Water Science & Technology

 Water Science & Technology Vol 90 No 4, 1290 doi: 10.2166/wst.2024.276

Optimized wastewater management utilizing multivariate statistical analysis: a case study of the Mascara wastewater treatment plant, Algeria

Imène Benstaali *, Amel Talia and Laouni Benadela

Laboratory of Water Science and Techniques, University of Mascara, BP 763, Mascara 29000, Algeria
*Corresponding author. E-mail: i.benstaali@univ-mascara.dz

IB, 0009-0000-4215-9413

ABSTRACT

Effective wastewater management is crucial in regions experiencing water scarcity and environmental stressors, such as pollution and climate change. Optimizing treatment processes is essential for achieving environmental sustainability. This study aims to highlight the importance of effective wastewater management strategies, particularly in regions facing water scarcity. Our objective was to identify key factors influencing the treatment process. Therefore, we evaluated associations between physicochemical parameters using multivariate statistical methods, including Principal Component Analysis (PCA) and Hierarchical Ascendant Classification (HAC). Our findings categorize the monthly water samples into three distinct groups based on levels of organic pollution: the first group (July, August, and September) is characterized by high oxygenation levels and significantly low organic pollution, indicating optimal system operation. The second group (April, October, November, and December) exhibits low oxygenation and low organic pollution, promoting sludge settling and pollutant reduction. The third group (January, February, March, May, and June) shows significantly high organic pollution and low oxygenation, which corresponds to unfavorable environmental conditions. Our study demonstrates the effectiveness of multivariate statistical methods in optimizing wastewater treatment processes, providing crucial insights for environmental sustainability and water resource management.

Key words: HAC, organic pollution, PCA, physicochemical parameters, wastewater

HIGHLIGHTS

- Use of multivariate statistical analysis for characterizing treated wastewater.
- Identification of specific seasons or periods when organic pollution levels are notably high, facilitating targeted interventions and resource allocation.
- Optimization of wastewater management, enabling improved practices and enhancing overall efficiency and sustainability of treatment processes.

INTRODUCTION

Water is fundamental to the development of nations, as a consistent supply of fresh water is necessary for establishing permanent communities (Hashem & Qi 2021). Securing sufficient access to water resources is crucial for fostering sustainable development (Bhaduri *et al.* 2016).

In many regions worldwide, ongoing population growth and rising water demand are leading to more frequent freshwater shortages (Hussain *et al.* 2019).

Algeria, marked by its arid and semi-arid climate, experiences irregular rainfall and high evapotranspiration rates (Guergueb & Ferhat 2021; Derdour *et al.* 2022), which contribute to significant challenges in managing its water resources. These challenges include adverse climatic conditions, overconsumption exceeding the natural renewal rate of the resource, and diffuse or occasional pollution (Masmoudi *et al.* 2016). With an annual water potential of approximately 23.2 billion m^3 comprising 10.2 billion m^3 of surface water and 13 billion m^3 of groundwater (MRE 2020) and a population of approximately 43 million, Algeria confronts substantial constraints on freshwater resources (Hamiche *et al.* 2015). Water availability in Algeria falls below the recognized benchmark of 1,700 m^3 per person per year. According to the World Health Organization (WHO 1989), availability is less than

1,000 m^3 per person per year. Projections suggest a severe water shortage post 2050, with an anticipated 40% decrease in water availability per person to less than 500 m^3 per year (Terink *et al.* 2013).

According to MRE (2020), Algeria utilized 4.06 billion m^3 of water for drinking purposes and 7.6 billion m^3 for irrigation. As in many other countries, irrigation is the primary use of water resources in Algeria, a trend driven by the long-term decline in water availability observed in recent decades (Drouiche *et al.* 2012).

To address these challenges, Algeria has implemented a new strategy to mobilize and secure various water resources, ensuring sustainability and integrated, rational water resource management at the national level. The reuse of treated wastewater is considered a primary alternative for expanding water resources, particularly in arid regions (Chojnacka *et al.* 2020; Hussain & Qureshi 2020). Wastewater reuse for irrigation provides an additional, reliable, and safe source of water and nutrients (Becerra-Castro *et al.* 2015), aiding in natural resource preservation and promoting integrated water management.

Ensuring the safety of reused wastewater for crop irrigation is a critical issue globally (Ghernaout & Elboughdiri 2020). Given the variability in wastewater characteristics based on its source, thorough examination during treatment and reuse processes is essential (Ilori *et al.* 2019). Adhering to established quality standards for wastewater discharge into natural environments can mitigate the harmful effects of inadequate treatment. Improperly treated wastewater poses significant risks to public health, the environment, and the economy (Libutti *et al.* 2018). Therefore, it is crucial to adopt effective treatment and utilization methods to protect public health (Hashem & Qi 2021).

Despite advancements in the design and operation of urban wastewater infrastructure over the past decade, challenges related to effluent quality persist (Zhang *et al.* 2021). This emphasizes the need for better management and interpretation of data from wastewater treatment facilities. Enhanced data analysis can provide critical insights for optimizing treatment processes and ensuring compliance with safety standards.

Continuous analysis campaigns play a pivotal role in improving the understanding and management of wastewater quality. However, they often generate vast amounts of complex and heterogeneous data that can be challenging to interpret meaningfully. To address this challenge, multivariate statistical techniques such as Principal Component Analysis (PCA) and Hierarchical Ascendant Classification (HAC) are indispensable. These methods not only establish relationships among interconnected data but also quantify wastewater quality and compress large databases of quantitative variables to extract relevant information (Aguado & Rosen 2008). Specifically, these approaches have proven particularly effective in analyzing seasonal variations in wastewater flows generated by domestic and industrial sources (Ouali *et al.* 2009), thereby providing crucial insights for optimizing wastewater treatment processes.

Numerous studies worldwide have used multivariate statistical analysis to establish relationships between quality parameters and pollution sources in wastewater treatment data (Aguado & Rosen 2008; Lefkir *et al.* 2016; Abbaa *et al.* 2021; Rekrak *et al.* 2021; Mouhtady *et al.* 2022; Rahmat *et al.* 2022; El Aatik *et al.* 2023; Moussaoui *et al.* 2023; Peng *et al.* 2023; Newhart *et al.* 2024). These methods are recognized for their ability to effectively handle large, complex, and diverse datasets, providing comprehensive insights into the relationships among variables influencing wastewater quality. Multivariate statistical analysis enables precise modeling and identification of key factors impacting the performance of wastewater treatment facilities. This approach ensures more accurate and effective management of wastewater treatment processes.

This study aims to highlight the importance of effective wastewater management strategies at an activated sludge treatment plant in Mascara, a region facing water scarcity. The main objective is to identify key factors influencing the treatment process and to highlight seasonal variations in organic pollution levels to understand the operational patterns at the plant. To achieve this objective, two types of multivariate statistical methods were applied. The first involves PCA, which reduces the dimensionality of datasets and evaluates associations between physicochemical parameters. The second method is HAC, which groups parameters into homogeneous categories. By identifying critical factors and analyzing seasonal variations, this research offers an innovative approach to optimize wastewater treatment processes, crucial for environmental sustainability. The final aim of this study is to inform policymakers and operators on necessary adaptations to optimize water resource management under diverse environmental conditions.

