International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

Linear Regression based Water Quality Analysis through Spatial Distribution of Parameters for Vembanad Lake

K A Mohandas[1], J Brema[2]

[1]*Research Scholar, Karunya Institute of Technology and Sciences, Coimbatore*
[2]*Professor, Department of Civil Engineering, Karunya Institute of Technology and Sciences, Coimbatore*

Email: [1]arackaalmohandass@yahoo.com

Abstract: Vembanad Lake is a Ramsar sitethat stretches about 80 kms and covers an area of 200 Sq.kmslocated at Alleppey, Kottayam and Eranakulam districts of Kerala, India. The quality of water of this lake has to be analysed as it serves almost 55,000 ha of agricultural land. Water quality can be analysed through parameters viz. pH, turbidity, BOD, TDS, TSS etc. In this research paper, the watershed map of the study area has been created and the land is classified. The spatial distribution of water quality parameters of the Vembanad Lake has been plotted. A linear regression model has been created from the spectral reflectance values of these parameters and were analyzed. The study determined the parts of the lakethat are more contaminated than others. This study and the corresponding analysis will be beneficial in implementing steps to improve water quality and manage waste outflow into the lake.

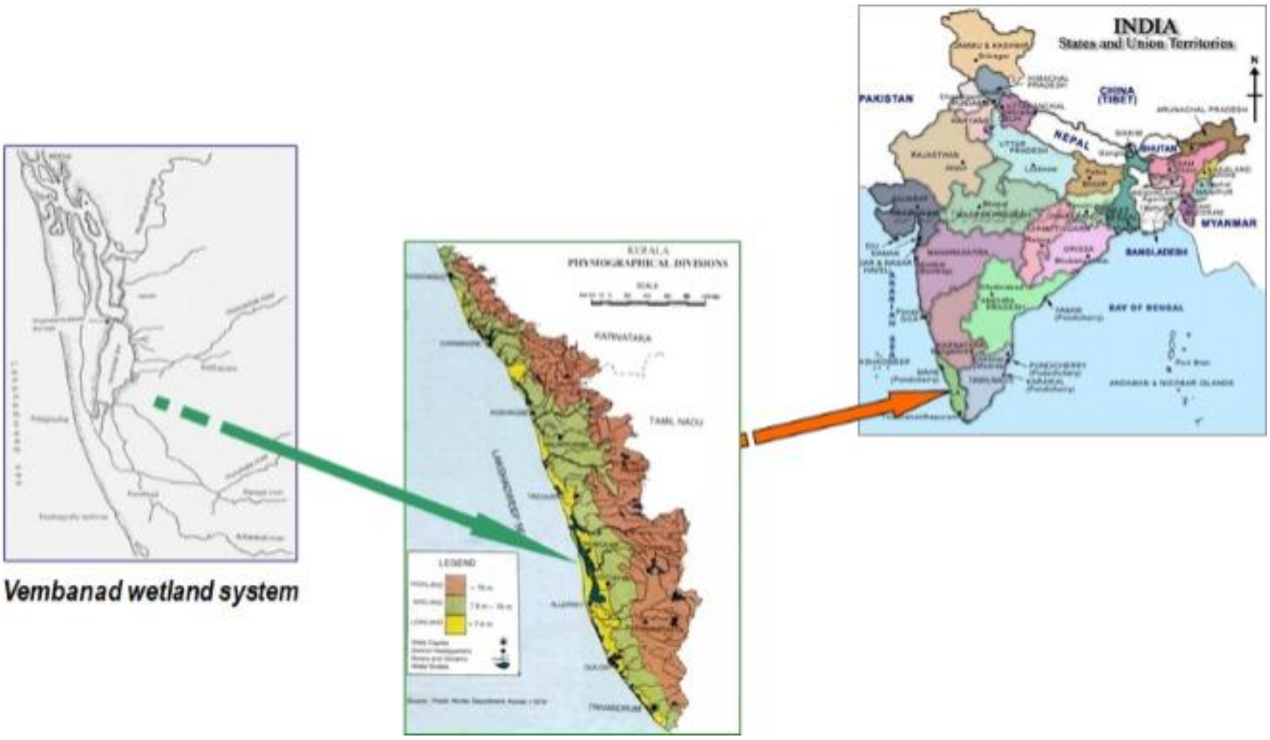
Keywords: Vembanad Lake, Spatial Distribution, Linear Regression, Water Quality, Spectral Reflectance

1. INTRODUCTION

Water quality testing is an essential component of environmental monitoring. Poor water quality has an effect on the whole environment. It is influenced by a number of factors, which can be investigated using advances in remote sensing and geographic information systems (GIS) (Charan Kumar et al 2020). Lakes play a major role in human habitation as they are major sources of rain water harvesting systems. Water quality of lakes are declining day by day as a result of climate change, population growth, inappropriate resource use, lack of long-term management practices and anthropogenic behavior.Many methods are available to analyze the water quality. GIS ease the way of analyzing the quality of water with the help of images obtained from satellites and through remote sensing. With the help of GIS and remote sensing system, the quality of lake water can be assessed easily. From these systems, spatial distribution of the water quality parameters can be plotted. Spatial graph analysis was first used by John Snow in 1854, in order to analyze the spread of cholera disease (Kobayashi, 1994). Gyananath et al (2001) utilized spatial graph algorithm and assessed the environmental parameters of the ground water quality. Dormann et al (2007) reviewed the various spatial regression methods and classified them into three categories in which they could be modeled. The classification is based on the effect of spatial model which includes covariate predictors, error terms and response of the variables replacement by transforming the original data. Spatial autocorrelation was by Fan C and Myint (2014) used to measure the landscape metrics in urban landscape planning. In this research work, spatial distribution analysis has been carried out to study the water quality analysis of Vembanad Lake, a Ramsar Site of Kerala State, India.

