Application of multivariate statistical techniques in the evaluation of large-scale water treatment plants in Baghdad.

Nisreen Y. Mohammed¹ and Khalid A. Abdulrazzaq²

¹M.Sc. student, civil engineering department, college of engineering, University of Baghdad, Iraq. Email: n.mohammed1101@coeng.uobaghdad.edu.iq
²Assistant Professor, Civil engineering department, college of engineering, University of Baghdad, Iraq. Email: aleouboady@coeng.uobaghdad.edu.iq

Abstract. This paper aims to evaluate large-scale water treatment plants' performance and demonstrate that it can produce high-level effluent water. Raw water and treated water parameters of a large monitoring databank from 2016 to 2019, from eight water treatment plants located at different parts in Baghdad city, were analyzed using nonparametric and multivariate statistical tools such as principal component analysis (PCA) and hierarchical cluster analysis (HCA). The plants are Al-Karkh, Sharq-Dijlah, Al-Wathba, Al-Quadisiya Al-Karama, Al-Dora, Al-Rasheed, Al-Wehda. PCA extracted six factors as the most significant water quality parameters that can be used to evaluate the variation in drinking water quality and responsible for 73.389% of the variance in the data set. Based on this selection criterion, the most significant water quality parameters that can be used to evaluate the variation in drinking water quality parameters are the mineral-related parameters (e.g., Ca²⁺, Mg²⁺, salinity, hardness), the nutrient parameters (i.e., dissolved nitrate and nitrite and orthophosphate), and a physical parameter. HCA analysis was able to group water treatment plants with similar raw water and treated water quality based on the water quality data from eight WTPs into three clusters.

Keywords: Cluster analysis; Principal component analysis; Multivariate statistics; Water treatment plants.

1. Introduction

Baghdad water supply and sanitation are discriminated against by poor water quality and lack of service. Three decades of war, along with limited environmental cognizance, have destroyed Baghdad's water resources management system. 91% of the population has access to drinking water, with significant variations between governates and urban and rural areas. Only 77% of the population has access to improved drinking water sources in ruler areas, while 98% in urban areas. Also, 6.2 of the population does not use an enhanced sanitation facility. Besides the inadequacy of potable water, the threat of waterborne diseases is critical public health issues. Preventable health risks are directed to poor water management services. UNICEF (2012)[1]. Clean and safe water results from a careful evaluation of source water quality variation and the plant's performance monitoring. The operators of water treatment plants (WTPs) should regularly analyze water treatment plants' performance and ensure the system operates with the most efficient equipment and technology. Evaluating the plant's performance helps identify the factors that inhibit the treatment plant (treatment processes) from producing acceptable quality water. Adequate and representative performance indicator systems are required to ensure the proper functioning of each water treatment unit. Usually, performance indicator (PI) methodologies have been proposed on performance
evaluation of WTPs, which is a simple, understandable tool. Performance indicator system (PIs) can be used to evaluate the performance historically over previous periods to show improvement or deterioration in performance so that remedial measures can be taken before service is affected (Marques & Monterio (2001)[2]; Perotto et al. (2008)[3]; Ofwat (2004)[4]).

Vieira et al. (2008)[5] conducted a performance indicator system to be used as a standardized methodology for performance assessment in drinking water treatment plants. They identified 80 PI over seven domains: treated water quality, plant reliability, natural resources and raw materials; by-products management; safety; human resources; and economic and financial resources. Sadiq et al. (2010)[6] assessed small water utilities' performance using Ordered Weighted Averaging (OWA) operators and fuzzy set theory based on various performance indicators.

Further on, some studies have focused on the multi-parameter water quality index (WQI) (Chang et al. (2007)[7]; Quevedo-Castro et al. (2018)[8] for evaluation WTPs. In recent years, Multivariate statistical tools, particularly hierarchical cluster analysis (HCA) and principal component analysis have been successfully used in the water field to analyze the complex structure that underlies many analytical data. The PCA is a powerful tool capable of handling massive amounts of data, making it smaller uncorrelated significant variables while still keeping as much information as possible, explaining the variance observed. Olawoyin et al. (2014)[9]. The cluster analysis is an analytical approach comprised of a family of techniques that grouping similar variables to make informed decisions. Many types of research have used the multivariant analysis for different purposes, including evaluating the performance of water treatment plants (Flaieh & Abdul-ahad, (2014)[10]; Melo et al. (2019)[11]; Issa & Alrwai, (2018)) [12]; Saad & Hamdan, (2020)[13], wastewater treatment plants (Bayo & López-Castellanos (2016)[14]; Aguado and Rosen, (2008)[15]; Ismail et al.(2013)[16], monitoring the quality of rivers (Al-Badaii et al. (2013)[17], Dabgerwal & Tripathi, (2016)[18], Alves et al. (2018) [19], and studying the distribution of organic pollutants Zhang et al. (2013).[20].

This study has been performed to study the levels changes of different chemical and physical parameters due to the discharge of the city wastes and verify its compliance with the Iraqi standards and WHO standards; also, the quality of the Tigris River water for irrigation purposes was evaluated. Finally, PCA and HCL were performed to explore a reduction in data that enables the grouping of water quality variables into selected factors with common features to describe the plants' behavior in terms of pollution sources and treatment efficiency.

