ANALYZING SPASIAL GROUNDWATER SALINITY USING MULTIVARIATE ANALYSIS AND MULTIPLE LINEAR REGRESSION MODELS

Kristin Ina Binna[1], Ramadhani Yanidar[1]*, Sheilla Megagupita Putri Marendra[1], Herika Muhammad Taki[2], Ariani Dwi Astuti[1]

[1]Department of Environmental Engineering, Universitas Trisakti, Indonesia
[2]Department of Planology, Universitas Trisakti, Indonesia

Abstract

The increase in the amount of groundwater withdrawal will inevitably pose a threat of seawater intrusion. The purpose of this research was to identify the distribution of shallow groundwater salinity in North Jakarta, West Jakarta and Central Jakarta and to develop a regional model of shallow groundwater salinity distribution. The data used in this study was that of the groundwater quality monitoring, obtained from the Regional Environment Status Book (SLHD), published by The Environment office of Greater Jakarta released in 2022, involving a total of 121 sample points in North Jakarta, West Jakarta, and Central Jakarta. The primary data was taken at 6 (six) sampling locations for model validation purposes. The study began with data grouping, using the Hierarchical Cluster Analysis (HCA) method. The results of identifying the highest distribution of salinity are in cluster 3 (three). A model was subsequently developed, after removing the outliers, with multiple linear analysis methods using the variable the distance from the coastline (X_1), well depth (X_2) and hardness (X_3), to determine the influence of EC, TDS and salinity distribution in shallow groundwater. The results obtained are as follows; EC Models: $Y_{EC3} = -1.879+ (1.19.X_1) + (5.08.X_3)$. TDS models: $Y_{TDS3} = -2.211.30 + (0.81.X_1) + (101.41.X_2) + (4.07.X_3)$. Salinity models: $Y_{salinity3} = -0.07+ (6.75\times10^{-5}.X_1) + (2.4\times10^{-4}.X_3)$. Model verification results for $R^2_{EC3} = 0.70$; $R^2_{TDS3} = 0.92$; $R^2 salinity_3 = 0.88$. Validation results produce 21.14% for EC, 8.21% for TDS, and 22.87% for Salinity. This needs further research by increasing the number of primary samples.

Keywords: *electrical conductivity, model, salinity, shallow groundwater, total dissolved solid*

Introduction

The existence of water is key to the sustainability of life. Water sources can come from various places, such as rainwater, rivers, lakes, or water trapped in the soil or aquifers. The water found within the soil is known as groundwater. This groundwater moves between soil particles, seeping into the ground and combining to form an aquifer (Davis & Cornwell, 2008). In the natural hydrological system, the balance of groundwater in an area is maintained between input and output. Groundwater input comes from rainwater infiltration, while groundwater output includes springs and the use of dug wells or boreholes. The balance of groundwater can be disrupted, if it is continuously exploited, which is inversely related to the availability of water resources.

Based on the data from the Central Statistics Agency (BPS) in 2021, the area of DKI Jakarta is approximately 664.01 km2, with a population of 10,644,776. As the population increases and the economy grows, the demand for clean water also continues to rise, both for daily use and industrial needs. However, the water supply for drinking water in DKI Jakarta is only partially

*)Corresponding Author:
E-mail: ramadhani@trisakti.ac.id

Received: 23 February 2024
Revised : 16 March 2024
Accepted: 16 March 2024
DOI: 10.23969/jcbeem.v8i1.12708

provided by the Jakarta Water Company (PAM), serving approximately 64% of the Jakarta population. The water supply provided by PAM DKI Jakarta cannot yet meet the water needs of the entire Jakarta population. To meet the demand for clean water, the people of Jakarta rely more on groundwater from shallow or deep aquifer storage layers.

Groundwater quality is often influenced by various factors, one of which is salinity. Groundwater salinity levels can increase due to natural processes, such as seawater intrusion or human activities (Gurmessa et al., 2022). Excessive extraction of groundwater in DKI Jakarta area can disrupt the subsurface environment by mixing shallow and deep aquifers and causing seawater intrusion in coastal areas (Kagabu et al., 2010). Variations in human as well as natural pressures lead to groundwater salinization to change over space and time, it is critical to regularly monitor and assess the variability of physical and chemical parameters in groundwater (Parisi et al., 2023).