DATA AND METHODS

Study area

Mascara is a city located in northwestern Algeria, between 0° and 0°15′ East longitude and 35°15′ and 35°30′ North latitude, as shown in Figure 1. The climate of the area is semi-arid, characterized by irregular seasonal precipitation and an extended

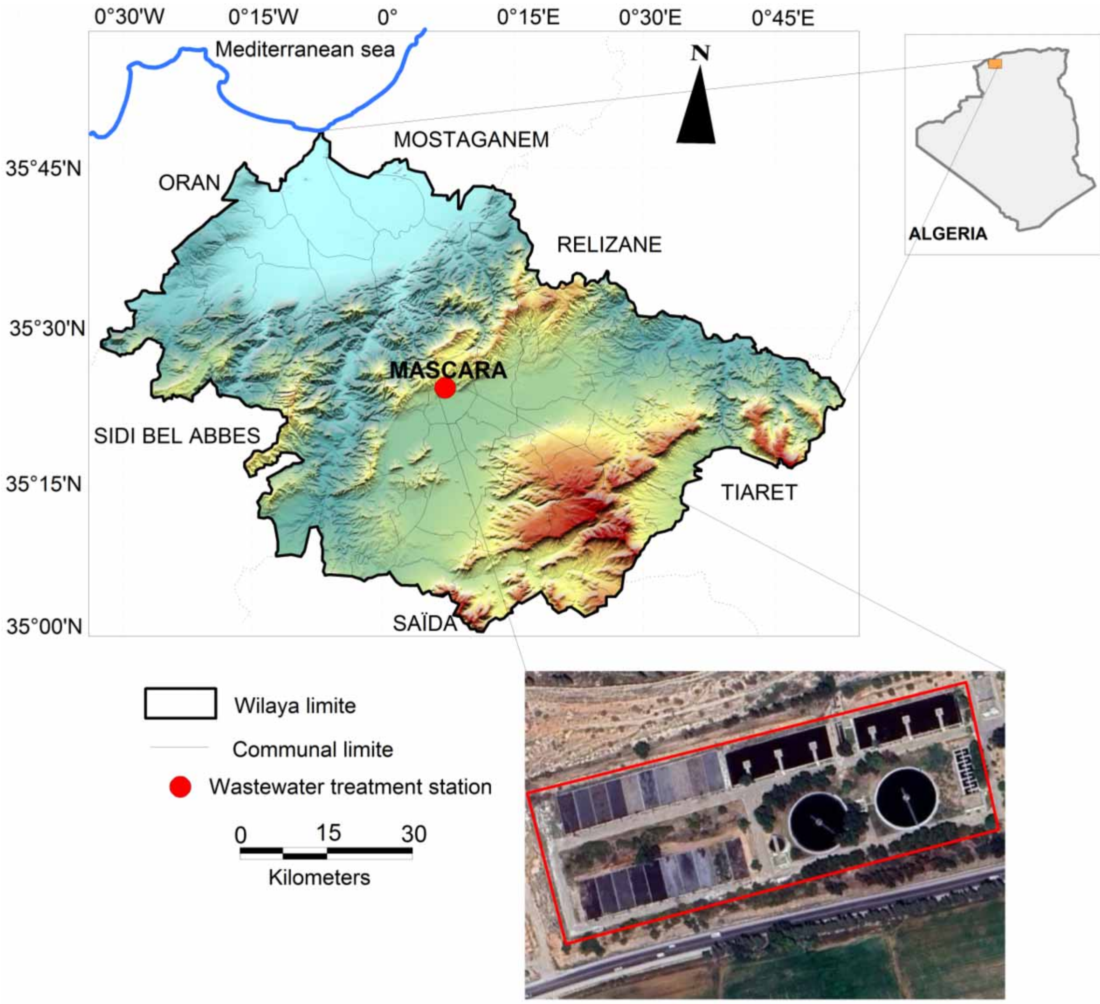

Figure 1 | Location of the Mascara WWTP.

period of summer drought. Annual rainfall typically falls below 367 mm, with an average yearly temperature of around 17.4 °C.

The city of Mascara is situated in an agricultural region covering an area of 75 km^2, with 71% of the territory consisting of agricultural land, totaling 55 km^2. Within this area, 35 km^2 is located in the Ghriss plain, where groundwater resources have been significantly depleted due to increased exploitation and decreased rainfall. However, the treated wastewater from urban areas could be used in agriculture to compensate for the deficit and free up conventional resources for drinking water supply (Chadli *et al.* 2022).

Description of the mascara activated sludge treatment plant

The Mascara wastewater treatment plant (WWTP) is located 2 km south of the provincial capital, near the Kouair River, which receives the treated water. The plant is designed to serve a population equivalent of 100,000 inhabitants with a treatment capacity of 13,000 m^3 per day. It treats wastewater from both urban and industrial sources using a low-loaded activated sludge process (or extended aeration). The treatment plant operates in three phases: pre-treatment, biological treatment using low-loaded activated sludge, and sludge treatment (drying beds). At present, the treated wastewater from the sewage treatment plant is specifically intended for irrigating an extensive agricultural area spanning 1,095 ha in the Ghriss Plain. This innovative and sustainable approach involves the efficient treatment of wastewater, facilitating its reuse to address local agricultural water requirements.

Collecting data

The available database for this study includes physicochemical parameters of raw (influent) and treated (effluent) waters from the plant, with a monthly time step. The database covers the period from January to December 2020. These data are measured and provided by the National Sanitation Office of the Mascara province. Measurements are conducted four times per month, and an average value is assigned for each month. Additionally, the Ghriss meteorological station provides data on air temperature (T_{air}) and wind speed (Wind).

The physicochemical parameters include potential of hydrogen (pH), volume (V), water temperature (T_w), chemical oxygen demand (COD), biochemical oxygen demand over 5 days (BOD_5), suspended solids (SS), dissolved oxygen (DO), ammoniacal nitrogen (NH_4^+), nitrates (NO_3^-), and orthophosphates (PO_4^{-3}).

Statistical analysis methods

The adopted methodological approach relies on the use of multivariate statistical methods: PCA and HAC. These methods were applied using STATISTICA.

Many studies have applied HAC and PCA for the classification and characterization of wastewater quality (Ouali *et al.* 2009; Rekrak *et al.* 2021; Mouhtady *et al.* 2022; Rahmat *et al.* 2022; El Aatik *et al.* 2023; Moussaoui *et al.* 2023). This technique helps establish a relationship between quality parameters and sources of pollution in wastewater treatment data (Rahmat *et al.* 2022). It identifies the variables that contribute the most to effluent quality, providing crucial information for operators and decision-makers to modify treatment processes accordingly (Abbaa *et al.* 2021).