2. Material and Methods
2.1 Sample collection and dataset

Data records of physicochemical parameters of raw water and treated water for the period 2013-2016 were obtained monthly from eight water treatment plants (WTPs) located in Baghdad. The main characteristics of the WTPs are summarized in Table 1 below and the location of the mentioned WTPs presented in figure 1.

| Plant   | Extension Date | Capacity (million L/day) | Distance from a reference point | Average Percentage of Available Capacity | Percentage of Supplying Baghdad City |
|---------|----------------|--------------------------|-------------------------------|----------------------------------------|--------------------------------------|
| Al-Karkh| 1986           | 1365                     | 0                             | 82%                                    | 55.8%                                 |
| Year       | Name        | Median Flow | Total Flow | Treatment Efficiency | Loss Efficiency |
|------------|-------------|-------------|------------|----------------------|-----------------|
| 1988       | Sharq-Dijlah| 30          | 540        | 74%                  | 20%             |
| 1985       | Al-Karama   | 40.76       | 200        | 94%                  | 7.5%            |
| 1961       | Al-Wathba   | 34.47       | 60         | 90%                  | 3.5%            |
| 1978       | Al-Qadisiya | 55.02       | 135        | 84%                  | 4.9%            |
| 1982       | Al-Dora     | 58.78       | 105        | 76%                  | 3.8%            |
| 1952       | Al-Wehda    | 64.15       | 50         | 83%                  | 2.8%            |
| 1969       | Al-Rasheed  | 67.09       | 45         | 77%                  | 1.7%            |

Figure 1 location of water treatment plants.
following physicochemical water quality parameters were included in the dataset: pH, alkalinity (Alk.), turbidity (Turb.), total dissolved solids (TDS), total hardness (TH), calcium (Ca²⁺), magnesium (Mg²⁺), chloride (Cl⁻), sulfate (SO₄²⁻), aluminum (Al³⁺), total dissolved solids (TDS), temperature (Temp.), electrical conductivity (E.C), Nitrite (NO₂⁻), Nitrate (NO₃⁻), Ammonia (NH₃), Iron (Fe³⁺), fluoride (F⁻), Silica as (SiO₂), and Orthophosphate (PO₄³⁻).

3. Data analysis
Approximately 1672 data points of both raw water and treated water were subjected to descriptive statistical analysis using Microsoft Excel 2019 and SPSS (Statistical Package for Social Sciences) 26.0. Table 3 and Table 4 represent mean and standard deviation corresponding to the eight water treatment plants' measured parameters. Pearson correlation test was performed to investigate the possible relationship between the physicochemical parameters of water.

**Principal factor analysis (PFA)** was conducted to determine the significant parameters that describe the variation in the total data set. PCA is a dimension reduction tool used to identify a smaller number of uncorrelated variables (principal components) from a large set of data with minimum loss of original information Vialle et al., (2011)[21]. Kaiser-Meyer-Olkin (KMO) test of sampling adequacy was used to study the data's fit degree to factor analysis. If KMO above 0.5, this indicates the Correlation matrix should and can be factored analyzed. Kaiser et al., (1974)[22] suggested guidelines for KMO value, and it does have below 0.5 is unacceptable. KOM value in 0.5 level is miserable, they also write that KOM in the 0.6 level is acutely mediocre, even in the KOM 0.7 level is middling, in the 0.8 level is meritorious and certainly in the 0.9 level is marvelous. Bartlett’s test of sphericity was performed to examine the null hypothesis that the resulting correlation matrix is an identity matrix. The PCA was calculated using the following steps: 1. Standardize the variable data to make sure they all have equal weights in the analysis. 2. This step followed by Calculation of the covariance matrix. 3. Find the eigenvalues and the corresponding vectors. 4. Eliminate any components that contribute to only a small percentage of the difference in data sets. 5. Establish the factor loading matrix and conduct a rotation of varimax on the factor loading matrix to conclude the principal stations. Manly, (1986)[23].

**Hierarchical Clustering (CA)** is a multivariate approach used to see how similar or distinct each plant is from others by considering each plant is their own cluster and iteratively merge the similar clusters until all items belong to one cluster using agglomeration or bottom-up approach. (Wunderlin et al., 2001)[24]. The dendrogram pictorially represents the result of HCL. (Shrestha and Kazama, 2007)[25]. The dendrogram is a tree-like structure represent the HCL techniques. The similarity of two objects is shown by low linkage distance, whereas a large distance shows the dissimilarity. Huang et al., (2014)[26]. In this study, owing to wide variations in data dimensionality, the data was standardized through z scale transformation. Standardization eliminates classification errors and eliminates the effect of various measurement units, and makes the data dimensionless. The Ward method was used to evaluate distances between clusters as an ANOVA approach, and the means of squared Euclidean Distances as a measure of the differences between analytical values from two separate samples.

4. Results and discussion
4.1 Data analysis
The descriptive statistics were developed after the data consistency analysis to verify compliance with the IQS and WHO standards in Table 2. Table 3 and Table 4 show the results for raw water and treated water parameters under each WTP study, respectively.
4.1.1 PH
For raw water data, the PH was ranged from 8.04 to 7.33. It shows the water at the Tigris river is neutral to sub-alkaline. As most crops can withstand a pH range of 4.5 to 9.0 Lund, (1971)[27], so it is evident from these findings that issues with irrigating with such water are unlikely to be experienced, as far as the pH is concerned. The PH value for drinking water was within permissible for both WHO and IQS.