2. STUDY AREA

Vembanad Lake, is linked to drainage basins in the humid tropical area between 09°00' - 10°40'N and 76°00'-77°30'E. In November 2002, it was designated as a Ramsar Site, i.e., wetland of international significance due to its special geology, physiography, hydrology, climate, land use fauna, and flora. Monsoon-fed rivers connect Vembanad and are short, fast-flowing, and wide. Backwater and connecting Kol lands are part of the lake and its wetland. The lake covers about 24,000 hectares of wetlands, accounting for more than half of the state's total backwater. The location of the Vembanad Lake is shown in Figure 1.



(a) Location

(b) Study Area
Figure 1. Vembanad Lake

The Vembanad lake eco system spans three Kerala districts: Eranakulam in the north, Alappuzha in the south, and Kottayam in the east. The lake's northern zone is dominated by salt water, while the southern zone is dominated by freshwater. The Thanneermukkom bund separates the two zones. The lake's width ranges from 500 metres to four kilometers, with the bund being the narrowest point. By 1975, the Thanneermukkom Bund had been built. It keeps salt water out of the paddy fields and encourages double cropping. This bund serves a low-lying field farming area of 55,000 hectares. The Vembanad lake's depth varies between 1 and 12 metres. During monsoons, floods occur, and in 1995, a spillway was built at Thottappally to help prevent them. It contributes to the rapid draining of water from the lake into the Arabian Sea (Mohandas &Brema 2021).

Since 1.6 million people live on the lake's bank, the Vembanad wetlands and the surrounding area have a variety of fiscal values and functions. It creates a social-ecological environment that sustains human life (Purandara BK, 2008; Purandara BK &Doray YL, 1987). It's a delicate ecological convergence region between the aquatic and terrestrial worlds. The lake produces a significant number of commercial fish and shellfish. Sea fishes and shrimps are among the other fish species supported. The lake supports rice farming in Kuttanad, Kerala's rice bowl, which covers 1,100 km^2 and is a reclaimed portion of the lake. This lake and its surrounding areas provide food and livelihood to a large living community, either directly or indirectly. According to research conducted on the lake area, the lake can be used for mining, fishing, tourism, and lime shell deposits. The surrounding lands are used for agriculture, with plans to grow rice, plantation crops, and other crops, as well as tourist resorts. These practices are contingent on the lake's and its environs' environmental integrity. The lake's environmental conditions are changing dramatically as a result of human interaction and hydrological variations. It has an effect on the local population's wellbeing. Fisheries practices, such as unscientific fishing with smaller mesh fishing nets, fishing at high tide, and so on, have an effect on the lake's endogenic environment. The development of Wellington Island, according to previous studies, is the primary cause of environmental pollution in Vemanad Lake. Budding works in the Kuttanadarea, as well as the construction of the Thotapally spillway to divert floodwaters from Achankovil, Pamba, Manimala, and Meenachil to the sea. The Thanneermukkon bund prevents salinity from entering the Kuttanad agricultural region during the summer season (Purandara BK, 2008; Purandara BK &Doray YL, 1987).The above mentioned intrusions altered the lake's flow pattern and ecology, making it unfit for human habitation. This Ramsar Site – Vembanad Lake was chosen as the study area in this research work for these reasons with the following objectives.

- Watershed map delineation and Land use classification using Sentinel 2A Data
- Spatial Distribution of water quality parameters using interpolation technique – Inverse Distance Weighted (IDW) Technique
- Development of linear regression model through Sentinel 2A reflectance values.

3. MATERIALS AND METHODS

3.1 Watershed Map

Watershed map indicates the area of land that drains into a common outlet along a stream. Watershed model is created with the help of hydrology toolset, which extracts the hydrological information from a Digital Elevation Model (DEM). ArcMap is the hydrological tool that is used in this research work. The model created will be useful for watershed analysis, which plays a vital role in regional planning, agriculture and forestry. Figure 2 shows the process involved in watershed map creation using Arc Map software.

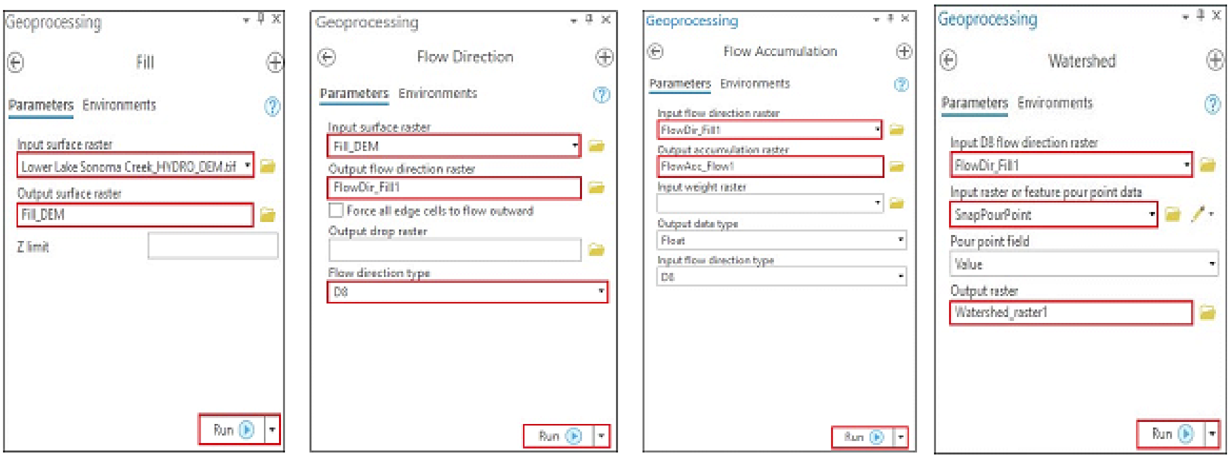


Figure 2. Process to create Watershed map

The watershed map of the study area has been depicted in Figure 3. The DEM data implemented on the map, indicates the water stream lines by which the sources of water enter the lake, its flow and the basin. This map is used to classify the study area through the Sentinel 2A data.

