/journal.pone.0257715

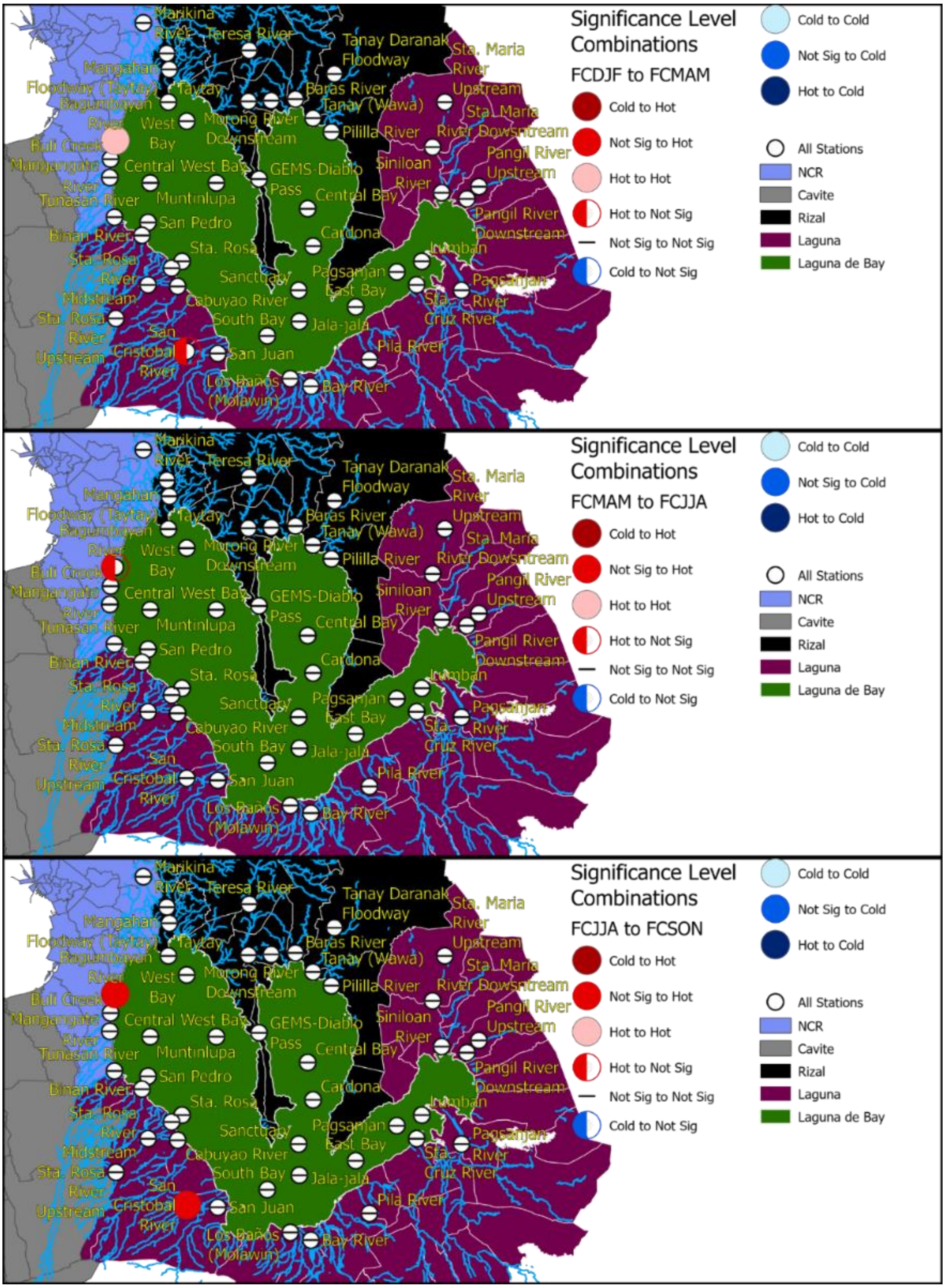
[49] Washington State Department of Ecology. (2005). *Issue what's so important about fecal coliform bacteria? - Washington*. https://apps.ecology.wa.gov/publications/documents/0210010.pdf



Appendix 1: Map comparing Spots of BOD



Appendix 2: Map comparing Spots of DO



Appendix 3: Map comparing Spots of FC