4.1.2 Total dissolved solids
Apart from Al-Karkh WTP, there was an increase in TDS concentration throughout the other WTPs. These substances are introduced to the water from agricultural, industrial, and domestic waste. The obtained results of this study show that the overall average concentration of TDS is 568.037 mg/l, which put it in class I concerning its impact on crop yield (no problem) and class II (increasing problem) regarding its impact on soil permeability according to FAO standards [28]. Irrigation with such water should post no problem on crops; however, the way the Iraqi farmer practices irrigation leads to built-up salts in the soil. They are applying a large amount of water and paying no attention to water use for different crops. As the temperature goes up to 50 Cº during the summer, water evaporation from the soil becomes critical, leaving the salt behind Mutlak (1980)[29]. For drinking water, the current study of drinking water matched the Iraqi standards for safe drinking water.

4.1.3 Turbidity
Turbidity showed a wide range of fluctuations along the studied period reaching its maximum value at Al-Wehda WTP (105.125 NTU). The higher values of Al-Wehda and Al-Rasheed, especially in winter is probably due to erosion from upland caused by rainfall, organic matter from sewage discharges during vegetable oil plant and electrical power plant bypass. phosphorus from different sources can increase algae growth, resulting in increased turbidity. Sources of phosphorus may include treatment facilities for wastewater, nutrient runoff from cropland and other sources. High turbidity can drastically reduce the river aesthetic quality, harming recreation and tourism. It can raise the cost of drinking water treatment, affecting the irrigation, and damage fish and other aquatic life via reducing food sources, preventing the efficient growth of eggs and larvae of fish; and influencing gill function. Quality (2008)[30], Hassan & Mahmood (2018)[31], Al-Obaidi, (2009)[32], Abdulhay et al. (2018)[33]. For drinking water, the result did not exceed the limits permitted by the WHO standard as the highest turbidity rate was reached at Al-Wehda (3.25 NTU). The increase and decrease in turbidity levels depend on the contents of the river water in terms material causing turbidity, the age of the project, the efficiency of operation and maintenance of the project, as well as the water consumption by citizens in quantities more than the productive capacity of the project as the water does not have sufficient time to stagnate in the sedimentation basins or use low-quality alum. (Mahmood, 2018)[34]. The results agreed with a study conducted by Al-Fatlawy, (2001)[35] who indicated during his study that Al-Wehda project has the least filtration efficiency, compared to the rest of the WTPs in Baghdad.

4.1.4 Alkalinity
Except for Sharq-Dijlah and Al-Karkh, the alkalinity of raw water exhibits high concentration above the standard limits. Higher levels of alkalinity in surface waters can buffer acid rain and other acid waste and stop changes in pH that are detrimental to marine life. Oram, (2020)[36]. For drinking water, the average alkalinity exceeded the permissible limit at Al-Karama water treatment plant, whereas the average other parameters were within the limit.
4.1.5 Total hardness
The hardness concentration was slightly increased as the water passes through Baghdad, ranging from 282.167 to 340.250 mg/l. The results obtained from this study for drinking water was ranged from 284.188 to 337.542 mg/l and accomplice with the WHO standards.

4.1.6 Calcium and Magnesium
The overall average Ca^{2+} and Mg^{2+} concentrations in Tigris river were 81.26 and 29.34 mg/l respectively. The results present in Table 2 shows that there was a slight increase in Mg^{2+} and Ca^{2+} concentration in the river due to city discharge. These ions are beneficial when the water is used for irrigation purposes. (Rengasamy, 1987)[37]

4.1.7 Iron and Fluoride
The overall mean concentrations of Fe^{2+} and F are 1.724 and 0.116 mg/l, respectively. The average concentration is within the permissible limit. The mean concentrations obtained are unlikely to affect crops and irrigation purposes. The recommended concentration of Fe and F by the Committee on Water Quality Criteria[38], are 5 and 1 mg/l respectively for water continuously used on the soil. These levels may be significant in Iraq, where the PH of the soil is above 7.

4.1.8 Inorganic nitrogen
The results show that the concentration of NO_{2}^{-}, NO_{3}^{-}, and NH in raw water was found to be 0.00918, 0.959, and 0.182 mg/l respectively. In contrast, the average concentration of these components in treated water was 0.0016, 1.0365, and 0.0268 mg/l Consecutive, and these values are within the limits of the Iraqi specifications. An increase in the concentration of these elements beyond the permissible limit in drinking water may cause Methemoglobinemia and cyanosis in children young age. Wall (2013)[39]

4.1.9 Chloride and sulfate
The average concentration of Cl and SO_{4} in raw water 68.25 and 197.12 mg/l, respectively. Both Cl and SO_{4} exhibits variation, which is likely due to the variation in the river discharge. In drinking water 69.1 and 198.62 mg/l. It is evident from these finding that the concentration of Cl and SO_{4} in drinking is more significant than in raw water. This increase is caused by the adding of alum and chlorine to the water. Despite this increase in the concentration of SO_{4} and Cl in water, it remains with the WHO limitations.

