

# **Multivariate statistical-based surface water quality assessment of the Tigris and Diyala Rivers at their confluence in Baghdad**

Alhassan H Ismail<sup>1\*</sup>, Muntasir A Shareef<sup>1</sup>, Hatem H Hussein<sup>2</sup>, Havan Hassan Salman<sup>2</sup>, Nazar Kadhim Naama<sup>3</sup>, Ahmed Hatif Salim<sup>3</sup>, Anmar Ghalib Matar<sup>3</sup>

<sup>1</sup> Middle Technical University, Institute of Technology-Baghdad, Water Resources Techniques Department <sup>2</sup>Department of Planning and Follow-up / Ministry of Water Resources <sup>3</sup>National Center for Water Resources Management / Ministry of Water Resources, \*Corresponding author's email: [alhassan\\_hayder@mtu.edu.iq](mailto:alhassan_hayder@mtu.edu.iq)

# **Abs tract**

 Recently, the Tigris and Diyala rivers have deteriorated due to the discharge of inadequately treated wastewater into the river, an increase in water demand, and climate change. This makes Iraq suffer from providing water for different uses. In this paper, an attempt has been made to apply multivariate statistical methods, factor analysis (FA), to identify the primary factors and pollution sources affecting the water quality at the confluence of the Tigris and Diyala Rivers. The water quality of 16 parameters was considered in 35 sampling stations during two different periods. The water quality parameters are Turbidity, Total Hardness, Calcium, Magnesium, Sodium, Potassium, Chloride, pH, Electrical conductivity, Total dissolved solids, Sulfate, Carbonate, Bicarbonate, Iron, Chromium, Nitrate. The results of factor analysis specified three factors representing 80.810% of the total variance in each water quality dataset for the first period, whereas, for the second period, 2 factors were identified representing 77.402% of the total variance in each water quality dataset. The main differences are related to anthropogenic activities, in addition to some agricultural activities and hydrochemical effects in the study area. Multivariate statistical methods such as factor analysis can be used to understand the complex water quality data and identifying the source of pollution, in addition to provide a better explanation of the relationship between the large numbers of variables in surface water. FA can determine priorities to improve water quality and is believed to aid decision-makers in assessing water quality.

**Keywords:** Water quality Factor analysis, Multivariate statistical analysis, Tigris River, Diyala River.



# **تقييم جودة المياه السطحية متعدد المتغيرات على أساس إحصائي النهري دجلة وديالى عند التقائهما في بغداد**

الحسن حيدر اسماعيل!.\* منتصر عبد الحميد<sup>1</sup> حاتم حميد<sup>2</sup> هافان حسن<sup>2</sup> نزار كاظم احمد<sup>3</sup> هاتف انمار غالب<sup>3</sup> 1 الجامعة التقنية الوسطى، معهد تكنلوجيا بغداد، قسم تقنيات الموارد المائية 2 دائرة التخطيط والمتابعة /وزارة الموارد المائية 3 المركز الوطني الدارة الموارد المائية /وزارة الموارد المائية \* alhassan\_hayder@mtu.edu.iq :المرسل المؤلف