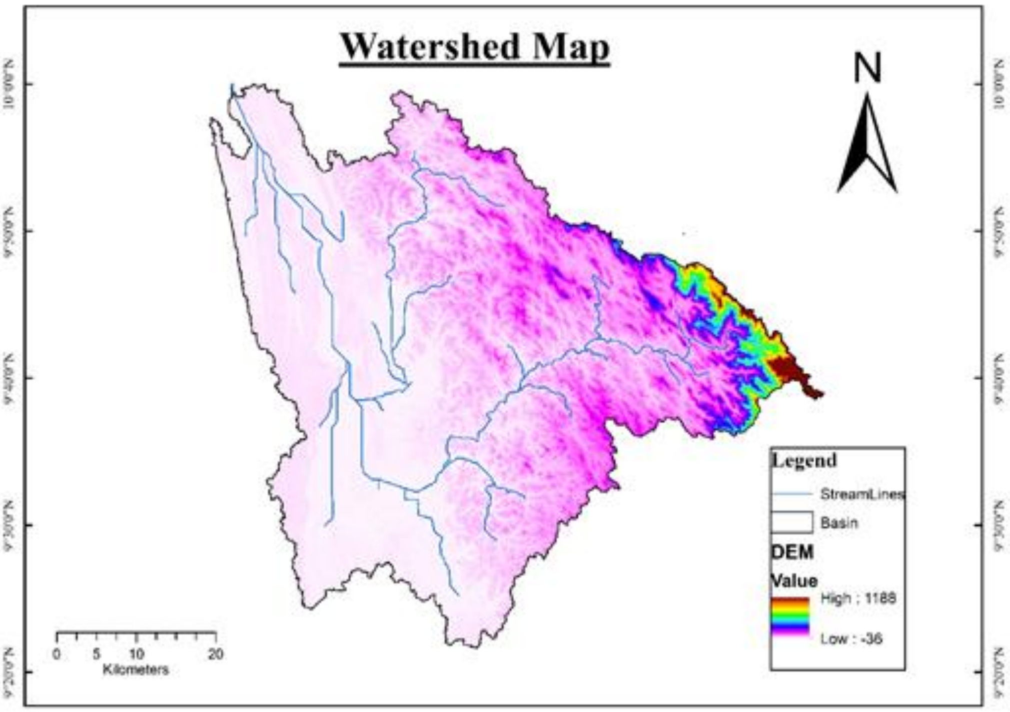


Figure 3. Watershed Map of Vembanad Lake

3.2 Land Classification

During the post-monsoon session, the last week of December 2020, 30 water samples from Vembanad Lake were collected. Figure 4 shows a few of the locations from where samples are obtained. The location points from where these samples are obtained are marked in the satellite images obtained through their latitude and longitude details and are used to obtain the land use map.

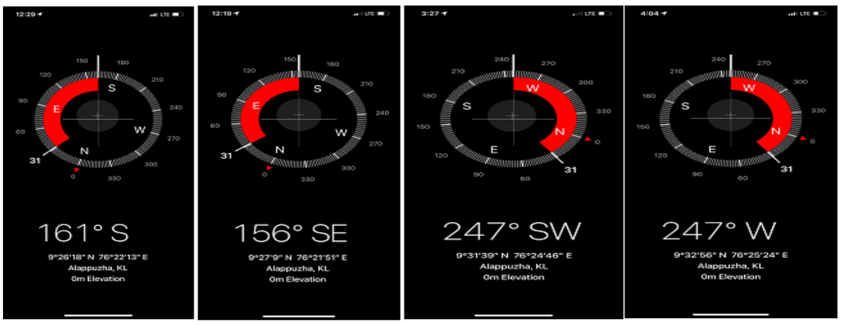
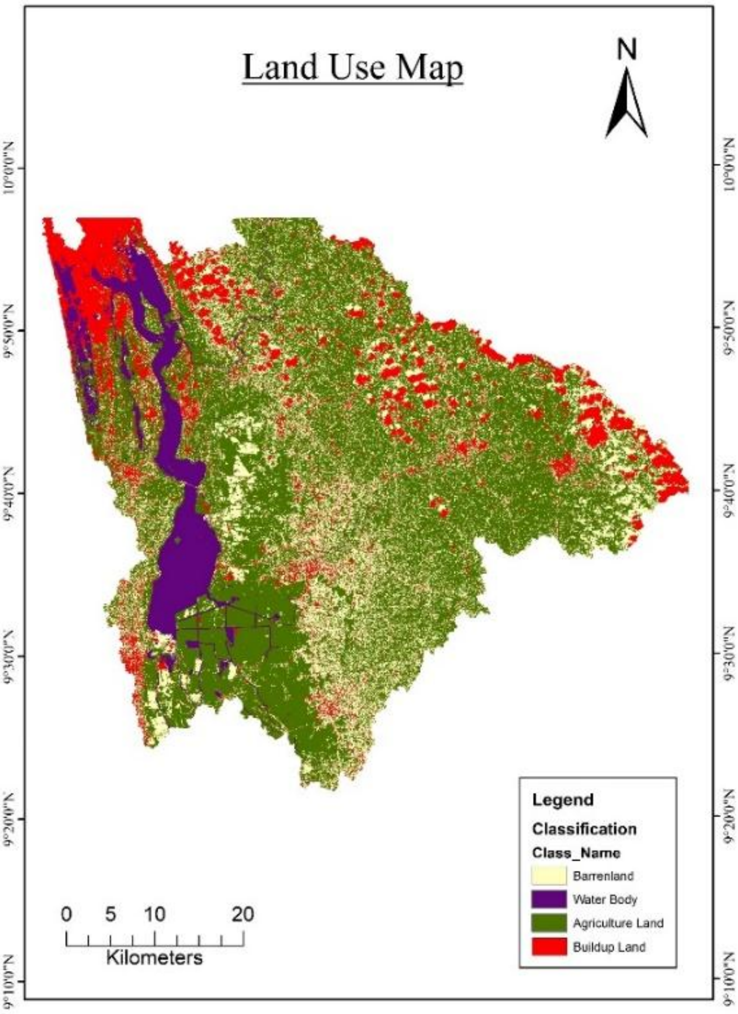


Figure 4 Locations where samples are collected

The high spatial resolution images are downloaded from United States Geological Survey (USGS) department for the location points from where the water samples are collected. These are obtained from Sentinel-2 satellite, whose images have a resolution range of 10 to 60 metres over land and coastal waters. The data is then subjected to atmospheric correction to eliminate the effects of atmospheric conditions. The spectral data for the different bands are extracted using the basin data extracted from DEM data. Using these data, classification of land happens through supervised learning method. In ArcGIS, software, image classification is an inbuilt tool, which is used to classify the study area. Loading the image of the study area and proper selection of pixels of the classification viz. barren land, water body, agriculture land and buildup land, the training samples can be made. Then applying it throughout the study area region, the land is classified to identify the barren land, water body, agriculture land and build up land as shown in Figure 5 (a) and the study area i.e. lake area extracted is depicted in Figure 5 (b).