PCA is a multidimensional statistical method used to establish, on one hand, a similarity assessment among samples (samplings), and on the other hand, an assessment of the relationships between variables (Hotelling 1933). The fundamental concept of PCA is to transform a large set of data containing associated variables into a smaller set of uncorrelated variables while maintaining the largest amount of information relating to the variation between the variables in the original dataset (Mouhtady *et al.* 2022). The method involves analyzing the correlation coefficients of each parameter successively with the first, second, and third principal components and representing the results graphically. Principal components generated during the analysis are arranged in such a manner that they correspond to a decreasing contribution of variance; i.e., Principal Component 1 (PC1) explains the highest amount of variance in the original data (Vieira *et al.* 2012).

It is important to note that PCA is sensitive to outliers (Garry 2007). To minimize the influence of these values, PCA is applied to centered and normalized data, giving equal weight to each variable. The initial classification of the parameters involves separating them based on their positive or negative correlations with the principal components. A factor loading greater than 0.75 is considered strong, while a range of 0.5–0.75 is considered moderate, and a range of 0.3–0.5 is considered weak (Schreiber 2020).

The HAC method is primarily used to complement the results obtained by PCA. It organizes a large dataset into clusters based on a given set of characteristics (Sharma *et al.* 2021). Cluster analysis facilitates the grouping of observations or variables according to their similarities or differences (Hussain 2004). The HAC output is represented by a dendrogram that categorizes observations or variables into groups or subgroups based on similarities. In this study, a dendrogram was obtained by performing Ward's method using squared Euclidean distance as a measure of similarity.

The PCA was conducted on the physicochemical data of treated wastewater at the Mascara activated sludge station and the climatic data from the Ghriss Meteorological Station. The analysis involved utilizing various physicochemical parameters of treated wastewater as variables, such as pH, volume (V), water temperature (T_w), BOD_5, COD, SS, DO, ammonium nitrogen (NH_4^+), nitrates (NO_3^-), ortho-phosphate (PO_4^{-3}), and various climatic parameters including air temperature (T_{air}) and wind speed (Wind). Each sampling month within the 12 month period was considered as an individual case in the PCA. HAC was applied to a matrix with 12 rows (months in a year) and six columns (physicochemical parameters). In this classification, we considered only the organic parameters necessary for assessing organic pollution. Figure 2 shows the methodology described above and used in this study.

RESULTS AND DISCUSSION

Physicochemical parameters

The physicochemical parameters of the effluent wastewater were analyzed and compared with standards recommended by Algerian norms (JORA 2012), as well as those set by the World Health Organization (WHO) and the Food and Agriculture Organization (FAO).

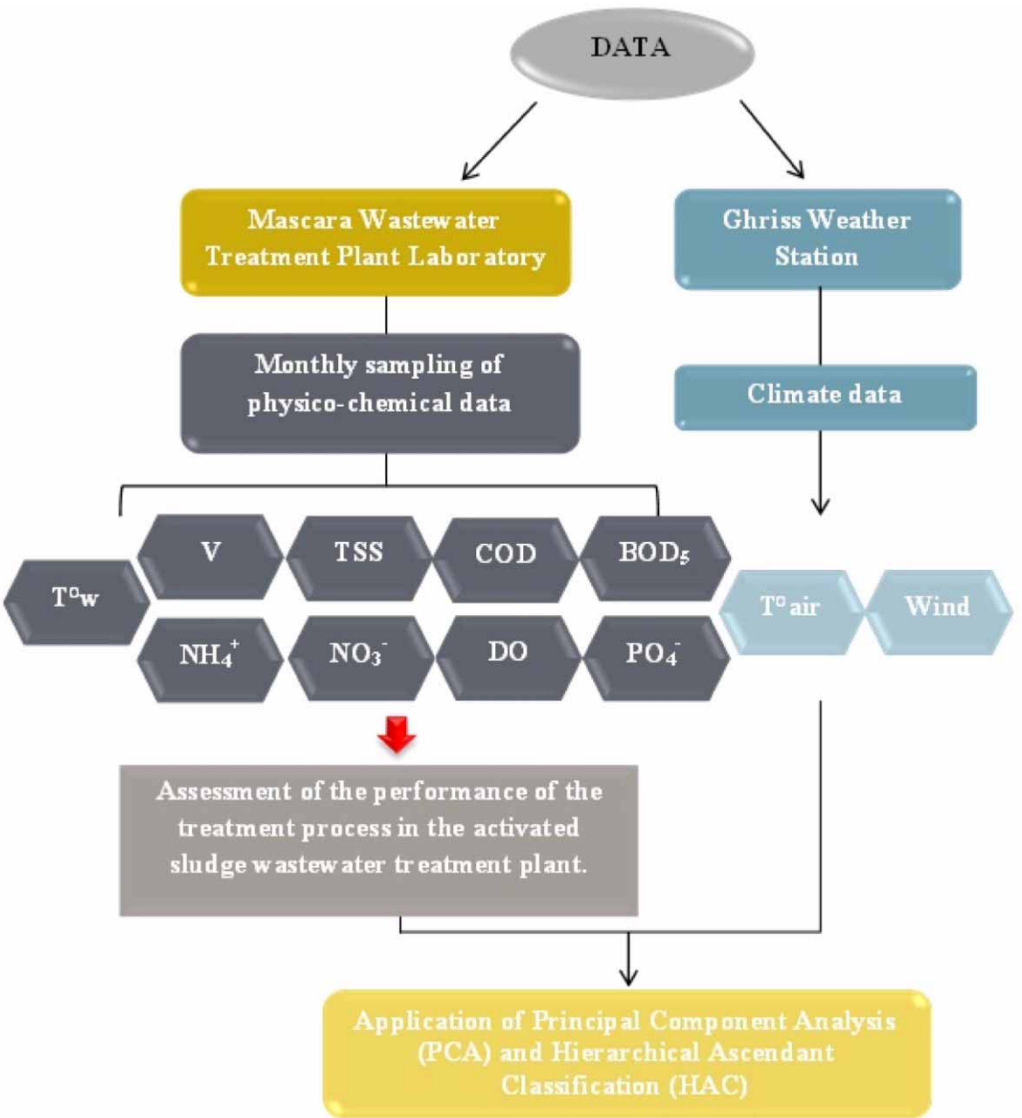

Figure 2 | The approach adopted in this study.

Table 1 provides a summary of basic descriptive statistics, including the range values, for the various physicochemical parameters of samples collected at the outlet of the Mascara WWTP. The analysis of Table 1 and Figure 3 reveals variability in various wastewater quality parameters.