The Jakarta aquifer structure, which is relatively flat, is a coastal type, making the Jakarta area vulnerable to saltwater intrusion. Saltwater intrusion is the movement of saltwater into freshwater aquifers, which can contaminate the source of drinking water. Saltwater intrusion can be identified by the distribution of salinity in groundwater. Decisive actions by the Jakarta government to protect the environment and preserve the quality and quantity of groundwater and the involvement of the community in environmental conservation (Wijaya et al., 2019) are required, to address this critical issue, so that, the Sustainable Development Goal (SDG) number 6, which is sustainable availability and management of clean water and sanitation for all, can be achieved.

Some important indicators for analyzing saltwater intrusion include electrical conductivity (EC), total dissolved solid (TDS), and salinity, in areas with a large amount of data. The electrical conductivity reaches >5,000 μS/cm (brackish groundwater) in the confined aquifer system (40 - 140 meters below mean sea level) in the northern part of the Jakarta Groundwater Basin (*Cekungan Air Tanah/ CAT*) (Setiawan et al., 2017).

Multivariate statistical modelling allows for the classification of groundwater quality data based on more than one dependent variable and large amounts of data is cluster analysis (Wali et al., 2022). It is used to group similar data into clusters, based on their characteristics. The purpose of this research is to identify the distribution of shallow groundwater salinity in North Jakarta, West Jakarta, and Central Jakarta and to develop a regional model of groundwater salinity distribution, considering the distance from the shoreline and well depth when calculating EC, TDS, and salinity.

Research Methodology

Description of Study Area and Water Quality Data

Monitoring shallow groundwater salinity is critical for managing water resources in low-lying coastal town. (Setiawan et al., 2022). These three (3) research areas are within the DKI Jakarta province, which is a low-lying area, with an average elevation of 7 meters above sea level. This research was conducted in the North Jakarta, West Jakarta, and Central Jakarta regions, with the following geographical conditions: The North Jakarta Administrative City is situated at 106°20'00"E and 06°10'00"S. The location of the West Jakarta Administrative City is at 106°48'0"E and 06°12'00"N, while the Central Jakarta Administrative City is located between 106°24'42"E and 106°58'18" E, and between 5°19'12"S and 6°23'54"S (Bappeda DKI Jakarta, 2018). The data used in this study was that of the groundwater quality monitoring, obtained from the Regional Environment Status Book (SLHD), published by The Environment office of Greater Jakarta DKI Jakarta released in 2022. A total of 121 secondary data points were analyzed (Figure 1).

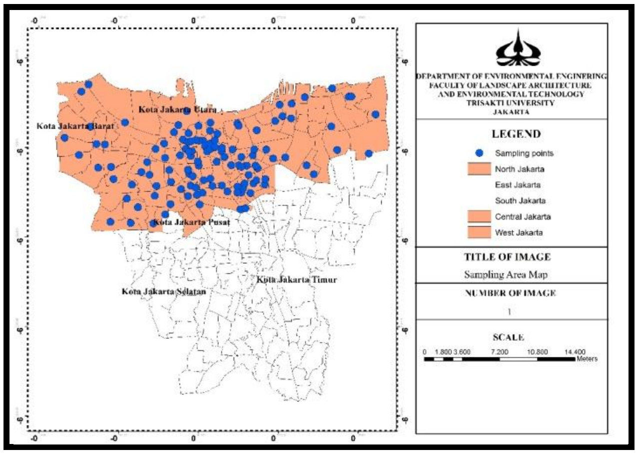

Figure 1. Sampling Area Map.

Based on the secondary data, the groundwater quality from the Environmental Agency (DLHK) was validated by collecting 3 (three) samples of community well water in each respective region. These samples were analysed in the Environmental Laboratory of the Department of Environmental Technology, Faculty of Landscape Architecture and Environmental Technology, Trisakti University.

Tools and Materials

SPSS 26.0 was employed as a statistical tool for data analysis. The analysis was performed to obtain descriptive values, with the goal of providing the primary characteristics of the data in a numerical manner.