### **الخالصة**

في الأونة الأخيرة، تدهورت حالة نهري دجلة وديالي بسبب تصريف مياه الصرف الصحي المعالجة بشكل غير مناسب إلى النهر، وزيادة الطلب على المياه، وتغير المناخ. وهذا ما يجعل العراق يعاني من توفير المياه لمختلف االسببتخدامات. في هذا البحث، جرت محاولة لتطبيق األسباليب اإلحصبائية متعددة المتغيرات، مثل التحليل العاملي(FA (، لتحديد العوامل األولية ومصبادر التلوث التي تؤثر على نوعية المياه عند التقاء نهري دجلة وديالي. تم أخذ جودة المياه في الاعتبار من خلال 16 متغيرا في 35 محطة وتم أخذ عينات المياه خلال فترتين مختلفتين. معايير جودة المياه هي العكارة، الصبلابة الكلية، الكالسبوري، المغنيسبوري، الصبوديوري البوتاسبيوم، الكلوريد، الرقم الهيدروجيني، التوصبيل الكهربائي، المواد الصبلبة الذائبة الكلية، الكبريتات، الكربونات، بيكربونات، الحديد، الكروم، والنترات. حددت نتائج تحليل العوامل ثلاثة عوامل تمثل 80.810% من إجمالي التباين في كل مجموعة بيبانات لجودة المياه للفترة الأولى، بينما في الفترة الثانية، تم تحديد عاملين يمثالن %77.402 من إجمالي التباين في كل مجموعة بيانات لجودة المياه. وترتبط االختالفات الرئيسبية باألنشبطة البشبرية، بالإضبافة إلى بعض الأنشبطة الزراعية والتأثيرات الهيدروكيميائية في منطقة الدراسبة. ويمكن اسبتخدام الأسباليب الإحصبائية متعددة المتغير ات مثل تحليل العوامل لفهم بيانات نو عية المياه المعقدة وتحديد مصبدر التلوث، بالإضبافة إلى تقديم تفسير أفضيل للعلاقة بين الأعداد الكبيرة من المتغيرات في المياه السطحية. يمكن لـ FA تحديد الأولويات لتحسين جودة المياه ويُعتقد أنها تساعد صناع القرار في تقييم جودة المياه.

.ا**لكلمات المفتاحية**: جودة المياه، التحليل العاملي، الاحصـاء متعدد المتغيرات ، نهر دجلة، نهر ديالي



## **1. Introduction**

 Surface water quality in Iraq has deteriorated during the last recent years. Iraq relies on surface water to provide water for drinking, irrigation, and industry. Several factors have influenced the surface water quality deterioration such as water shortage, water scarcity, and climate change (Abbas, et al., 2016; Ghalib 2017; Ismail, et al., 2014). Water project from neighbouring countries has also affected the water quota of Iraq, for example, the South-eastern Anatolia Project (GAP) project in Turkey (Ismail , et al., 2020).

 Iraq has two main rivers namely, Tigris and Euphrates. The Tigris River flows through Baghdad (the capital of Iraq) dividing the city into two parts, the Karkh and Rusafa. On the other hand, the Diyala River joins the Tigris River at the south of Baghdad (Ismail , et al., 2019). The present study attempts to evaluate the water quality at the confluence points of the Tigris and Diyala Rivers. The latter is heavily polluted as stated by many researchers. Diyala River receives wastewater discharge from wastewater treatment plants located at its lower region before its confluence with Tigris River (Ismail & Muntasir 2018). Accordingly, the water quality of the Diyala River has deteriorated over the past decades and affected the water quality of the Tigris River, which necessitates the development of an appropriate management plan by the authorities to maintain the quality of the river (Ismail , et al., 2013).

 Numerous studies have been carried out to evaluate the water quality of the Tigris and Diyala Rivers at their confluence in Baghdad. This area is characterized by a high pollution load in the river due to the different point source pollution join the river near the confluence point. Furthermore, previous attempts have been conducted to simulate water quality in river (Ismail  $\&$ Muntasir 2018; Abed et al. , et al., 2021; Ismail , et al., 2022), or assess the water quality using water quality index technique (Sabeeh, et al., 2023). The present study uses the multivariate statistical technique to assess the water quality at the confluence point. Numerous studies conducted around the world used multivariate statistics techniques to assess the quality of rivers in different regions of the world (Wunderlin et al, et al., 2001; Simeonov , et al., 2004; Singh , et al., 2004; Boyacioglu & Boyacioglu, 2007; Chapagain , et al., 2010). According to our knowledge, the present study is the first one use the aforementioned technique to assess the water quality with two periods.