(a) Classification of Study Area

(b) Lake Area Extracted from Classified Land

Figure 5. Classification of Study Area

3.3 Spatial Distribution of water quality parameters using interpolation technique – Inverse Distance Weighted (IDW) Technique

The spatial distribution of the water quality samples are plotted by using Inverse Distance Weighted (IDW) interpolation technique. To obtain the spatial distribution, first the elevation points of the locations from where the water samples are collected is needed. These are obtained by superimposing the DEM data on the classified land use map, particularly on the water body area. Those elevated points are reflected in Figure 6. To those elevation points, Inverse Distance Weighted (IDW) technique is applied, to get the spatial distribution.

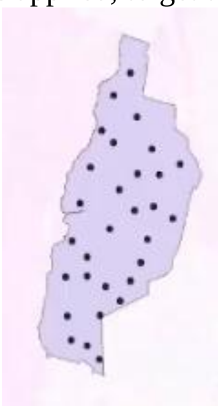


Figure 6. Elevation pints of locations from where water samples are collected

3.3.1 IDW Algorithm

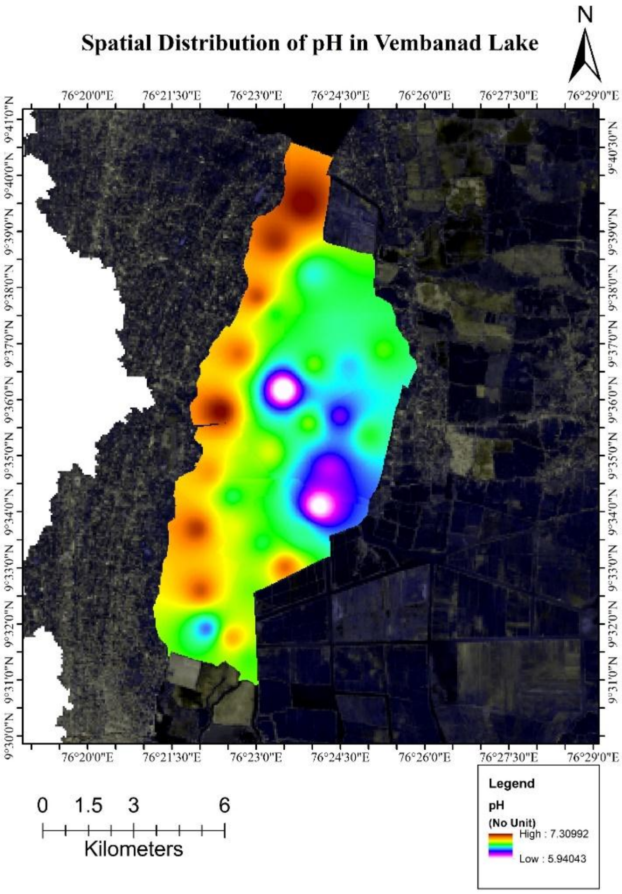
IDW is a deterministic multivariate interpolation approach that uses a known scattered set of points. A weighted average of the values available at the known points is used to assign values to unknown points. With the elevated points, this interpolation technique estimates the neighborhood values by averaging the values. The neighborhood values are influenced through the weight of the averaging process and decreases with distance from the sampled location. The neighborhood value calculated using IDW depends on the selection of the power value (p) and the search strategy. More is the emphasis on the nearest points, with increase in the power value. This results in more detailed neighborhood function. IDW for this research work is defined by the Equations 1, 2 & 3.

$$u(x) = \frac{\sum_{i=1}^{N} W_i(x) u_i}{\sum_{i=1}^{N} W_i(x)} \; if \; d(x, x_i) \neq 0 \; for \; all \; i \quad (1)$$

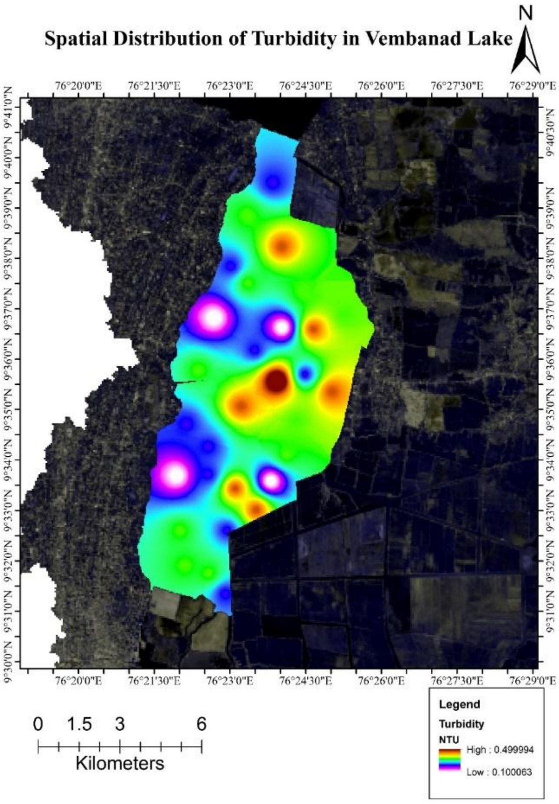
$$u(x) = u_i \; if \; d(x, x_i) = 0 \; for \; some \; i \quad (2)$$