4.1.10 Aluminium
The Aluminium concentration in water was ranged from 0.01 to 0.016 mg/l and in treated water ranged from 0.046 to 0.113 mg/l. The increasing in aluminium concentration in drinking water depends on its concentration in raw water and the use of alum as a coagulant. (WHO, 1989)[40]

1.1.9 Phosphate and Silica
The PO_{4} concentration has varied from 0.01 mg/l at Sharq-Dijlah to 0.43 mg/l at al-wehda. All value of PO_{4} has not exceeded the permissible value. The average concentration of silica in raw water and treated water was 4.40451, 4.075 respectively mg/l which was almost the same. The SiO2 concentration have doubled in Al-Wehda as compared with Sharq-Dijlah. metal silicate in the earth crust. Due to its complexity, silica fouling is difficult to control. (Park, 2020)[41]
| Parameter          | Symbol | Unit            | WHO, 4th edition, 2011 | IQS/417, 2009 |
|--------------------|--------|-----------------|------------------------|--------------|
| Turbidity          | Turb.  | NTU             | 5                      | 5            |
| Alkalinity         | Alk.   | (mg/l) as CaCO$_3$ | 150                   | 150          |
| Total Hardness     | T.H    | (mg/l) as CaCO$_3$ | 500                   | 500          |
| Calcium            | Ca$^{+2}$ | (mg/l)       | 200                   | 150          |
| Chloride           | Cl$^-$ | (mg/l)         | 250                   | 350          |
| Magnesium          | Mg$^{+2}$ | (mg/l)     | 150                   | 100          |
| Iron               | Fe$^{+2}$ | (mg/l)     | 0.3                   | 0.3          |
| PH                 | PH     | -              | 6.5-8.5               | 6.5-8.5      |
| Electric conductivity | E.C   | (µS/cm)       | 1000                  | 1500         |
| Sulfate            | SO$_4^{2-}$ | (mg/l)    | 250                   | 400          |
| Total dissolved solids | TDS   | (mg/l)         | 500                   | 1000         |
| Ammonia            | NH$_3$ | (mg/l)         | 0.2                   | 0.5          |
| Orth-Phosphate     | PO$_4^{3-}$ | (mg/l) | 1                      | 1            |
| Nitrite            | NO$_2^-$ | (mg/l)       | 3                      | 3            |
| Nitrate            | NO$_3^-$ | (mg/l)       | 50                     | 50           |
| Fluoride           | Fl     | (mg/l)         | 1                      | 1.5          |
| Aluminum           | Al$^{+3}$ | (mg/l) | 0.1                   | 0.2          |
| Parameter | Al-Karkh | Sharq-Dijlah | Al-Watha | Al-Karanna | Al-Qadisiya | Al-Dora | Al-Rasheed | Al-Wahda | Overall Mean |
|-----------|----------|-------------|----------|------------|-------------|---------|------------|---------|--------------|
| Mean      | 23.000   | 22.413      | 6.389    | 21.771     | 5.347       | 21.583  | 5.434      | 23.021  | 6.086        |
| SD        | 4.780    | 4.287       | 8.058    | 6.534      | 6.040       | 69.331  | 96.625     | 105.125 | 142.992      |
| Turb.     | 88.583   | 88.848      | 90.393   | 35.467     | 67.354      | 80.458  | 64.000     | 144.125 | 150.125      |
| Mean      | 12.269   | 11.180      | 14.921   | 12.978     | 15.950      | 12.078  | 15.520     | 15.833  | 12.225       |
| SD        | 1.191    | 1.287       | 8.278    | 11.961     | 10.924      | 15.833  | 10.942     | 15.393  | 14.341       |
| Alk.      | 282.167  | 321.174     | 46.037   | 35.333     | 46.946      | 327.083 | 46.342     | 325.500 | 45.716       |
| Mean      | 64.875   | 67.453      | 10.786   | 13.493     | 13.715      | 82.750  | 12.042     | 85.298  | 13.318       |
| SD        | 6.781    | 76.435      | 10.786   | 8.396      | 13.715      | 82.750  | 12.042     | 85.298  | 13.318       |
| T.H       | 45.896   | 67.543      | 15.284   | 66.521     | 14.489      | 74.292  | 14.433     | 76.854  | 15.648       |
| Mean      | 28.083   | 31.717      | 5.277    | 30.875     | 4.551       | 30.646  | 4.970      | 29.833  | 4.892        |
| SD        | 9.087    | 8.009       | 0.048    | 8.012      | 0.068       | 8.042   | 0.061      | 7.883   | 0.193        |
| PH        | 7.879    | 8.009       | 8.012    | 8.042      | 7.883       | 7.991   | 0.103      | 7.881   | 0.080        |
| E.C       | 712.917  | 853.217     | 133.915  | 876.250    | 138.433     | 876.750 | 135.890    | 905.042 | 142.794      |
| Mean      | 119.396  | 182.196     | 41.468   | 207.188    | 54.068      | 213.146 | 40.936     | 230.438 | 53.102       |
| SD        | 37.999   | 41.468      | 41.468   | 207.188    | 54.068      | 213.146 | 40.936     | 230.438 | 53.102       |
| TDS       | 454.271  | 64.351      | 572.174  | 88.863     | 587.229     | 92.471  | 572.625    | 98.133  | 573.688      |
| Mean      | 0.911    | 0.747       | 2.111    | 1.069      | 1.976       | 1.875   | 1.102      | 0.353   | 2.596        |
| SD        | 0.071    | 0.048       | 0.048    | 0.048      | 0.048       | 0.048   | 0.023      | 0.023   | 0.111        |
| Fe²⁺      | 0.097    | 0.018       | 0.100    | 0.009      | 0.100       | 0.009   | 0.011      | 0.011   | 0.011        |
| Mean      | 0.013    | 0.020       | 0.010    | 0.009      | 0.010       | 0.013   | 0.011      | 0.011   | 0.011        |
| SD        | 0.004    | 0.020       | 0.007    | 0.008      | 0.008       | 0.006   | 0.006      | 0.006   | 0.006        |
| NO₃⁻      | 1.006    | 0.298       | 0.985    | 0.349      | 1.162       | 0.318   | 0.997      | 0.294   | 0.938        |
| Mean      | 0.025    | 0.023       | 0.199    | 0.116      | 0.044       | 0.023   | 0.008      | 0.163   | 0.183        |
| SD        | 0.349    | 0.349       | 0.349    | 0.349      | 0.349       | 0.349   | 0.349      | 0.349   | 0.349        |
| NH₄⁺      | 2.333    | 0.884       | 2.590    | 0.691      | 3.131       | 0.929   | 5.333      | 0.835   | 3.267        |
| Mean      | 0.400    | 0.040       | 0.040    | 0.018      | 0.007       | 0.077   | 0.042      | 0.016   | 0.030        |
| SiO₂⁺     | 5.784    | 5.784       | 5.784    | 5.784      | 5.784       | 5.784   | 5.784      | 5.784   | 5.784        |
| Mean      | 0.101    | 0.138       | 0.138    | 0.138      | 0.138       | 0.138   | 0.138      | 0.138   | 0.138        |
| PO₄³⁻     | 0.430    | 0.430       | 0.430    | 0.430      | 0.430       | 0.430   | 0.430      | 0.430   | 0.430        |