The inter-monthly irregularities of these parameters are assessed by their coefficients of variation. Notably, the coefficients of variation for PO_4^{-3} (14%) and NH_4^+ (23%) are slightly low, indicating that the data have relatively low dispersion around their mean. This suggests that the inter-monthly irregularities of these parameters are less significant. Conversely, the coefficients of variation are slightly high for T_w (33%), TSS (34%), BOD_5 (35%), and DO (33%), indicating slightly greater dispersion of values around their mean. This highlights the variability in the data and the potential for extreme values. Specifically, the coefficient of variation for NO_3^- stands at 118%, indicating a strong monthly irregularity. This suggests considerable fluctuation in NO_3^- values from month to month, with significant dispersion relative to the monthly mean. Furthermore, the low coefficient of variation for pH (3%) implies a higher degree of stability in this parameter. A low coefficient of variation for pH suggests that this aspect of water quality tends to remain close to its average, indicating more consistent and predictable behavior. Conversely, these irregularities indicate significant deviations from the mean, suggesting dynamic changes in these

Table 1 | Descriptive statistics of physicochemical parameters of samples at the outlet of the Mascara WWTP

Parameter	Unit	Mean	Min	Max	Standard deviation	Cv (%)	SD Algerian Standards
T_w	°C	15.22	9.75	23.85	5.07	33.3	25
pH	–	7.60	7.33	8.35	0.26	3.5	6.5–8.5
BOD_5	mg/L	42.56	22.50	78.00	14.91	35	30
COD	mg/L	92.36	57.75	133.50	19.45	21	90
TSS	mg/L	37.90	22.00	61.75	12.95	34.2	30
DO	mg/L	3.20	1.67	5.22	1.05	32.8	–
NH_4^+	mg/L	6.84	4.13	10.49	1.64	24	–
NO_3^-	mg/L	0.95	0.09	4.24	1.12	118	30
PO_4^{-3}	mg/L	2.64	2.13	3.49	0.39	15	1à 2

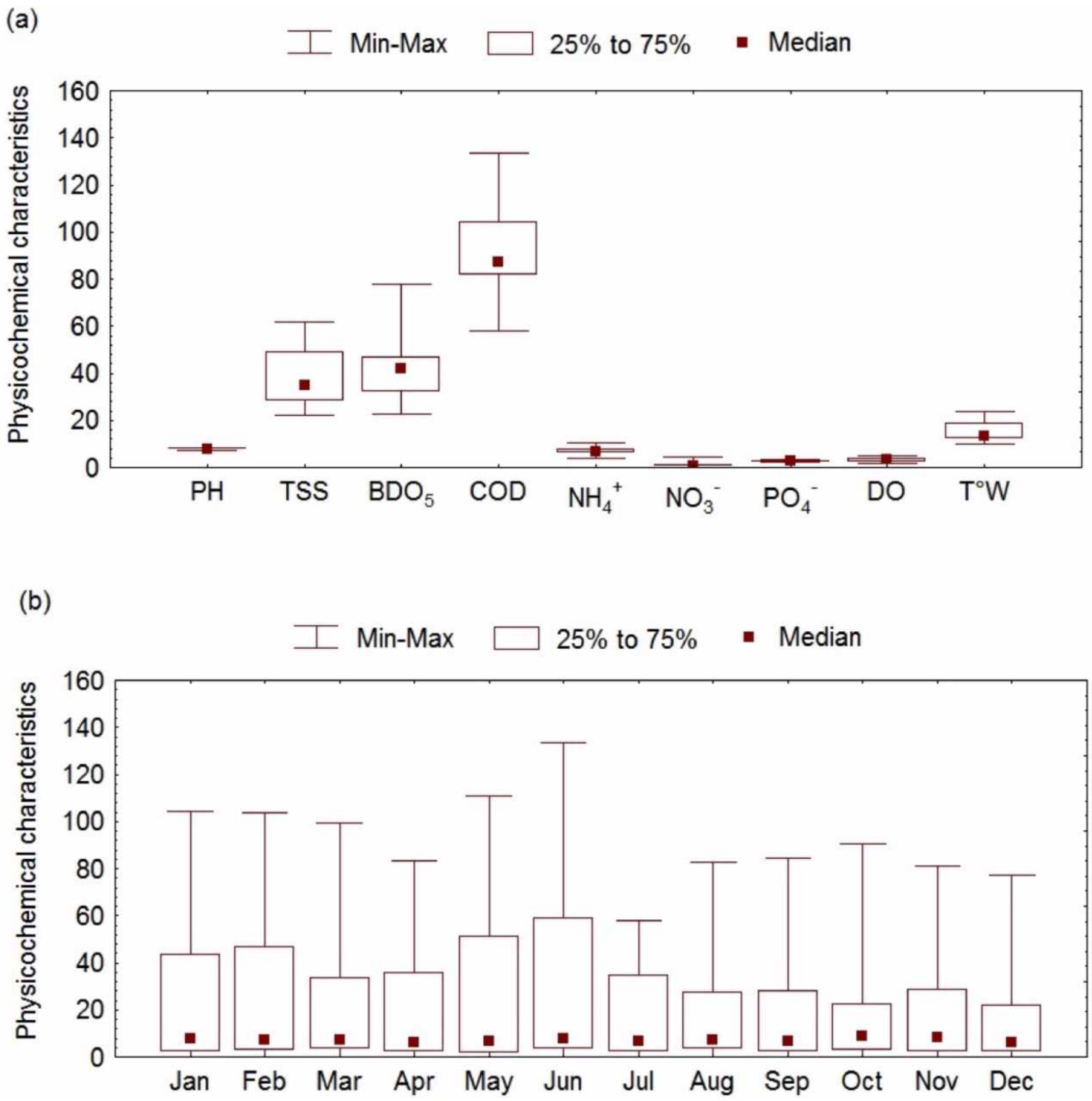

Figure 3 | Boxplot of the physicochemical characteristics of wastewaters.

water quality indicators. Hence, parameters with a low coefficient of variation, such as pH, along with NO_3^- exhibiting a high coefficient of variation, were not considered in the PCA and HCA (Ketrouci *et al.* 2023).

Figure 3 shows a substantial variation in the physicochemical composition of wastewater between different months due to the significant differences between medians and maximum values. The results indicate substantial variability in concentration rates both across different months and among various parameters.

This underscores the importance of considering both temporal and parameter-specific variations when assessing the composition of wastewater. Understanding these variations is crucial for effective wastewater management and treatment strategies, as it provides insights into the specific challenges posed by different months and parameters in maintaining water quality standards.

Water temperature

Temperature is an important ecological factor that positively affects chemical and biochemical reactions, as well as the growth of living organisms, especially microorganisms, in water (Rodier 2009). It also affects other parameters and influences the reactions of organic matter degradation and mineralization (Moussaoui *et al.* 2023). The temperature values range from 9.7 to 23.8 °C, with an average temperature of 15.2 °C (Figure 4(a)). These results indicate that the temperature of the treated wastewater is below the recommended limit of 25 °C for irrigation water, according to Algerian standards (JORA 2012).