$$W_i(x) = \frac{1}{d(x, x_i)^p} \quad (3)$$

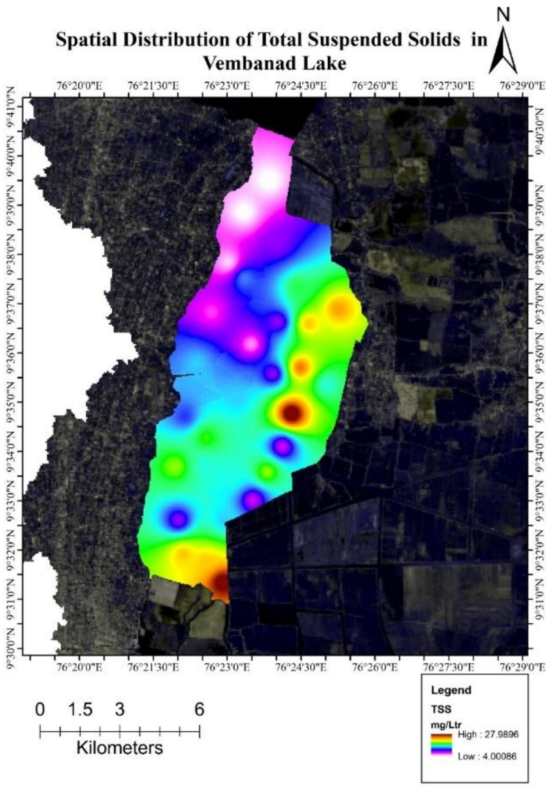
Where,
u(x) is the value of neighborhood interpolated location
u(i) is value of the nearest sampled location
N is the total number of samples
i is the sampled location
x is the neighborhood location
w is the weightage
d is the distance between the sampled location and neighborhood location
The water quality parameters for which spatial distribution aimed includes pH, turbidity, total suspended solids, total hardness, salinity and BOD. Through IDW technique, these parameters for the neighborhood functions are estimated and the spatial distribution is plotted as shown in Figure 7.



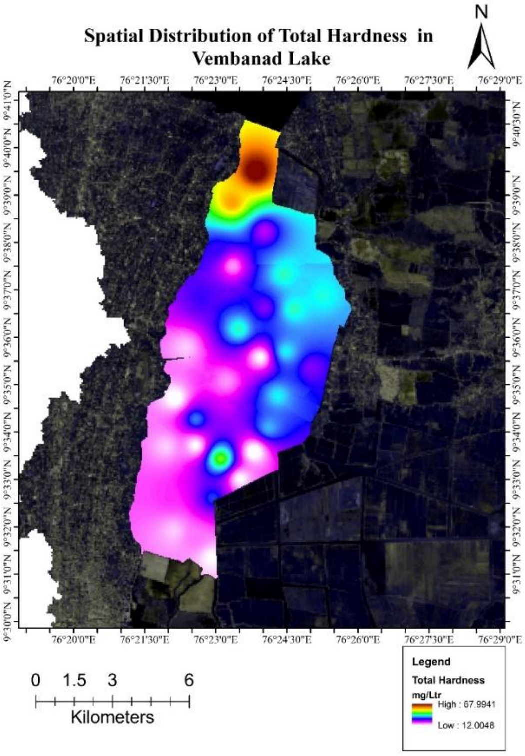
(a) pH



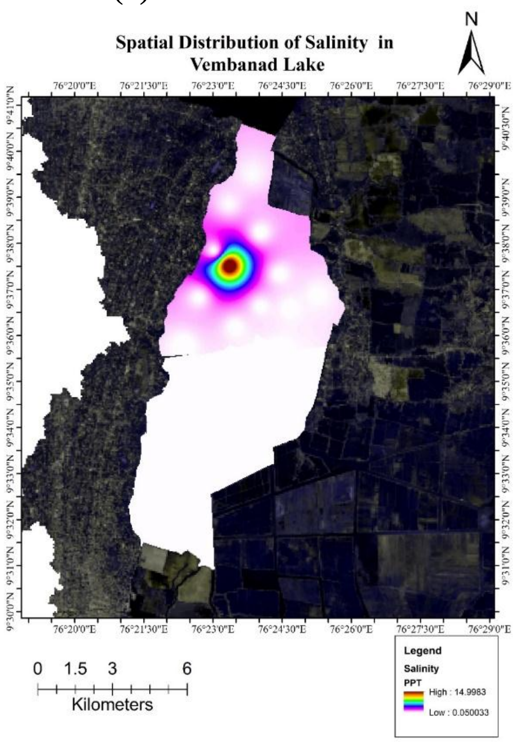
(b) Turbidity



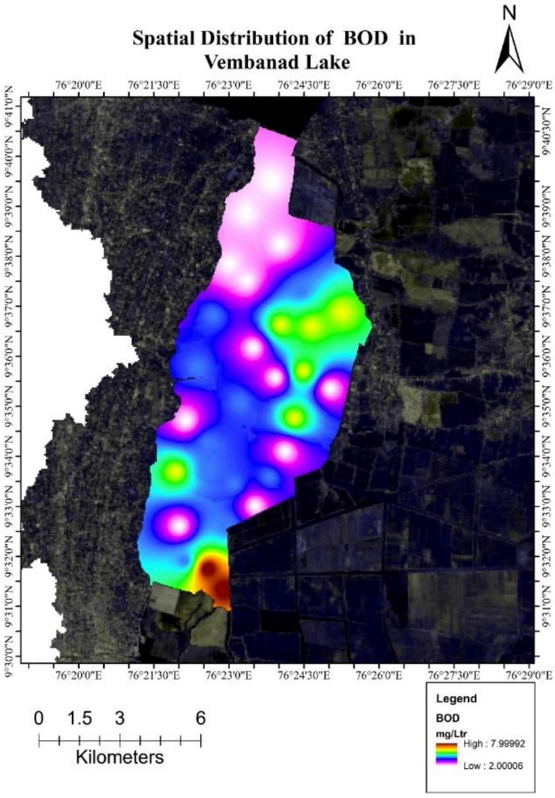
(c) TSS



(d) Total Hardness



(e) Salinity



(f) BOD
Figure 7. Spatial Distribution of Water Quality Parameters

3.3 Regression Analysis

The IDW algorithm assumes that the phenomenon being modeled is driven by local variation. It can be captured or modeled by defining an adequate neighborhood. But this technique does not provide prediction standard errors. In this way, justifying this model would be difficult. Regression analysis could be useful to predict the quality of water (Sri Dhivya Krishnan &Bhuvaneswari, 2017). In order to overcome this, a linear regression model has been developed from the reflectance values of the parameters considered for study from the Vembanad Lake area. The relationships between independent and dependent variable is defined by Regression analysis. It results in an expression, whose coefficient represent the correlation between the variables. The reflectance values of the water quality parameters are obtained from the sentinel-2A data. They are measured as a function of wavelength (Shafique, N et al, 2003& Warren M A et al, 2019) and are depicted in Figure 8.