Table 2. Descriptive statistics of the physicochemical parameter of raw water data for each WTP.
### Table 3. Descriptive statistics of the physicochemical parameter of drinking water data for each WTP.

| Parameter | Al-Karkh | Shaq-Dijlah | Al-Warba | Al-Karama | Al-Qadisiya | Al-Dora | Al-Rasheed | Al-Wahda | Overall mean |
|-----------|----------|-------------|---------|-----------|-------------|---------|------------|----------|--------------|
| Mean      | 22.813   | 22.435      | 22.313  | 6.488     | 21.917      | 6.171   | 22.188     | 5.417    | 23.479       |
| SD        | 4.841    | 4.400       | 4.400   | 6.888     | 6.171       | 6.171   | 5.417      | 6.039    | 6.502        |
| Turb.     | 1.304    | 0.561       | 3.035   | 1.118     | 2.117       | 0.486   | 2.733      | 0.734    | 2.379        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| Alk.      | 119.896  | 17.389      | 139.783 | 7.636     | 140.375     | 12.189  | 155.417    | 11.594   | 148.396      |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| T.H       | 284.188  | 31.545      | 320.717 | 45.854    | 326.771     | 45.795  | 337.542    | 46.239   | 328.854      |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| Ca²⁺      | 65.500   | 6.798       | 76.717  | 10.901    | 83.792      | 13.451  | 84.604     | 13.628   | 83.458       |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| Cl⁻       | 46.708   | 10.040      | 67.804  | 15.260    | 69.167      | 15.323  | 74.875     | 14.441   | 71.333       |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| Mg²⁺      | 28.125   | 3.868       | 31.966  | 5.278     | 30.667      | 4.674   | 30.625     | 4.854    | 29.813       |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| PH        | 7.548    | 0.073       | 7.586   | 0.087     | 7.532       | 0.085   | 7.740      | 0.074    | 7.658        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| E.C       | 718.729  | 91.790      | 855.022 | 133.592   | 885.979     | 140.580 | 882.146    | 137.813  | 873.333      |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| SO₄²⁻     | 124.646  | 38.524      | 182.935 | 40.603    | 209.042     | 53.051  | 214.875    | 40.459   | 235.375      |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| TDS       | 459.563  | 63.726      | 573.739 | 88.474    | 593.958     | 94.112  | 576.417    | 98.287   | 574.917      |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| Fe²⁺      | 0.069    | 0.071       | 0.167   | 0.107     | 0.078       | 0.051   | 0.165      | 0.066    | 0.048        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| F         | 0.098    | 0.019       | 0.092   | 0.010     | 0.087       | 0.030   | 0.157      | 0.021    | 0.091        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| Al³⁺      | 0.097    | 0.046       | 0.113   | 0.041     | 0.079       | 0.029   | 0.061      | 0.001    | 0.008        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| NO₂⁻      | 0.001    | 0.001       | 0.001   | 0.001     | 0.001       | 0.003   | 0.001      | 0.001    | 0.001        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| NO₃⁻      | 1.112    | 0.297       | 1.152   | 0.344     | 0.864       | 0.364   | 0.962      | 0.257    | 1.054        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| NH₄⁺      | 0.031    | 0.143       | 0.011   | 0.005     | 0.020       | 0.001   | 0.010      | 0.005    | 0.019        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| SO₄²⁻     | 4.246    | 0.652       | 2.338   | 0.713     | 2.608       | 1.154   | 4.627      | 0.806    | 3.167        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |
| PO₄³⁻     | 0.484    | 0.052       | 0.010   | 0.034     | 0.011       | 0.004   | 0.042      | 0.019    | 0.053        |
| Mean      |          |             |         |           |             |         |            |          |              |
| SD        |          |             |         |           |             |         |            |          |              |

Notes: Mean: Average value; SD: Standard Deviation; Temp: Temperature; Turb: Turbidity; Alk: Alkalinity; T.H: Total Hardness; Ca²⁺: Calcium; Mg²⁺: Magnesium; E.C: Electrical Conductivity; SO₄²⁻: Sulfate; TDS: Total Dissolved Solids; Fe²⁺: Iron; F: Fluoride; Al³⁺: Aluminum; NO₂⁻: Nitrite; NO₃⁻: Nitrate; NH₄⁺: Ammonium; SO₄²⁻: Sulfate; PO₄³⁻: Phosphate.
4.2 Correlation analysis