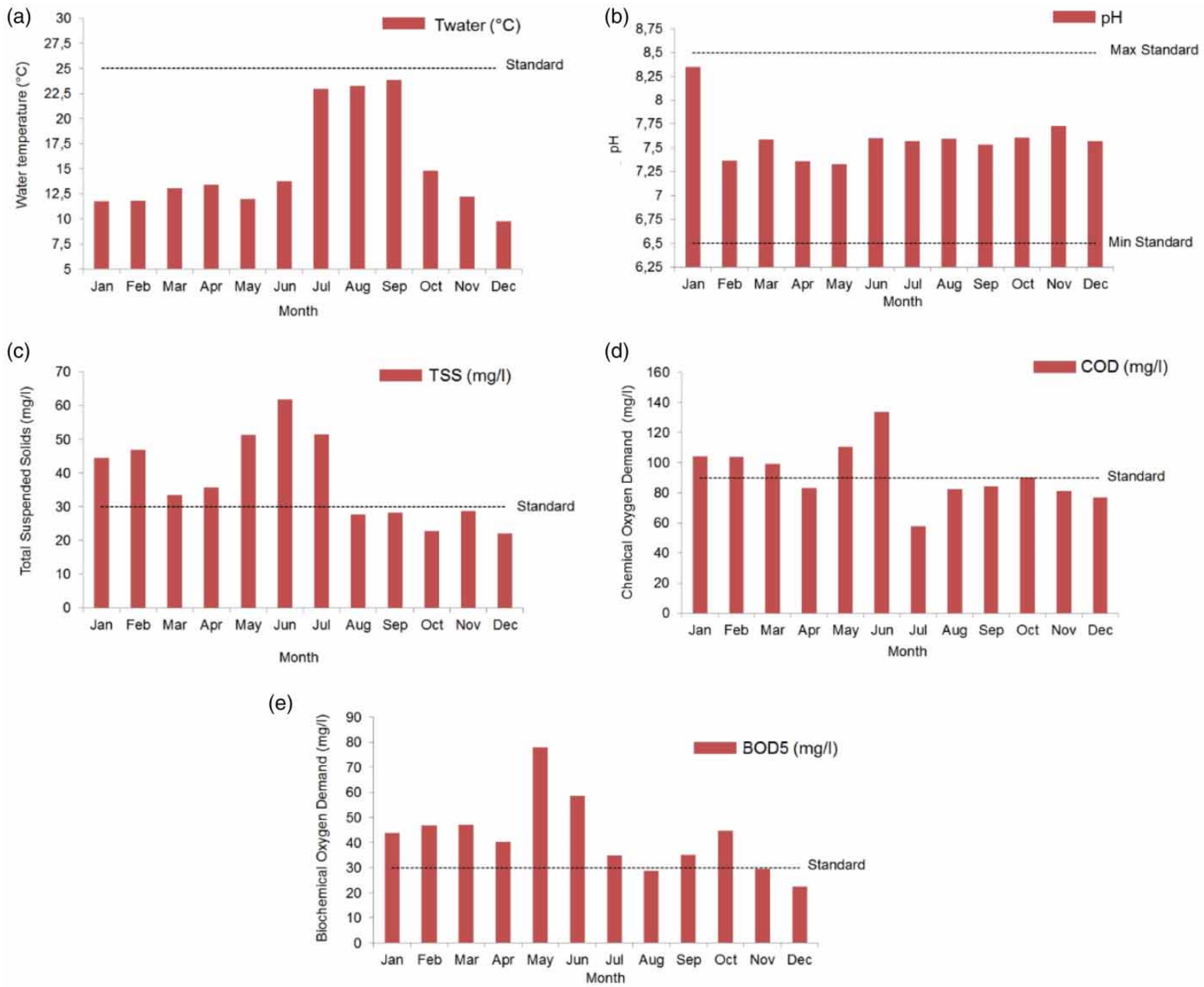

Figure 4 | Monthly variation of effluent parameter measurements in the treated wastewater: (a) T_w, (b) pH, (c) TSS, (d) COD, (e) BOD_5.

Potential of hydrogen

The pH of the raw wastewater at the outlet exhibits slight alkalinity, ranging from 7.33 to 8.35, with an average of 7.60. This alkalinity in the effluent may be attributed to the composition of domestic waste, such as soap water, feces, and urine (Paliwal *et al.* 1998). This range supports the growth of bacteria necessary for the biological breakdown of organic pollutants, as noted in a previous study by Kadouche and colleagues (Kadouche *et al.* 2018). The pH values of the treated wastewater illustrated in Figure 4(b) conform to the recommended range of 6.5–8.5 according to Algerian standards (JORA 2012).

Total suspended solids (TSS)

TSS constitutes a mixture of organic matter, inorganic matter, and microorganisms present in wastewater. TSS can negatively impact water transparency and decrease light penetration, which can ultimately affect photosynthesis. Sedimentation and filtration processes can be used to remove solid particles from wastewater. At the outlet of the basin, the concentrations range from 22 to 66.7 mg/L, with an average of 37.89 mg/L (Figure 4(c)). Although SS in treated waters have decreased, several sampling months recorded values that exceeded the limit allowed by Algerian regulations for irrigation water. This is due to filter clogging caused by high levels of particles in urban effluent. The recorded values for the months of August, September, October, November, and December are below the permissible limit set by Algerian regulations for irrigation water (<30 mg/L) (JORA 2012). This reduction is attributed to the settling of suspended matter and physical filtration, which involves the retention of coarse and fine materials on the surface (Chachuat 1998).

Chemical oxygen demand

COD is a method used to determine the concentration of dissolved or suspended organic or mineral substances in water (Clescerl *et al.* 1998). It measures the amount of oxygen required for their complete chemical oxidation (Rodier 2009). The COD values vary between 57.7 and 133.5 mg/L, with an annual average of 92.36 mg/L (Figure 4(d)). Concentrations in January, February, March, May, and June exceed the standards required by Algerian legislation for irrigation water (90 mg/L) (JORA 2012). The lowest COD values are observed during July, August, September, and December, where concentrations fall below the authorized limit for discharge into the natural environment. The reduction in COD concentration is primarily caused by the physical retention of organic matter in the discharge filters and, to some extent, its oxidation by microbial flora in the presence of oxygen.

Biochemical oxygen demand

BOD is a critical water quality parameter that measures the amount of DO consumed by microorganisms in breaking down organic matter in water (Arcand *et al.* 1989). Therefore, BOD_5 is a crucial parameter used in wastewater treatment to assess the level of organic pollution in water. The concentrations of BOD_5 at the outlet range from 78 to 22. 5 mg/L, averaging 42.6 mg/L, as shown in Figure 4(e). During August, November and December, the recorded values fell below 30 mg/L, attributed to the degradation of biodegradable organic matter by microorganisms (Kadouche *et al.* 2018). However, throughout the rest of the year, concentrations significantly exceed the standards set by Algerian legislation for irrigation water, which is 30 mg/L (JORA 2012).