 These techniques allow extracting hidden information from the data set to obtain information about the environment's potential impacts on water quality and offer greater possibilities to aid decision-making. Factor analysis (FA) attempts to explain the relationship between observations



regarding fundamental factors, which cannot be directly observed (Singh , et al., 2004). FA can help in identifying the main components of water quality the most important variables that cause variation in water quality and the impact of potential sources on water quality. The output of FA may provide a helpful guide for decision-makers to identify and prevent sources of pollution in the studied area. Therefore, the aim of this paper is to assess the water quality at the confluence point of Tigris and Diyala Rivers in two different periods and identify the main source of pollution in the study area.

## **2. Materials and Methods**

## **2.1. Study area**

 Tigris River is one of the largest rivers in the Middle East, extending for more than 1,900 km, including 1,415 km inside Iraq, and its area is 235,000 km<sup>2</sup>. The Tigris River enters Baghdad from the north in the Tarmiyah area and extends to the south to divide the city into two areas: Al-Karkh (right) and Rusafa (left). The study area is characterized by an arid to semi-arid climate with hot, dry summers and cold winters and an average annual precipitation of about 151.8 mm (Ismail , et al., 2020).

 On the other hand, Diyala River is the fifth tributary and the third largest tributary of the Tigris River, formed by the confluence of the Sirwan and Tangro rivers in Lake Darbandikhan in the Sulaymaniyah Governorate in northern Iraq. The river passes through Iran and Iraq and has a total length of 445 km. The river originates in the Zagros Mountains and joins the Tigris River, south of Baghdad. It passes through three Iraqi governorates, namely Sulaymaniyah, Diyala, and Baghdad (Al-Rubaie & Al-Musawi 2019). The Diyala River can be classified according to the topography of the region into four regions: Upper Darbandikhan Dam, Upper Diyala, Central Diyala, and Lower Diyala (Abed et al. , et al., 2021). The last area is considered the most polluted reach due to the presence of many sewage drains that discharge their wastewater into the river, such as the Nahrawan irrigation drainage, the Rustamiya wastewater treatment plant, the Army Canal, and others (Ismail & Muntasir 2018; Ismail , et al., 2022). The Diyala River suffers from significant pollution, which has caused it to deteriorate its physical, chemical and biological properties. The pollutants released into the river vary according to the source of those pollutants, including industrial, domestic, and animal waste. These pollutants are transferred from the Diyala River to the Tigris River when it empties into it south of Baghdad, which negatively affects the



properties of the Tigris River (Sabeeh et al. 2023). The two rivers meet for a distance until the Tigris River gradually regains its characteristics.

 The population of the city of Baghdad is about nine million people, which has increased the demand for water, and the river in Baghdad suffers from the deterioration of its water quality due to the discharge of municipal and industrial liquid waste into the river without adequate treatment. Thus, the area south of Baghdad is one of the most important areas that cause pollution of the Tigris River's water. Moreover, the study area is charecterized by organic pollution and BOD reach to more than 25 mg/L while DO depleted to less than 1 mg/L (Abed , et al., 2021). The map of the study area is shown in Figure (1).





Figure (1): Map of the study area.

### **2.2. Sampling and water quality analysis**

 Following the standard method for water sampling and analysis of the American Association for Public Health (APHA 1998), thirty-five (30 cm depth) grab water samples were collected from Tigris and Diyala River in two different date, the first date was on 26/04/2021 and the second date was on 31/08/2021. Fourteen water samples were collected from Diyala River while twenty-one samples were collected form Tigris River (before and after the confluence point). It should be noted that some samples were taken on the transverse side of the river from one site to make the



water samples more representative of the actual situation in the river. Then, the water samples were transferred to the Labs of the National Center for Water Resources Management to perform the water quality analysis. Table (1) shows the locations of water samples in the study area. Table (2) shows the analytical method, abbreviation, and units for water quality parameters.



Table (1) : The coordinates of water samples in the study area.