Among the various physicochemical parameters of water, only a few parameters showed significant correlations. In the present study, the values of total hardness (TH) shown a statistically significant positive correlation with calcium (.912**), magnesium (.759**), chloride (.891**), E.C (0.944**) SO4 (0.821), and total dissolved solids (.911**). The electric conductivity in water has exhibited a strong positive relationship with total dissolved solids (.964**), chloride (.942**), calcium (.852**), and moderate correlation magnesium, according to Singh et al., (2015)[42]. The increase in impurities concentration in water leads to an increase in the number of free ions, causing an increase in EC. The total dissolved solids in water in the present study have a significant correlation with total hardness (.911**), Chloride (.915**), Calcium (.852**), SO4 (.863**), and E.C (.964**) and moderate correlation with Magnesium (0.687). This means with an increase in the total dissolved values, the values of chloride, magnesium, conductivity, and SO4 increase. PH was poorly correlated with Ca and Cl and negatively correlated with the temperature. The chloride concentration in water has a positive correlation with calcium (.819**), magnesium (.680**).

|     | Temp | Turb | Alk | T.H  | Ca⁺² | Mg²⁺ | PH  | E.C | SO₄ | TDS | F   | NO₂ | NO₃ | NH₃ | SiO₂ | PO₄ |
|-----|------|------|-----|------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Temp|      | 1    |     |      |      |      |     |     |     |     |     |     |     |     |     |     |
| Turb| -1.48**| 1    |     |      |      |      |     |     |     |     |     |     |     |     |     |     |
| Alk | -0.261**| 0.163**| 1   |      |      |      |     |     |     |     |     |     |     |     |     |     |
| T.H | -0.616**| 0.024 | 0.216**| 1   |      |      |     |     |     |     |     |     |     |     |     |     |
| Ca⁺²| -0.527**| -0.005 | 0.285**| 0.912**| 1  |     |     |     |     |     |     |     |     |     |     |     |
| Cl  | -0.506**| 0.106 | 0.336**| 0.891**| 0.819**| 1  |     |     |     |     |     |     |     |     |     |     |
| Mg²⁺| -0.524**| 0.035 | 0.071 | 0.759**| 0.452**| 0.680**| 1  |     |     |     |     |     |     |     |     |     |
| PH  | -0.236**| 0.028 | 0.266**| 0.175**| 0.113**| 0.140**| 0.229**| 1  |     |     |     |     |     |     |     |     |
| E.C | -0.586**| 0.030 | 0.236**| 0.944**| 0.884**| 0.942**| 0.698**| 0.141**| 1  |     |     |     |     |     |     |     |
| SO₄ | -0.440**| -0.031 | 0.274**| 0.821**| 0.803**| 0.847**| 0.585**| 0.083 | 0.855**| 1  |     |     |     |     |     |     |
| TDS | -0.555**| 0.012 | 0.297**| 0.911**| 0.852**| 0.915**| 0.687**| 0.125 | 0.964**| 0.863**| 1  |     |     |     |     |     |
| F   | -0.024 | 0.217**| 0.118**| 0.026 | 0.133**| 0.134**| -0.160**| -0.144**| 0.086 | 0.082 | 0.111**| 1  |     |     |     |     |
| NO₂ | -0.056 | 0.140**| 0.249**| 0.104 | 0.149**| 0.171**| -0.013 | 0.025 | 0.147**| 0.092 | 0.165**| 0.165**| 1  |     |     |     |
| NO₃ | -0.201**| 0.168**| 0.167**| 0.007 | -0.087 | -0.076 | 0.181**| 0.335**| -0.035 | -0.070 | -0.054 | -0.569**| -0.008 | 1   |     |     |
| NH₃ | -0.094 | 0.111**| 0.271**| 0.145**| 0.307**| 0.216**| -0.083 | 0.214**| 0.164**| 0.236**| 0.644**| 0.265**| -0.476**| 1   |     |     |
| SiO₂| 0.074 | 0.066 | 0.070 | -0.023 | -0.073 | -0.012 | 0.026 | 0.102 | -0.058 | -0.073 | -0.098 | -0.580**| -0.063 | 0.468**| -0.438**| 1   |
| PO₄ | -0.008 | 0.124**| 0.115**| 0.007 | 0.157**| 0.098 | -0.249**| -0.219**| 0.070 | 0.055 | 0.101**| 0.850**| 0.222**| -0.679**| 0.738**| -0.683**| 1   |

***: Correlation is significant at the 0.01 level.
4.3 Multivariate analyses of raw and treated water

The dendrograms in figure 2 illustrated the categorization of the water treatment plants based on the similarity of the levels of the physicochemical parameters of raw water and treated water. Ward's method and the square Euclidean distance were used as a measure of dissimilarity.

Al-Karkh differed from the other WTPs and remain isolated while the other plants formed a larger group. Sharq-Dijlah, Al-Wathba, and Qadisiya got paired up as a sub-cluster. It is possible to notice that the smallest Euclidean distance between these three plants, they shared the same characteristic features and the same source of contaminations. The three WTPs joined Al-Dora and Alkarma to form one cluster.