Z-scores were also utilised to standardise the variables, minimising variations in variable size and units.

A total of 121 secondary data points were analysed using cluster analysis. In this analysis, the agglomerative ward hierarchical method was applied to group similar data points and generates a dendrogram. To examine the correlation between Electrical Conductivity (EC), Total Dissolved Solid (TDS), and salinity parameters with the research variables, namely well depth, the distance from the sampling point to the coastline, and hardness, multiple linear regression was employed [Eq. (1)].

$$Y=aX_1+bX_2+cX_3+C \qquad (1)$$

where:

Y = the concentration of parameter (EC, TDS, salinity)

X_1 = well depth

X_2 = the distance from the coastline

X_3 = hardness

Results and Discussion

To understand the initial characteristics of the research area, a descriptive analysis was conducted, and the results are as in Table 1. Maximum TDS value of 3,460mg/l indicates brackish water (>3000mg/l) and, maximum EC concentration =4,620 µS/cm (<5,000 µS/cm) slightly brackish in some locations.

Table 1. Descriptive Values of the Water Quality Parameter and Research Area Variabel

Parameters	Unit	Quality Standards	Min	Max	Mean	Std. Deviation
Color	TCU	50	2.0	222.00	24.37	31.75
Fe	mg/l	1	0.03	2.23	0.19	0.37
Hardness	mg/l	500	25.70	1,305.86	227.79	159.05
Nitrite	mg/l	10	0.01	2.00	0.06	0.20
pH		6.5-8.5	6.21	8.58	7.26	0.41
TDS	mg/l	1.000	151.0	3.460	732.89	610.28
EC	µS/cm		0.98	4.620	1,021.28	825.71
Salinity	$\%_0$		0.01	0.41	0.06	0.06
Turbidity	NTU		0.00	223.00	6.23	23.05
Coastline	m		441.07	13,109.77	5,710.27	2,735.68
Color	TCU	50	2.0	222.00	24.37	31.75

Cluster Analysis

Cluster Analysis, using the backward method, was employed to determine the similarity of water quality parameters, resulting in a dendrogram as shown in Figure 3.1. The cluster analysis indicates that the 121 sampling locations were grouped into four main clusters. Cluster 1 is composed of 56 sampling points, cluster 2 includes 43 sampling points, cluster 3 comprises 16 sampling points, and cluster 4 contains 6 sampling points.

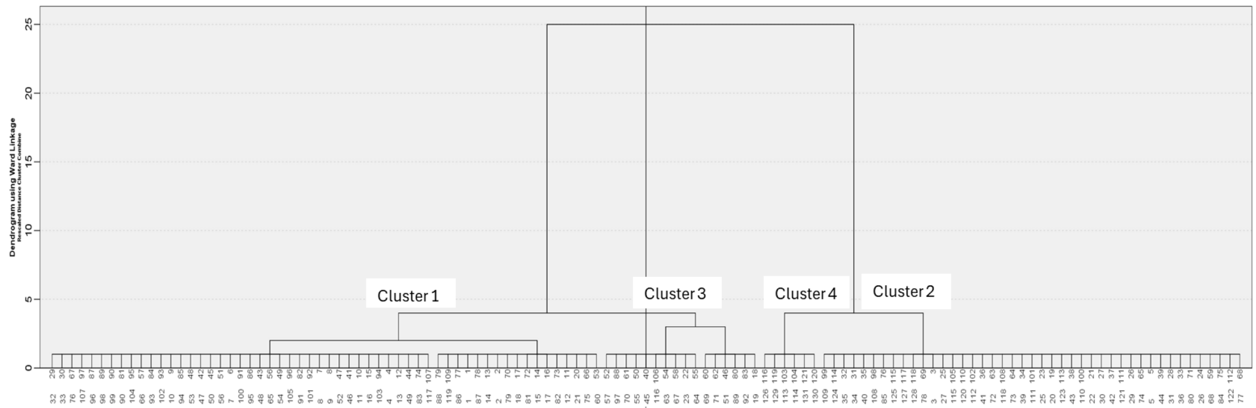

Figure 2 Dendrogram based on agglomerative ward hierarchical clustering for 121 sampling sites