Table (2): The analytical method, abbreviation, units for water quality parameters



### **2.3. Data pre-processing and multivariate analysis**

 The multivariate statistics has been used as a successful tool in many applications and scientific disciplines for many years. As for its use in the field of assessing the quality of river water, it has been widely used in many researches and studies in order to obtain better information about the quality of river water, such as (Wunderlin , et al., 2001; Simeonov , et al., 2004 Singh , et al., 2004; Ismai & Robescu 2019). In this study, factor analysis was used to interpret the results, which has a reasonable interpretation of complex water quality data. Factor analysis (FA) may be



useful in searching for solutions to water pollution challenges and can be an effective tool for water resource management. FA was applied through three stages which are generating the correlation matrix for all water quality parameters, and then extracting the initial set of factors using the principal component analysis (PCA) method, and finaly rotating the extracted factors by Varimax rotation (Ismail & Robescu, 2019). FA was conducted using the statistical program IBM SPSS 24. It should be noted that before starting to apply the factor analysis the normal distribution of water quality parameters was examined using the Shapiro–Wilk (W) test (Chapagain , et al., 2010).

## **3. Results and Discussion**

 The descriptive statistics of water quality data for the first and second periods are shown in the Table (3) and Table (4), respectively. It can be seen from Figure (2), which shows the scree plot for the first date (26/04/2021), three major factors have eigenvalues greater than one and explain 80.810% of the total variance in each water quality datasets. Figure (3) shows the scree plot for the second date (30/08/2021) and it is obvious that only two major factors have eigenvalues greater than one and explain 77.40% of the total variance in each water quality datasets.

<b>Parameter</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>SDV</b>
EC	0.85	2.86	1.49	0.76
pH	7.24	7.55	7.40	0.09
<b>TDS</b>	552.00	1847.00	967.91	492.80
Ca	44.00	152.00	85.69	32.27
Mg	26.40	153.60	69.24	34.76
Na	53.59	359.95	148.75	103.08
K	2.50	14.00	6.13	4.53
$SO_4$	172.80	864.00	430.90	233.35
Cl	74.55	390.50	178.51	112.41
CO <sub>3</sub>	0.00	12.00	6.51	6.07
HCO <sub>3</sub>	109.80	292.80	162.78	47.52
NO <sub>3</sub>	1.00	5.00	2.36	1.07
TH	0.85	270.40	154.93	62.70
Ba	7.24	1.27	0.23	0.34
Fe	552.00	0.04	0.01	$0.01\,$
Cr	44.00	0.07	0.01	0.02
Turb.	26.40	46.60	24.46	10.36

Table (3): Descriptive statistics of the results for the first period on 26/04/2021

Table (4): Descriptive statistics of the results for the second period on 30/08/2021







Figure (2) :Scree plot for the first period on 26/04/2021





Figure (3): Scree plot for the second period on 30/08/2021

 Table 5 shows the output results of factor analysis (FA). The three factors that were extracted in this paper for the first date are responsible for explaining 80.81% of the total variation in water quality in the study area. The first factor (F1) contributed to explaining 60.49% of the total variance; the second factor (F2) contributed 11.24%, while the third factor (F3) contributed to explaining 9.08% (Table (5)) . The correlation coefficient values were considered significant and strong if the value was equal to 0.75 or more, and moderate if it ranged from 0.60 to 0.74. Form Table 5, it can be noted that the important variables for each extracted factor have high positive correlation values, as F1 has strong positive correlation with the variables: EC, TDS, Ca, Mg, Na, K, SO4, Cl, HCO3, TH, Ba and medium positive loading factor with Cr. The presence of variables such as Ca, Mg, Na, K and Cl in this factor indicates the geohydrochemical variables in the study area. On the other hand, the presence of variables such as TDS and EC may indicate the problem of water scarcity in the river due to low flow, which led to increased salinity in the river (Yürekli , et al., 2021). As for SO4, the most important sources are the use of agricultural fertilizers and pesticides and their entry into the water during runoff. Also, the presence of SO4 in water at a high percentage in drinking water can cause diarrhea for humans and livestock and also cause blockages in water networks in high concentrations (Bascaron, 1979). The first factor could be called the human, agricultural and hydrochemical pollution factor, which contributed the largest percentage of the total variance value (80.81%). F2 has moderate positive correlation with the variables CO3, and NO3 and moderate negative correlation with pH (Table (5). Changes in pH and nitrates indicate that the source of water pollution with nitrates in the study area is different from the