Ammonia nitrogen (NH_4^+)

Ammonia nitrogen is a reliable indicator of pollution and is produced by the breakdown of animal proteins, domestic effluents (urea), and urban runoff. The concentration of ammonium in the influent ranges from 4.12 to 10.5 mg/L, with an average of 6.84 mg/L (Figure 5(f)). These values slightly exceed the typical range for water intended for irrigation (0–5 mg/L) according to FAO standards (Ayers & Westcot 1985), except for the months of December and April, when the value does not exceed the standards. The increased levels of ammonia can be attributed to various factors, including seasonal variations that influence microbial activity and decomposition. An increase in ammonium nitrogen concentrations at the WWTP was caused by an influx of organic waste or biomass (Moussaoui *et al.* 2023). Ammonia levels can be affected by nitrification processes, influent characteristics, and treatment performance (Roy *et al.* 2017).

Nitrate (NO_3^-)

Nitrates are the end product of nitrogen oxidation and can be present in treated water. High levels of nitrates in surface waters can lead to eutrophication, while also disrupting the disinfection process for drinking water. Algerian regulations have set a limit of 30 mg/L for nitrate concentration in irrigation water to prevent groundwater pollution (JORA 2012). The nitrate

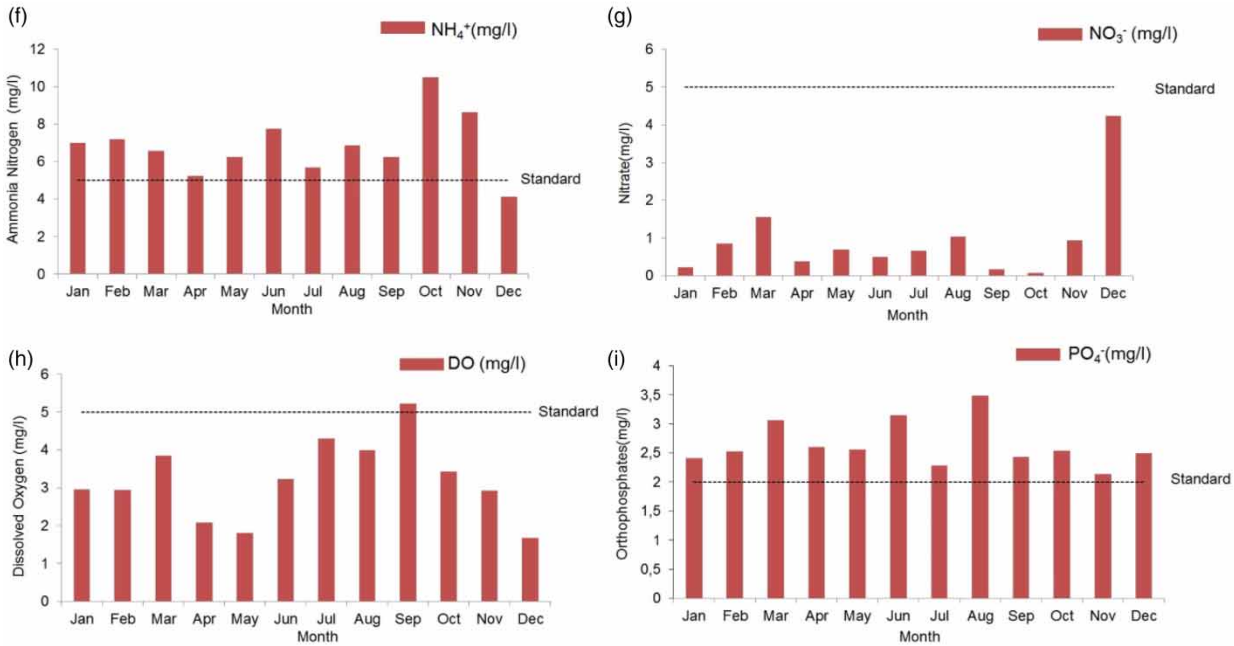

Figure 5 | Monthly variation of effluent parameter measurements in the treated wastewater: (f) NH_4^+, (g) NO_3^-, (h) OD, (i) PO_4^{-3}.

values observed in this study range from 0.08 to 4.24 mg/L (Figure 5(g)). The nitrate concentration in the treated water is below the acceptable limits for irrigation.

Dissolved oxygen

DO in water is the amount of oxygen available for aquatic organisms. It is a crucial parameter for assessing water quality and determining its ability to support aquatic life. The recorded DO values in the effluent at the outlet of the basin are high, ranging between 1.67 and 5.21 mg/L, with an average of 3.20 mg/L (Figure 5(h)), indicating the efficiency of the biological treatment process in the WWTP and the satisfactory degradation of organic matter (Haidara *et al.* 2022). This value falls within the standard limit of 5 mg O_2/L, as stated by WHO standards for direct discharge into the environment (WHO 1989).

Orthophosphates (PO_4^{-3})

Phosphorus is a crucial nutrient for the growth of plants and microorganisms, and it can exist in various oxidized forms. Elevated levels of phosphorus in wastewater can contribute to eutrophication. The results for orthophosphates in the treated wastewater in this study vary between 2.13 and 3.5 mg/L, with an average of 2.64 mg/L of phosphate (Figure 5(i)), which complies with the Algerian standard of 2 mg/L (JORA 2012). The reduction in phosphate levels could be attributed to the consumption of phosphate by bacteria during the treatment process (Rekrak *et al.* 2021).

Correlation matrix

The Pearson correlation matrix analyzes the relationships between different components, quantifying how well the variance of each component is explained by their correlations with one another (Singh *et al.* 2004). A correlation coefficient close to 1 indicates a strong positive relationship, meaning that as one variable increases, the other also increases. Conversely, a coefficient close to -1 signifies a strong negative relationship, where an increase in one variable corresponds to a decrease in the other. Values near 0 indicate little to no linear relationship between the variables, suggesting that changes in one variable do not predict changes in the other.

Table 2 shows the Pearson correlation matrix between the parameters. The analysis reveals that temperature has a positive correlation with DO, and a negative correlation with TSS, BOD_5, and COD. Moreover, the positive correlation between temperature and parameters indicates their increase during the warm period (Bahlaoui *et al.* 1998). Additionally, the positive correlation between DO suggests an increase in its levels during the warm period. Conversely, during cold periods, an increase in TSS, BOD_5, and COD concentrations is suggested by the negative correlation observed with temperature. Furthermore, BOD_5 exhibits a high correlation with COD and a moderate correlation with TSS.

Application of PCA

In order to identify the main factors influencing the treated wastewater, a PCA was applied to the analyzed physicochemical parameters data. The analysis extracted four principal factors following Kaiser's rule based on eigenvalues greater than one (Hurley *et al.* 2012). However, the analysis revealed that the first three components (PC1, PC2 and PC3) effectively explain the dataset's characteristics, accounting for 78% of the total variance. The remaining components do not follow the general trend of actions between variables.