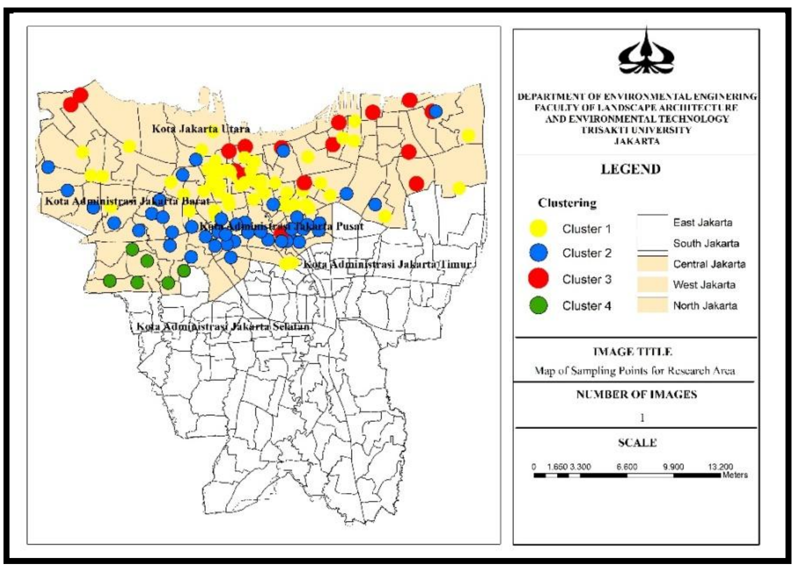

Figure 3 Mapping of Clustering Sampling Point

Based on the clustering analysis, the results are as follows: Cluster 1 is dispersed across the North, West, and Central Jakarta regions, comprising a total of 56 sample points. The sampling locations for Cluster 1 are situated between 2.0 km and 6.5 km from the coastline, with an average distance of approximately 4.5 km. The well sample depths also vary from 1 meter to 250 meters, with an average depth of 22.72 meters.

Cluster 2 is similarly dispersed in the North, West, and Central Jakarta areas, with a total of 43 sample points. The sampling locations for Cluster 2 range from 6.5 km to 10.0 km from the coastline, averaging about 8.0 km. Well sample depths vary from 4 meters to 200 meters, with an average depth of 26.94 meters.

Cluster 3 is scattered in the West and Central Jakarta areas, primarily concentrated in the North Jakarta region, totalling 16 sample points. The sampling locations for Cluster 3 range from 0.4 km to 1.9 km from the coastline, averaging about 8.0 km. Well sample depths vary from 2 meters to 15 meters, with an average depth of 8 meters.

Cluster 4 is distributed in the West Jakarta area, comprising a total of 6 sample points. The sampling locations for Cluster 4 are between 10.9 km and 13.0 km from the coastline, with an average distance of around 12.0 km. Well sample depths also vary from 20 meters to 25 meters, with an average depth of 21 meters.

The quality groundwater analysis of each variable within each cluster, the results are obtained using the Box and Whisker Plot graph, to provide a clearer view of the Interquartile Range (IQR) in the boxplot for each cluster, as shown in Figure 4, 5, and 6.