sources of sulfate presence (Ismail ,et al., 2014). Previous studies indicated that the nonagricultural source of nitrate in the Tigris and Diyala rivers is due to wastewater discharge and reuse of wastewater for irrigation purposes (Abed , et al., 2021). This factor (F2) can be called the effluent factor. The third and final factor has moderate positive correlation with Fe and water turbidity. Water turbidity increases with the increase in water velocity in rivers, in addition to other factors such as soil erosion, so it usually increases during rainy seasons. The presence of Fe may indicate pollution with heavy metals that comes from the discharge of industrial wastewater into the rivers (Ismail & Robescu ,2019). F3 could be called erosion and heavy metals factor.

 As for the second period, on 8/30/2021, only two factors were extracted, responsible for explaining 77.40% of the total variation in water quality in the study area. The first factor contributed to explaining 61.50% of the total variance, and the second factor contributed 15.90% (Table (6)). It is noted that there is no significant difference in the results of the factor analysis in the second period from the first period, and this confirms that the results obtained from the inferential statistics used in this study.















# **4. Conclusions**

 The application of factor analysis to evaluate water quality in the study area gave important results for water quality management and to obtain better information about water quality in the study area. The results revealed that, in the study area, seventeen water quality parameters can be grouped under three factors for the first date and two factors for the date period. The three factors that were extracted in this study for the first date are responsible for explaining 80.81% of the total variation in water quality in the study area. The first factor contributed the largest percentage of the total variance value (80.81%) and was responsible for human and agricultural pollution and the change in hydrochemical properties. This confirms that the sources of change or deterioration in water quality are mainly human activities, in addition to some agricultural activities in addition to hydrochemical properties. The second factor was responsible for discharging wastewater, liquid waste from on-site sewage, and reusing wastewater for irrigation purposes and was called the effluent factor. The third factor, which was called the erosion and heavy metals factor, contributed to explaining a small percentage of the total variance, and this indicates that the process of erosion and contamination with heavy metals has a lesser impact than the sources mentioned above. The factors that were identified in the second period did not change significantly from the factors in the first period, which indicates that these factors prevail in the river in most seasons of the year. The study recommends periodic monitoring of the water quality parameters studied for more than one season, as well as adopting important variables such as biological oxygen demand (BOD) and dissolved oxygen (DO) to be included in the factor analysis, as the reason for not adopting it in the statistical analysis is that it is not measured in all sampling sites.

## **References**

Abbas, N., Wasimi, S. A., & Al-Ansari, N. ,2016, Impacts of climate change on water resources in Diyala River Basin, Iraq. *J Civ Eng Archit*, 10, 1059-1074.

Abed, B. S., Daham, M. H. & Ismail, A. H. ,2021, Water quality modelling and management of Diyala river and its impact on Tigris River. *Journal of Engineering Science and Technology*, 16.1, 122-135.

Al-Rubaie, F., & Al-Musawi, N. ,2019, The Effect of Diyala River Water Quality on the Quality of Tigris River Water using GIS Mapping. *Journal of Engineering*, 25(10), 71-87.



APHA AWWA WEF, 1998, Standard methods for examination of water and wastewater, 20th Edition, American Public Health Association, Washington.

Bascaron M. ,1979, Establishment of a methodology for the determination of water quality. Bol Inf Medio Ambient 9:30–51. CIMA, MOPU, Madrid.