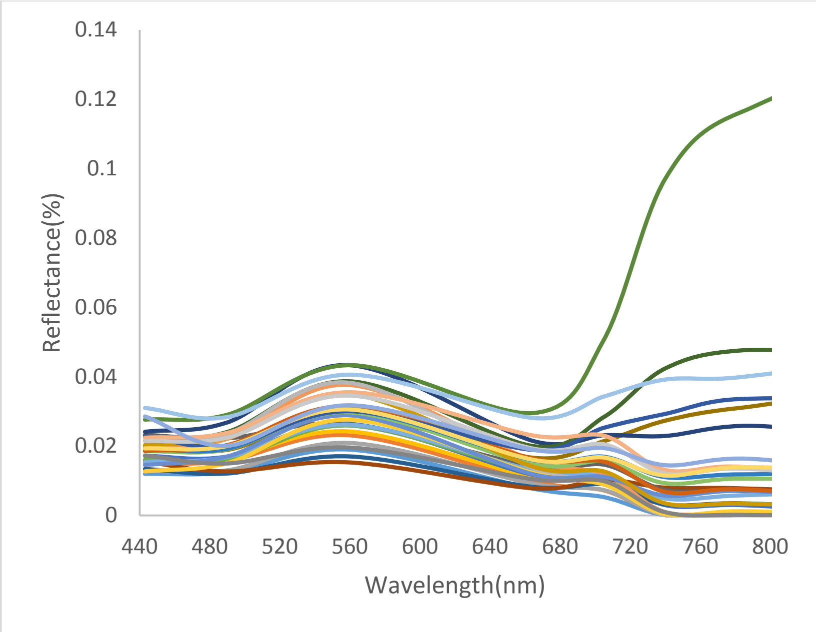


Figure 8. Spectral Reflectance

The spectral reflectance demonstrates that the water quality varies with the wavelength. Optical water quality parameters like suspended solids, turbidity and chlorphyll are inferred from the spectral reflectance curve. The peak refelctrnace values at 440 – 490 nm indicates the presence of chloprhyll, 550-590 nm indicates the presence of total suspended solids and 690 – 740 nm (Toming K 2016) refers to the turbidity of the water content. The spectral reflectance shows that the consistency of the water varies with wavelength. The spectral reflectance curve is used to infer optical water quality parameters such as suspended solids, turbidity, and chlorphyll. Peak reflectance values between 440 and 490 nm indicate the presence of chloprhyll, 550 and 590 nm indicate the presence of total suspended solids, and 690 and 740 nm indicate the turbidity of the water content (Toming K 2016).

4. RESULTS & DISCUSSIONS

4.1 Results

The values of the water quaity parameters given by the spectral development are analyzed using regression analysis. The band of ths spectrum is divided into bands 1 to 8 and the water qaulity parameters for each band are predicted. Form this the regression model is developed. The model predicts the relation between the band ratio and the water quality parameter. The R^2 value indicates the best fit expression and the exact value of the parameter. The water quality parameters of the sampled 30 loactions are tabulated in Table 1. The best fit model observed for those water quality paramters are tabulated in Table 2. Table 3 indicates the comparison of resultant water quality parameters corresponding to the best fit regression values with the standard BIS values. The values of the parameter obtained from regression models are compared with the BIS standard values (Soumya Singha et al, 2015) and the comparative analysis is depicted in Figure 9 (a) and 9 (b).

Table 1. Water Quality Parameters of the Sampled Locations

Location	Location reference no:	pH at 25° C	Turbidity	Electrical Conductivity	Total Dissolved Solids	Total Hardness as CaCo3	Total Suspended Solids	BoD @ 27° C for 3 days	Dissolved Oxygen	Chloride
		Value	Value NTU	Value (μs/cm)	Vlaue (mg/L)	Value mg/L	Value (mg/L)	Value (mg/L)	Value (mg/L)	Value (mg/L)
Location 1	ALT/JAN/21/W-26307	7.31	0.2	750	428	68	9	4	6.1	165
Location 2	ALT/JAN/21/W-26308	7.23	0.3	590	336	56	4	2	7	145
Location 3	ALT/JAN/21/W-26309	7.18	0.2	480	274	28	5	2	6.7	110
Location 4	ALT/JAN/21/W-26310	7.14	0.1	400	228	28	7	4	6.9	95
Location 5	ALT/JAN/21/W-26311	7.29	0.3	210	120	16	13	4	6.7	50
Location 6	ALT/JAN/21/W-26312	7.07	0.2	190	108	13	11	2	6.9	45
Location 7	ALT/JAN/21/W-26313	7.21	0.1	200	114	16	18	6	6.7	40
Location 8	ALT/JAN/21/W-26314	7.17	0.3	210	120	21	9	2	6.5	50
Location 9	ALT/JAN/21/W-26315	7.04	0.3	170	97	19	21	8	5.8	30
Location 10	ALT/JAN/21/W-26316	6.91	0.4	154	88	18	13	4	6	35
Location 11	ALT/JAN/21/W-26317	6.93	0.2	200	114	14	14	4	6.4	40