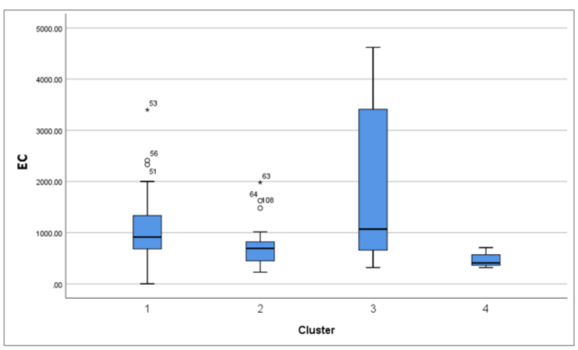

Figure 4. EC Boxplot for Each Cluster

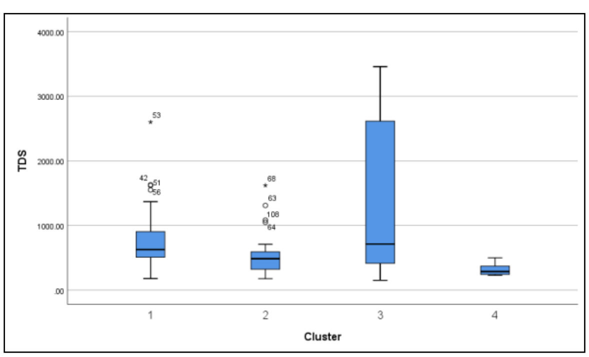

Figure 5. TDS Boxplot for Each Cluster

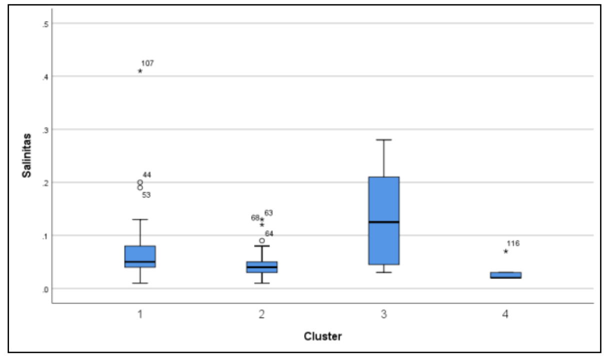

Figure 6. Salinity Boxplot for Each Cluster

In the analysis results depicted in Boxplots from Figure 4 to Figure 6, it is evident that Cluster 3 (three) has a higher potential for saltwater intrusion. This is indicated by higher levels of hardness, pH, EC, TDS, and salinity, compared to Clusters 1 (one), 2 (two), and 4 (four). This is likely influenced by the geographical conditions of Cluster 3 (three), which tend to be distributed in the North Jakarta region. In this study, boxplot analysis was employed to examine the distribution of pH, Electrical Conductivity (EC), Total Dissolved Solids (TDS), and salinity data in three research areas: North Jakarta, West Jakarta, and Central Jakarta.

The results revealed significant differences in pH data, with point 30 as an outlier in cluster 1, and points 71 and 83 as outliers in clusters 2 and 3, respectively. In cluster 4, there were no extreme values for pH. The highest concentration of EC was found in cluster 3, while cluster 4 had the lowest EC values. Regarding TDS, significant differences were observed in clusters 1 and 2, with several points outside the whisker limits, such as points 53, 42, 51, and 56 in cluster 1, and points 68, 63, 108, and 64 in cluster 2.

The highest salinity distribution was observed in cluster 3. For salinity, outliers were identified at points 107, 44, and 53 in cluster 1, and points 63, 68, and 64 in cluster 2. Cluster 4 had only one outlier at point 116. In conclusion, the highest distribution for pH, EC, TDS, and salinity was found in cluster 3, with some outliers requiring special attention. In the analysis results depicted in Boxplots from Figure 4 to Figure 6, it is evident that Cluster 3 (three) has a higher potential for saltwater intrusion. This is indicated by higher levels of hardness, pH, EC, TDS, and salinity, compared to Clusters 1 (one), 2 (two), and 4 (four). This is likely influenced by the geographical conditions of Cluster 3 (three), which tend to be distributed in the North Jakarta region. In this study, boxplot analysis was employed to examine the distribution of pH, Electrical Conductivity (EC), Total Dissolved Solids (TDS), and salinity data in three research areas: North Jakarta, West Jakarta, and Central Jakarta.

The results revealed significant differences in pH data, with point 30 as an outlier in cluster 1, and points 71 and 83 as outliers in clusters 2 and 3, respectively. In cluster 4, there were no extreme values for pH. The highest concentration of EC was found in cluster 3, while cluster 4 had the lowest EC values. Regarding TDS, significant differences were observed in clusters 1 and 2, with several points outside the whisker limits, such as points 53, 42, 51, and 56 in cluster 1, and points 68, 63, 108, and 64 in cluster 2. The highest salinity distribution was observed in cluster 3. For salinity, outliers were identified at points 107, 44, and 53 in cluster 1, and points 63, 68, and 64 in cluster 2. Cluster 4 had only one outlier at point 116. In conclusion, the highest distribution for pH, EC, TDS, and salinity was found in cluster 3, with some outliers requiring special attention.