Boyacioglu H. & Boyacioglu H. ,2007, Surface water quality assessment by environmetric methods, *Environmental Monitoring and Assessment*, 131, 371–376.

Chapagain S. K., Pandey V. P., Shrestha S., Nakamura T. & Kazama F. (2010). Assessment of Deep Groundwater Quality in Kathmandu Valley Using Multivariate Statistical Techniques, *Water, Air, & Soil Pollution*, 210, 277–288

Ghalib, H. B. , 2017, Groundwater chemistry evaluation for drinking and irrigation utilities in east Wasit province, Central Iraq. *Applied Water Science*, 7(7), 3447-3467.

Ismail, A. H., & Abed, G. A. (2013). BOD and DO modeling for Tigris River at Baghdad city portion using QUAL2K model. *Journal of Kerbala University (Scientific)*, 3, 257 -273.

Ismail, A. H., Abed, B. S., & Abdul-Qader, S. ,2014, Application of Multivariate Statistical Techniques in the surface water quality Assessment of Tigris River at Baghdad stretch. *Journal of Babylon University*, 22, 450-462.

Ismail, A. H., & Muntasir, A. H. ,2018, Estimation of river Tigris dispersivities using a steady‑state numerical model. *Appl Water Sci*, 8-108.

Ismail A.H, Robescu D. ,2019, Application of multivariate statistical techniques in water quality assessment of Danube river, Romania. *Environ Eng Manag J* 18(3):719-726.

Ismail, A.H, Shareef, M.A, Hussein, H.H, Salman, H.H, Salim, A.H, Matar, A.G, 2022, Twodimensional simulation of pollutant dispersion at the confluence of Diyala river with Tigris river using CFD technique, *Journal of Water Resources and Geosciences*, 1 (2), pp 85 - 100.

Ismail, A. H., Shareef, M. A., & Alatar, F. M. ,2019, Hydrochemistry of Groundwater and its Suitability for Drinking and Irrigation in Baghdad, Iraq. *Environ Process*, 6, 543-560.

Ismail, A. H., Hassan, G., & Sarhan, A. H. ,2020, Hydrochemistry of shallow groundwater and its assessment for drinking and irrigation purposes in Tarmiah district, Baghdad governorate, Iraq. *Groundw Sustain Dev, 10-100300*, from [https://doi.org/10.1016/j. gsd.2019.100300.](https://doi.org/10.1016/j.%20gsd.2019.100300)

Ismail, A., Shareef, M., Hussein, H., Salman, H, Kadhim, N., Salim, A., & Matar, A. ,2022, Twodimensional simulation of pollutant dispersion at the confluence of Diyala river with Tigris river using CFD technique. *Journal of Water Resources and Geosciences*, 1(2), 85–100.



Sabeeh, K.G, Mustafa, M.T, Ismail, A.H, Boncescu, C, ,2023, Evaluation of Water Quality at the Confluence Region of Diyala and Tigris Rivers Based on the Weighted Arithmetic Method, Journal of Techniques,  $5(1)$ , pp  $58 - 65$ 

Simeonov V., Simeonova P. & Tsitouridou R. ,2004)., Chemometric quality assessment of surface water: Two case studies, *Chemical Engineering Ecology*, 11, 449–469.

Singh K.P., Malik A., Mohan D., Sinha S. ,2004)., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): A case study. *Water Research*, 38, 3980–3992.

Wunderlin D. A., Díaz M. P., Amé M. V., Pesce S. F., Hued A. C. & Bistoni M. A., 2001, Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba-Argentina), *Water Research*, 35, 2881–2894.

Yürekli, K., Erdoğan, M., Ismail, A. H, Shareef, M. A. ,2021, How far can hydrochemical characteristics of surface water meet drinking and irrigation criteria: a pragmatic study for the Euphrates river basin, Turkey, *Sustainable Water Resources Management*, 7:25.