Location 12	ALT/JAN/21/W-26318	6.83	0.5	230	131	14	9	2	5.9	40
Location 13	ALT/JAN/21/W-26319	6.74	0.2	300	171	32	17	4	6.4	60
Location 14	ALT/JAN/21/W-26320	6.77	0.3	360	205	18	12	2	6.5	80
Location 15	ALT/JAN/21/W-26321	6.82	0.1	400	228	26	19	6	6.8	75
Location 16	ALT/JAN/21/W-26322	6.8	0.3	450	257	37	22	6	6.2	100
Location 17	ALT/JAN/21/W-26323	6.76	0.4	510	291	26	14	2	6.4	115
Location 18	ALT/JAN/21/W-26324	6.79	0.2	210	120	14	27	8	6.2	35
Location 19	ALT/JAN/21/W-26325	6.69	0.1	210	120	12	18	4	6.7	40
Location 20	ALT/JAN/21/W-26326	6.51	0.3	280	160	16	21	4	6.9	50
Location 21	ALT/JAN/21/W-26327	6.63	0.4	260	148	24	9	2	6.4	45
Location 22	ALT/JAN/21/W-26328	5.94	0.2	330	188	38	6	2	6.5	60
Location 23	ALT/JAN/21/W-26329	6.04	0.3	310	177	34	8	2	6.3	65
Location 24	ALT/JAN/21/W-26330	6.38	0.2	320	182	36	23	6	6.2	60
Location 25	ALT/JAN/21/W-26331	6.57	0.4	300	171	38	27	6	6.7	55
Location 26	ALT/JAN/21/W-26332	6.69	0.3	300	171	38	16	4	6.3	55

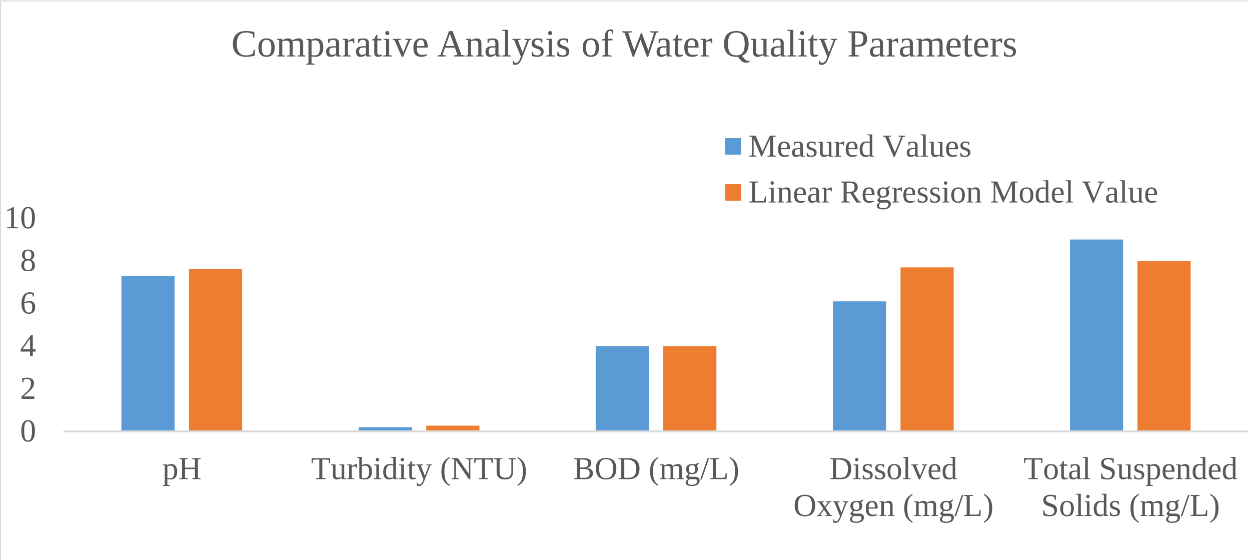
Location 27	ALT/JAN/21/W-26333	6.75	0.4	270	154	46	14	4	6.2	55
Location 28	ALT/JAN/21/W-26334	6.83	0.2	190	108	30	9	4	6.2	35
Location 29	ALT/JAN/21/W-26335	7.14	0.4	190	108	14	8	2	6.7	35
Location 30	ALT/JAN/21/W-26336	6.32	0.3	290	165	36	28	6	6.1	55

Table 2. Best Fit Model of the Water Quality Parameters from Linear Regression Analysis

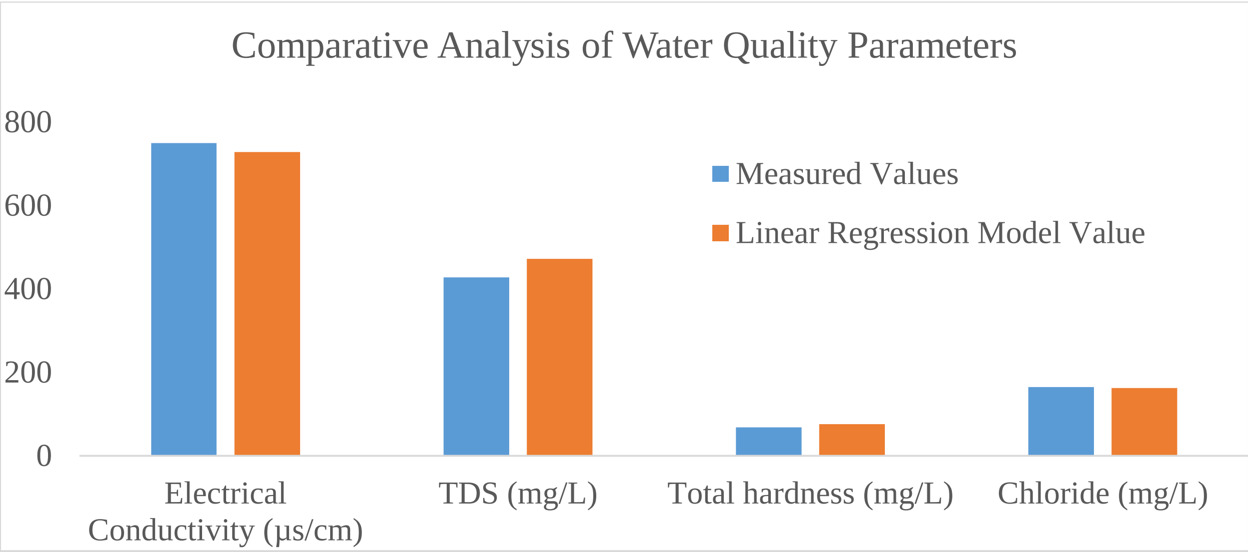
S. No.	Parameter	Model (band Ratio)	R^2 Value
1.	pH	pH = 5.0159 * (R_{490}/R_{665})	0.9798
2.	Turbidity	Turbidity = 0.195 * (R_{443}/R_{665})	0.8219
3	Electrical Conductivity	EC = 154.31 * (R_{560}/R_{705})	0.911
4	TDS	TDS = 171.01*($R_{665}+R_{705}$)	0.9223
5	Total hardness	TH = 26.439 * (R_{665}/R_{705})	0.8313
6	Total Suspended Solids	TSS = 19.984 * (R_{490}/R_{560})	0.8202
7	BOD	BOD = 5.7377 * (R_{490}/R_{560})	0.814
8	Dissolved Oxygen	DO = 3.0681 * (R_{490}/R_{705})	0.9666
9	Chloride	CL = 29.635 * (R_{560}/R_{705})	0.8941