Table 3 displays the correlations between variables and principal factors, component loading factors, cumulative percentages, and percentages of variance for each principal component.

Table 2 | Correlation matrix between different variables

	V	TSS	BOD_5	COD	NH_4^+	PO_4^-	T_w	OD	T_{air}	Wind
V	1.00									
TSS	0.19	1.00								
BOD_5	0.10	0.66	1.00							
COD	0.02	0.54	0.73	1.00						
NH_4^+	0.33	−0.07	0.18	0.29	1.00					
PO_4^-	0.33	0.07	0.13	0.38	0.01	1.00				
T_w	0.71	−0.08	−0.26	−0.23	0.23	0.16	1.00			
OD	0.59	−0.08	−0.30	−0.46	−0.04	0.20	0.82	1.00		
T_{air}	0.67	−0.14	−0.24	−0.49	0.06	0.12	0.77	0.95	1.00	
Wind	0.59	−0.07	−0.02	−0.14	0.57	0.18	0.55	0.43	0.53	1.00

Table 3 | Correlations between variables and principal factors selected by PCA with factor loadings >0.7

	Principal component		
Parameters	**F1**	**F2**	**F3**
V	**0.788**	0.281	0.330
TSS	0.026	**0.821**	−0.175
BOD_5	−0.196	**0.850**	0.136
COD	−0.334	**0.840**	0.228
NH_4^+	0.015	0.083	**0.954**
PO_4^-	0.320	0.455	0.003
OD	**0.869**	−0.087	0.222
T_w	**0.950**	−0.159	−0.091
T_{air}	**0.933**	−0.191	0.040
Wind	0.545	−0.040	**0.691**
Eingen value	4.01	2.53	1.29
% of variance	40.1	25.3	12.9
Cumulative % of variance	40.1	65.4	78.2

The representation of the circular variable in the first factorial plane allows for a quick visualization of correlations between variables (Figure 6). Therefore, PCA enables the definition of two groups of variables: The climatic conditions and physical variables, and the variables directly linked to organic pollution. These two groups of variables are nearly independent of each other, with their vectors being perpendicular. Moreover, the PO_4^{-3} variable is intermediate between the two previous groups of variables. NH_4^+ is not well represented in this first factorial plane, as it correlates with the third component.

The first axis, which alone explains 40.10% of the total inertia, is strongly constructed by physical variables (DO, Tw and V) and climatic conditions (T_{air} and Wind). Their relative contribution to this axis is 87, 95, 79, 93, and 54%, respectively (Table 3 and Figure 6). The first principal component (PC1) appears to indicate effluent associated with warm months and rich in DO.

The second axis accounts for 25.30% of the initial data variance, as depicted in Table 3. This axis is strongly correlated with TSS, COD, and BOD_5, parameters associated with pollution (Table 3 and Figure 6). It represents variations in parameters linked to organic load. The second principal component (PC2) appears to indicate effluent rich in organic matter, suggesting the degree of organic pollution in treated wastewater.

The third axis represents 12.9% of the total variance and is positively correlated with levels of NH_4^+ (Table 3 and Figure 6). The presence of ammonia on this axis indicates a malfunction in the nitrification activity of a wastewater treatment system, which involves the oxidation of ammonia into nitrite. PC2 expresses an effluent rich in ammonia, indicating nitrogen pollution in treated wastewater.

The observations at each month are projected onto the plane formed by the first two axes of the PCA, as shown in Figure 7. The seasonal cycle reveals a clockwise hysteresis, where DO is generally higher during periods of the year when levels of TSS, COD, and BOD_5 are low. This hysteresis curve helps elucidate the variability between the first two components, providing a better understanding of the complex interactions that influence the performance of wastewater treatment systems. Additionally, it can assist in identifying the seasonal factors that most significantly impact the treatment process, aiding in the development of more effective management strategies to maintain optimal performance throughout the year.

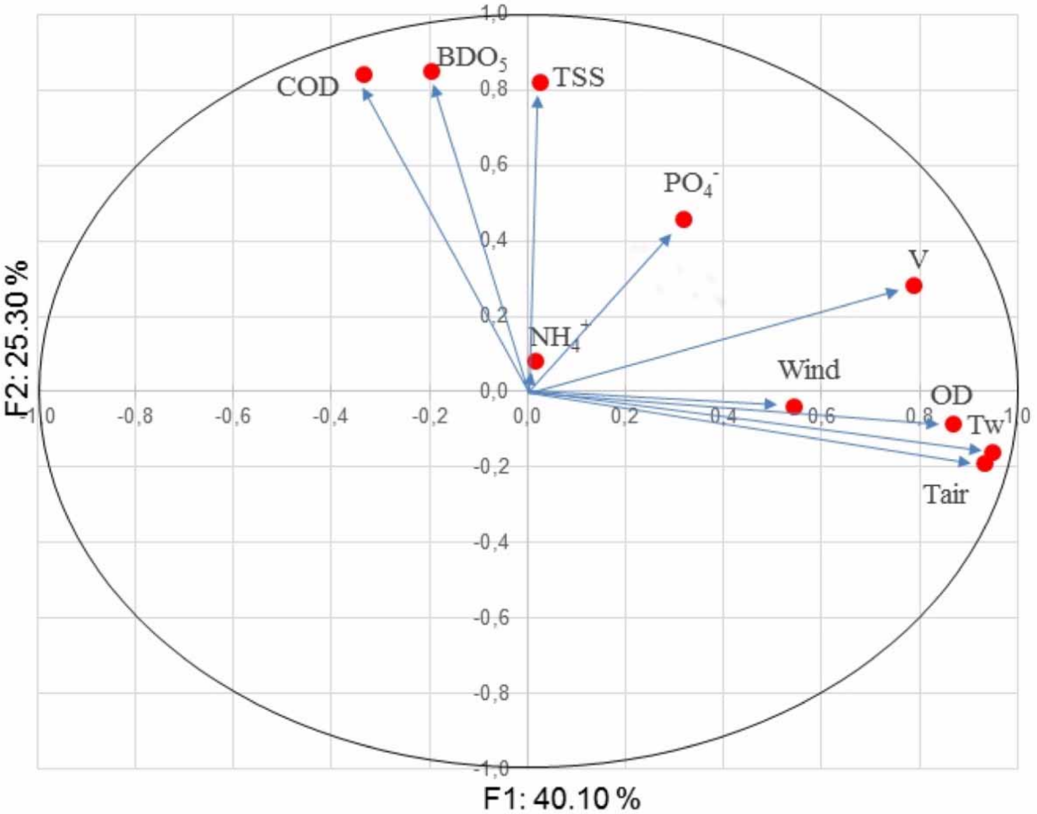

Figure 6 | Presentation of variables on the first two axes of the PCA.