The geological structure of an area can also influences saltwater intrusion, such as the impact of groundwater basins and rock types (Persaud & Levison, 2021). Identifying saltwater intrusion in groundwater basins is a crucial process in the management of groundwater resources. The discussion of geological conditions serves as one way to analyze the presence of saltwater intrusion. Each location has its own unique geological conditions, including lithology. The geological conditions in the Jakarta region are entirely formed by sedimentary rocks from the early Miocene to the early Pliocene, volcanic rocks, and current-age surface deposits.

Ground Water Quality Using Multiple Linear Regression Analysis Model

It has been identified that the highest values for EC, TDS, and salinity are in cluster 3. Consequently, a multiple linear regression analysis model was developed, and the results are as in Table 2.

Table 2. The Multiple Linear Model of Cluster 3

Parameters	**R^2**	**C**	**a**	**b**	**c**	**Significant F(16.95)**
EC	0.55	-1.879.5	1.19	-	5.08	0.020
TDS	0.83	-2.211.3	0.81	101.41	4.07	0.002
Salinity	0.59	-0.07	6.75×10^{-5}	-	2.42×10^{-4}	0.030

In Cluster 3 (three), there are a total of 16 sampling points distributed across 2 points in Central Jakarta, 10 points in North Jakarta, and 4 points in West Jakarta. The Cluster 3 models are as follows:

1. EC: $Y_{EC3} = -1{,}879.54 + (1.19X_1) + (5.08\ X_3)$.
2. TDS: $Y_{TDS3} = -2{,}211+0.8X1+101.4X2+4.1X3$
3. Salinity: $Y_{salinity3} = -0.07 + (6.8 \times 10^{-5}X_1) + (2.42 \times 10^{-4}X_3)$.

The verification results of the model at each primary sample collection point, compared to the secondary data are as following figures.