Table 3. Compariosn of Water Quality Parameters for the Best Fit Linear Regression Model with Satndard BIS Values for sample one location

S. No.	Parameter	Value as per sample	Value as per Linear regression model	BIS Standard Value
1.	pH	7.31	7.62	8.5
2.	Turbidity (NTU)	0.2	0.28	5
3	Electrical Conductivity (µs/cm)	750	728	300
4	TDS (mg/L)	428	472	500
5	Total hardness (mg/L)	68	76	300
6	Total Suspended Solids (mg/L)	9	8	500
7	BOD (mg/L)	4	4	5
8	Dissolved Oxygen (mg/L)	6.1	7.7	5
9	Chloride (mg/L)	165	163	250



(a) pH, Turbidity, BOD, DO & TSS Values



(b) EC, TDS, TH & Cl Values

Figure 9. Comparative Analysis of Water Quality Parameters

4.2 Discussions

The results tabulated in Table 3 shows the comparison of the measured water quality parameters with the developed linear regression model and the standard BIS values. The comparison is portrayed in Figure 9. The analysis illustrates that pH, turbidity, Electrical Conductivity, Total Dissolved Solids, Total Suspended solids and chloride values are within safe limit, whereas BOD, dissolved oxygen and total hardness are higher than the BIS standard values. Through this regression model, these parameters are determined as the pollutants for Vembanad Lake. The samples, were taken during the partial lockdown period of the year 2020. Due to lockdown, the movements of human and boats were restricted. Owing to this reason the suspended solids were reduced to a greater amount and in turn the turbidity was also within the safe limit. The same may be the cause for other water quality parameters to be within the prescribed safe limits.

5. Conclusion

Water quality analysis of Vembanad Lake, Ramsar site of Kerala State has been attempted in this research work. The watershed map of the study area and the digital elevation model developed, classified the land of the satellite image. From the classified study areas water body, samples are obtained and the water quality parameters were analyzed. A linear regression model developed indicated the presence of pollutants across the lake. The parameters affecting the water quality of the lake has been determined. This paper concludes that linear regression model can be effectively used for water quality analysis.

5. REFERENCES:

[1] K A Mohandas & J Brema, "Remote Sensing Based Water Quality Analysis of Vembanad Lake", Journal of Green Engineering, Vol. 11, No.1, January 2021, pp. 1011-1020.

[2] Purandara BK, "Environmental issues of wetland ecosystem management: a case study", Ecol Environ Conserv., vol. 14, no. 2-3, pp. 439-448, 2008.

[3] Purandara BK, Dora YL, "Studies on texture and organic matter in the sediments of vembanad lake nearshore sediments", Proceedings of the national Seminary on Estuarine Management, 1987.

[4] Soumya Singha, C P Devatha, Sudhakar Singha & M K Verma, "Assessing Ground Water Quality Using GIS", International journal of Engineering Research & Technology, ISSN: 2278-0181, Vol. 4, No. 11, November 2015, pp. 689-694.

[5] Charan Kumar G, Shoba Latha, Rajyalakshmi K, "Identification of Influencing parameters of water quality and its analysis by using spatial graphs and autocorrelation function", Journal of critical reviews, vo. 7, No.5, 2020, pp. 614-619.

[6] Kobayashi, "On the spatial graph", Kodai Math Journal, 17, 1994, pp. 511-517.

[7] G. Gyananath, SR Islam and S.V.Shewdikar , Assessment of Environmental Parameter on ground water quality, Indian Journal Of Environmental Protection, Vol.21, No.4, April 2001, pp.: 289-294

[8] Dormann, CF., Mcpherson, J.M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohelemuller, R., Peres-Neto, P. R., reineking, B., Schroder, B., Schurr, F. M., Wilson, R., "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review", Ecogrpahy, 30, 2007, pp. 609-628.

[9] Fan, C and Myint S, "A Comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation", Land Scape and Urban Planning, 12, 2014, pp. 117-128.

[10] Shafique, N. A., Fulk, F., Autrey, B. C., &Flotemersch, J. (2003, October). Hyperspectral remote sensing of water quality parameters for large rivers in the Ohio River basin. In *First interagency conference on research in the watershed, Benson, AZ* (pp. 216-221).

[11] Warren, M. A., Simis, S. G., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., ... &Ansper, A. (2019). Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sensing of Environment*, *225*, 267-289.

[12] Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., &Nõges, T. (2016). First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sensing*, *8*(8), 640.
[13] K. Sri Dhivya Krishnan and P. T. V. Bhuvaneswari, "Multiple linear regression based water quality parameter modeling to detect hexavalent chromium in drinking water," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, pp. 2434-2439, doi: 10.1109/WiSPNET.2017.8300197.