The raw water of these plants is affected by urban wastes and anthropogenic activities in a mediocre state. Again, Al-Rasheed joined along with Al-Wehda WTP and formed cluster 2. The raw water quality at AL-Wehda WTP is affected by the wastes discharged into the river from the vegetable oil plant. Also, the raw water at Al-Rasheed plant was influenced by the impact of the south Baghdad electrical power station and AL-Rasheed electrical power station. Cluster 1 and cluster 2 are combined with Al-Karkh to form cluster 3.

The dendrogram in figure 3.b grouped the data of drinking water into three clusters. AL-Karkh also remains isolated while the other plans grouped because they presented slightly lower performance. Al-Qadisiya, Al-Dora, formed sub-cluster and joined Al-Wathba, Al-Karama, and Sharq-Dijlah to cluster 1, which includes relatively less polluted water. The second cluster contains Al-Rasheed and Al-Wehda. Both these plants are old, and witness no attempted to expand since 2008. In contrast, the third cluster formed by joining cluster 1, cluster 2, and Al-Karkh WTP.
4.4 Principal component analysis
The data were standardized, and the KMO test and Bartlett sphericity test were calculated, which was 0.795, and the Bartlett sphericity test is smaller than 0.001, which indicates the feasibility for PCA.
It can be seen from the scree plot in figure 3 that only six principal components have eigenvalues greater than unity and explain 73.389% of the total variance in the data set. The first component is accounted for about 34.018% of the total variance with high loading on T.H (.961), Ca$^{2+}$ (.905), Cl$^-$ (.942), Con. (.978), SO$^4$ (.902), TDS (.963) and Mg (.719). The high loading of these parameters is related to weathering and an agricultural source of SO$^4$ in surrounding farmlands.
The second component accounted for about 9.696% of the total variance and have a moderate positive loading for Fluoride (0.775) and PH (0.679). A high level of fluoride in water may cause the utmost prevalence of dental fluorosis.
The third PCA features high and moderate loading (.844), (.763), corresponding turbidity, and iron with nearly 9% variance. High iron and turbidity levels are related to soil erosion caused by surface runoff, and agricultural runoff also is related
Strong and moderate loading factors of NO$^3$ (.879) and alkalinity (.604) dominate the fourth component, representing about 8.027% of the variance. The source of NO$^3$ in water is the eutrophication of the dissolved nitrate.
The fifth component contains PO$^4$ as the most significant variable and signifying 6.919% of the variance. The sources of Phosphate are fertilizers, pesticides, industry, and cleaning compounds.
Lastly, the sixth component accounts for 6.578% of the total variance was highly and moderately correlated to SIO$^2$ (.847) and NO$_2$ (.66), respectively.
Table 6. Factor loading for the selected parameter in drinking water.

Rotated Component Matrix\textsuperscript{a}

| Component | 1     | 2     | 3     | 4     | 5     | 6     |
|-----------|-------|-------|-------|-------|-------|-------|
| Temp.     | -0.596| -0.109| 0.102 | -0.507| 0.181 | 0.089 |
| Turb.     | 0.129 | -0.012| \textbf{0.844}| -0.044| 0.018 | 0.018 |
| Alka.     | 0.201 | 0.074 | 0.416 | \textbf{0.604}| 0.35  | -0.043|
| T.H       | \textbf{0.961}| 0.103 | 0.038 | 0.051 | -0.069| 0.012 |
| Ca\textsuperscript{2+} | \textbf{0.905}| -0.086| 0.053 | 0.023 | 0.118 | 0.098 |
| Cl\textsuperscript{-}  | \textbf{0.942}| 0.037 | 0.133 | 0.038 | 0.051 | -0.032|
| Mg\textsuperscript{2+}  | \textbf{0.719}| 0.337 | 0.024 | 0.064 | -0.347| -0.152|
| PH        | 0.044 | \textbf{0.679}| -0.032| 0.149 | -0.257| -0.057|
| E.C       | \textbf{0.978}| -0.018| 0.031 | 0.038 | -0.027| 0.042 |
| SO\textsubscript{4}\textsuperscript{2-}  | \textbf{0.902}| 0.045 | 0.031 | -0.046| 0.105 | 0.01  |
| TDS.      | \textbf{0.963}| -0.068| 0.026 | 0.026 | -0.036| 0.047 |
| Fe\textsuperscript{2+}  | -0.032| 0.072 | \textbf{0.763}| 0.074 | -0.159| 0.063 |
| F         | 0.013 | \textbf{0.775}| 0.146 | 0.017 | -0.167| -0.167|
| Al\textsuperscript{3+}  | -0.432| 0.136 | 0.09  | -0.257| -0.364| -0.051|
| NO\textsubscript{2}\textsuperscript{-} | 0.087 | -0.36  | 0.056 | 0.119 | -0.039| \textbf{0.66}|
| NO\textsubscript{3}\textsuperscript{-} | -0.06 | 0.008 | -0.024| \textbf{0.879}| -0.044| 0.079 |
| NH\textsubscript{3}      | -0.027| -0.671| 0.012 | 0.065 | -0.164| -0.067|
| SiO\textsubscript{2}     | -0.016| 0.094 | 0.034 | -0.044| 0.092 | \textbf{0.847}|
| PO\textsubscript{4}\textsuperscript{3-} | -0.068| -0.105| -0.105| -0.016| \textbf{0.844}| 0.044 |
| Variance (%) | 34.018| 9.696 | 8.15  | 8.027 | 6.919 | 6.587 |
| Cumulative | 34.018| 43.713| 51.863| 59.89 | 66.808| 73.396|

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. \textsuperscript{a} Rotation converged in 6 iterations.