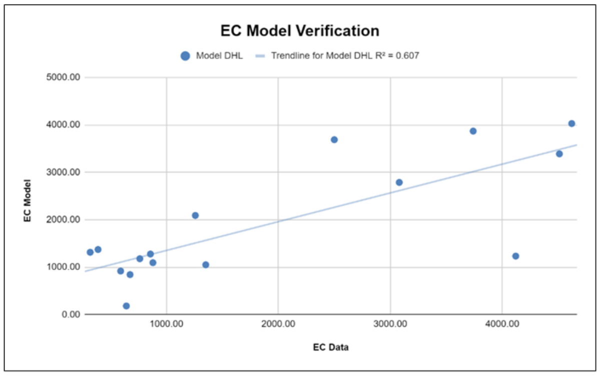

Figure 7. EC model verification

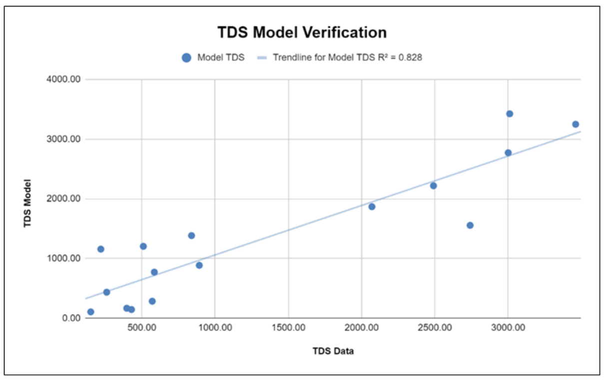

Figure 8. TDS model verification

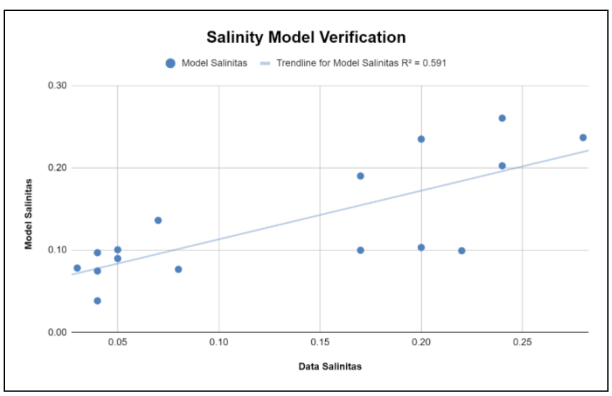

Figure 8. Salinity model verification

From the validation results, a deviation value between the data and the model was obtained, amounting to 21.14% for EC, 8.21% for TDS, and 22.87% for Salinity. These deviation values are still acceptable for EC.

Conclusions

The results of salinity distribution identification and analysis of the developed model using the multiple linear analysis method, considering coastal distance, well depth, and water hardness level indicate that the EC cluster 3 (three) model is influenced by coastal distance and hardness. The TDS model in cluster 3 (three) is influenced by coastal distance, well depth and hardness, while the salinity model in cluster 3 (three) is influenced by coastal distance and hardness. All the three variables are constrained within the range of coastal distance from 441.07 meters to 4,339.31 meters and hardness concentration from 142.90 mg/l to 618.76 mg/l for shallow groundwater. EC, TDS, and salinity in cluster 3 (three) are believed to have two influencing factors, namely saltwater intrusion, and geological conditions.

Acknowledgment

The authors would like to thank to Environment Department of DKI Jakarta Province, Indonesia, for their cooperation and for providing ground water quality monitoring data used in this study.

References

Davis M. L. & Cornwell, D. A. (2008). *Introduction to Environmental Engineering* (fourth ed., pp. 8–9). McGraw Hill.

Gurmessa, S. K., MacAllister, D. J., White, D., Ourdraog, I., Lapworth, D., & MacDonald, A. (2022). Assessing groundwater salinity across Africa. *Science of the Total Environment*, *828*, 154283. https://doi.org/10.1016/j.scitotenv.2022.154283

Kagabu, M., Delinom, R. M., Lubis, R. F., Shimada, J., & Taniguchi, M. (2010). Groundwater Characteristics in Jakarta Area, Indonesia. *Jurnal RISET Geologi Dan Pertambangan*, *20*(2), 69. https://doi.org/10.14203/risetgeotam2010.v20.35

Parisi, A., Alfio, M. R., Balacco, G., Güler, C., & Fidelibus, M. D. (2023). Analyzing spatial and temporal evolution of groundwater salinization through Multivariate Statistical Analysis and

Hydrogeochemical Facies Evolution-Diagram. *Science of the Total Environment*, *862*(December 2022). https://doi.org/10.1016/j.scitotenv.2022.160697

Persaud, E., & Levison, J. (2021). Impacts of changing watershed conditions in the assessment of future groundwater contamination risk. *Journal of Hydrology*, *603*(PD), 127142. https://doi.org/10.1016/j.jhydrol.2021.127142

Setiawan, I., Morgan, L., Doscher, C., Ng, K., & Bosserelle, A. (2022). Mapping shallow groundwater salinity in a coastal urban setting to assess exposure of municipal assets. *Journal of Hydrology: Regional Studies*, *40*, 100999. https://doi.org/10.1016/j.ejrh.2022.100999

Setiawan, T., Yermia, E., Purnomo, B. J., & Tirtomihardjo, H. (2017). Intrusi Air Laut Pada Sistem Akuifer Tertekan Cekungan Air Tanah Jakarta Berdasarkan Analisis Hidrokimia Dan Hidroisotop. *RISET Geologi Dan Pertambangan*, *27*(1), 1. https://doi.org/10.14203/risetgeotam2017.v27.430

Wali, S. U., Alias, N. B., Harun, S. Bin, Umar, K. J., Gada, M. A., Dankani, I. M., Kaoje, I. U., & Usman, A. A. (2022). Water quality indices and multivariate statistical analysis of urban groundwater in semi-arid Sokoto Basin, Northwestern Nigeria. *Groundwater for Sustainable Development*, *18*(August 2021), 100779. https://doi.org/10.1016/j.gsd.2022.100779