Figure 3 Scree plot of eigenvalues
5. Summary and Conclusion

This study was undertaken to evaluate the performance of eight water treatment plant in Baghdad city. The outcomes showed that the multivariate statistical techniques helped identify the significant input from a large data set that affects drinking water quality. In this study, the PCA in Table 6 highlighted 11 out of 19 as the Main contamination source in drinking water. The parameters including turbidity Cl, PH, SO4, TDS, E.C, Fl, Fe, NH3, SiO2, and PO4, are responsible for the variation of drinking water quality, which mainly introduced to water from weathering of minerals, agricultural waste, naturally existing organic matter, and domestic waste from the human activities. The hierarchical cluster analysis grouped the WTPS into three clusters with similar raw water and treated water clusters. The dendrogram illustrated the merging process. This merging process showing not only which data were joined at each point but when they merge in order of increasing cluster distance given the sequence of merges. Three clusters were formed corresponding to the similarity in raw water data cluster 1 contains Sharq-Dijlah, Al-Wathba, Qadisiya, Al-Karama, and Al-Dora. Cluster 2 contains Al-Rasheed and Al-Wehda. Cluster 3 joined Al-Karkh to cluster one and cluster 2. Regarding the similarity of the performance of the WTPs grouped into three clusters. Cluster 1 contains Al-Qadisiya, Al-Dora, Al-Wathba, Sharq-Dijlah, and AL-Karama. Cluster 2 contains Al-Wehda and Al-Wathba. Cluster 3 containing cluster 1, cluster 2, and Al-Karkh WTP.

6. References:
[1] ChildInfo, “Iraq - Multiple Indicator Cluster Survey 2011 - Sampling,” no. April, 2014.
[2] R. C. Marques and A. J. Monteiro, “Application of performance indicators in water utilities management - A case-study in Portugal,” Water Sci. Technol., vol. 44, no. 2–3, pp. 95–102, 2001.
[3] E. Perotto, R. Canzianzi, R. Marchesi and P. Butelli, "Environmental performance, indicators and measurement uncertainty in EMS context: a case study", Journal of Cleaner Production, vol. 16, no. 4, pp. 517-530, 2008.
[4] Ofwat, “Updating the overall performance assessment (OPA) - Conclusions and methodology for 2004-05 onwards March 2004,” no. March, 2004.
[5] P. Vieira, H. Alegre, M. Rosa and H. Lucas, "Drinking water treatment plant assessment through performance indicators", Water Supply, vol. 8, no. 3, pp. 245-253, 2008.
[6] R. Sadiq, M. Rodriguez and S. Tesfamariam, "Integrating indicators for performance assessment of small water utilities using ordered weighted averaging (OWA) operators", Expert Systems with Applications, vol. 37, no. 7, pp. 4881-4891, 2010.
[7] E. Chang, P. Chiang, S. Huang and Y. Lin, "Development and Implementation of Performance Evaluation System for a Water Treatment Plant: Case Study of Taipei Water Treatment Plant", Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management, vol. 11, no. 1, pp. 36-47, 2007.
[8] A. Quevedo-Castro, J. Rangel-Peraza, E. Bandala, L. Amabilis-Sosa, A. Rodriguez-Mata and Y. Bustos-Terrones, "Developing a water quality index in a tropical reservoir using a measure of multiparameters", Journal of Water, Sanitation and Hygiene for Development, vol. 8, no. 4, pp. 752-766, 2018.
[9] R. Olawoyin, B. Heidrich, S. Oywole, O. T. Okareh, and C. W. McGlothlin, “Chemometric analysis of ecological toxicants in petrochemical and industrial environments,” Chemosphere, vol. 112, pp. 114–119, 2014.
[10] H. M. Flaieh and M. Y. Abdul-ahad, “Assessing Tigris River Water Quality in Baghdad City Using Water Quality Index and,” Int. J. Eng. Sci. Res. Technol., vol. 3, no. 7, pp. 687–699, 2014.
[11] L. D. V. Melo, E. P. da Costa, C. C. Pinto, G. R. Barroso, and S. C. Oliveira, “Adequacy analysis of drinking water treatment technologies in regard to the parameter turbidity, considering the quality of natural waters treated by large-scale WTPs in Brazil,” Environ. Monit. Asses., vol. 191, no. 6, 2019.
[12] H. M. Issa and R. A. Alrwai, “Long-Term Drinking Water Quality Assessment Using Index and Multivariate Statistical Analysis for Three Water Treatment Plants of Erbil City, Iraq,” UKH J. Sci. Eng., vol. 2, no. 2, pp. 39–48, 2018.
[13] Z. A. A. L. Saad and A. N. A. Hamdan, “Evaluation of water treatment plants quality in Basrah Province , by factor and cluster analysis,” 2020.
[14] J. Bayo and J. López-Castellanos, “Principal factor and hierarchical cluster analyses for the performance assessment of an urban wastewater treatment plant in the Southeast of Spain,” Chemosphere, vol. 155, pp. 152–162, 2016.
[15] D. Aguado and C. Rosen, “Multivariate statistical monitoring of continuous wastewater treatment plants,” *D. Aguado C. Rosen, “Multivariate Stat. Monit. Contin.”*, vol. 21, no. 7, pp. 1080–1091, 2008.

[16] A. H. Ismail, “Cluster analysis of some ecological properties in Al-Rustamiyah wastewater treatment plant 3 rd expansion at Baghdad city, Iraq through (2006-2011),” *J. KerbalaUniversity*, vol. 11, no. 1, pp. 195