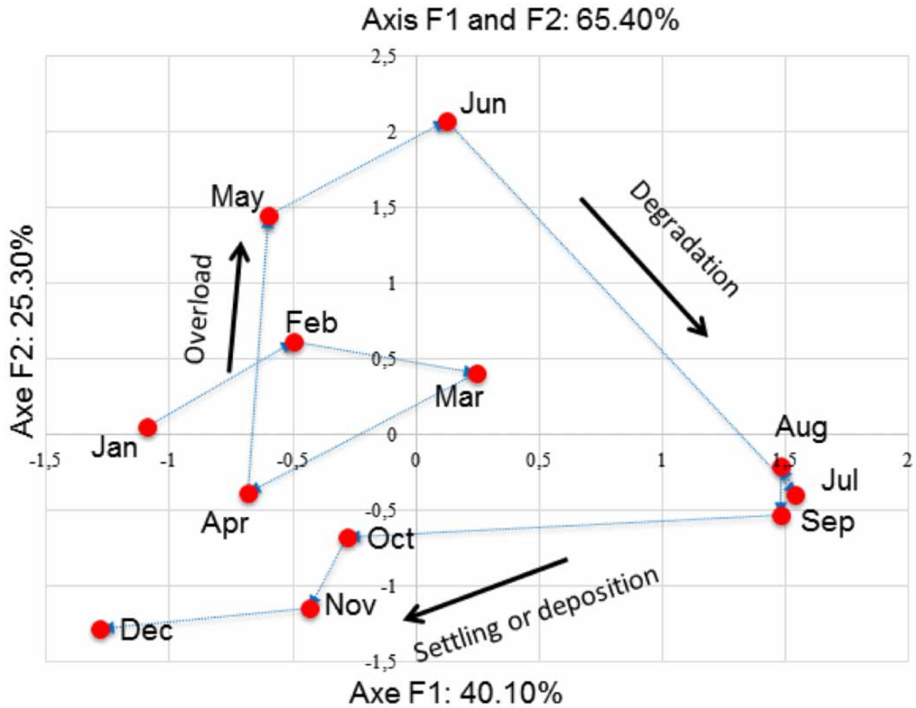

Figure 7 | Presentation of observations (months) on the first two axes of PCA.

At the onset of the cycle (Figure 7), a closed loop pattern is observed, which may be interpreted as follows: In January, February and March, the treated waters exhibit considerably higher levels of pollution compared to those in April. The treatment process appears to be functioning effectively in April, as the situation improves compared to the first three months of the year (effluents are less polluted in April). From April onwards, there is indeed a rapid return to a situation similar to that of January where the wastewater is more polluted.

The cyclical nature of the relationship between the first component (OD, air temperature, water temperature, wind and volume) and the second component (TSS, COD, and BOD_5) is summarized into three periods (Figure 7):

The first period (overload) is distinguished by a substantial increase in concentrations of TSS, BOD_5, and COD, alongside a low DO concentration. This period of the cycle potentially signifies a stage of dysfunction or overload within the treatment system. During this period, activated sludge wastewater treatment systems may encounter challenges in efficiently degrading organic matter present in the wastewater. Various factors contribute to this, including significant organic overload occurring when the amount of organic matter entering the treatment system surpasses its processing capacity. This leads to an accumulation of organic matter within the system, hindering microbial degradation despite the efforts of microorganisms within the treatment system. Unfavorable environmental conditions, such as extremely low temperatures, equipment or process malfunctions, in addition to reduced microbial activity levels due to these conditions, can exacerbate the situation. Furthermore, colder temperatures during these periods can decelerate biological processes in the wastewater treatment system, resulting in decreased efficiency in organic matter degradation. Consequently, organic load in wastewater may accumulate further, while the availability of oxygen for biological processes may diminish due to reduced oxygen solubility in water at lower temperatures. This period may also be associated with decreased activity of microorganisms responsible for decomposing organic matter, leading to an increased accumulation of organic compounds in treated wastewater.

The second period (degradation) is characterized by a decrease in the concentration of organic load and a significant increase in DO. This period of the cycle may correspond to a period where the treatment system operates optimally. It can be termed as a phase of stable performance or efficient operation of the activated sludge wastewater treatment process.

During this period, the operational conditions of the treatment system are optimal, enabling efficient degradation of organic matter by the microorganisms present in the activated sludge. The decrease in TSS, COD, and BOD_5 concentrations indicates an enhanced treatment efficiency, as organic materials are decomposed more thoroughly. The increase in DO concentration in the treated wastewater may result from effective aeration in the treatment system. Adequate aeration supplies sufficient oxygen to the microorganisms, enabling them to degrade organic matter effectively, consequently leading to an increase in DO concentration in the treated water.

The third period (settling) is characterized by a simultaneous decrease in the concentration of organic load and DO. This phase may correspond to a period of sludge settling in the wastewater treatment process. During this period, activated sludge, containing degraded organic matter and other suspended particles, settles to the bottom of the clarifier basin. This results in a reduction in the concentration of TSS, COD, and BOD_5 in the clarified water above the sludge. The decrease in DO concentration can also be observed during this period, as some of the DO is utilized by the microorganisms present in the activated sludge to degrade organic matter. Additionally, settling sludge may consume oxygen during the aerobic decomposition process of organic matter.

Hierarchical Ascendant Classification

The HAC method was used to classify parameters and months into more homogeneous groups in terms of organic pollution. This provides additional visualization to complement the information obtained from PCA. The Ward method was applied using the criteria of a significant Euclidean distance measure of standardized data. The results of this classification reveal three major statistical classes (Figure 8).

The first class, encompassing the months of July, August and September, is characterized by low concentrations of TSS, DOC and BOD_5, as well as high levels of DO, indicating robust oxygenation activity. These observations suggest satisfactory performance of the wastewater treatment process during these months, with operators effectively maintaining low levels of SS and dissolved organic matter while promoting adequate oxygenation of the water. This interpretation also suggests that environmental conditions during this period could be conducive to better water quality and optimal wastewater treatment operation.

The second class, comprising April, October, November and December, exhibits significantly lower concentrations of TSS, COD and $DOB_{5,}$ along with a reduced level of DO compared to the first class. This suggests a better performance of the treatment process during these months, with operators effectively reducing the levels of SS, COD, and BOD_5. The lower levels of DO may indicate efficient oxygen utilization by microorganisms during the treatment process.

The third statistical class, consisting of the months of January, February, March, May and June, is characterized by higher concentrations of BOD_5, COD, and TSS, indicating a higher level of organic pollution. The elevated concentrations of these parameters suggest that the wastewater from these months likely contains a significant amount of degradable organic matter. This may necessitate specific attention and treatment procedures to effectively reduce this organic load and comply with environmental standards for wastewater quality.

Optimizing seasonal management of activated sludge wastewater treatment

To optimize the wastewater management plan for the station, several key strategies should be implemented. First, ensure regular monitoring of system performance by tracking essential parameters such as turbidity, DO concentration and organic matter levels.

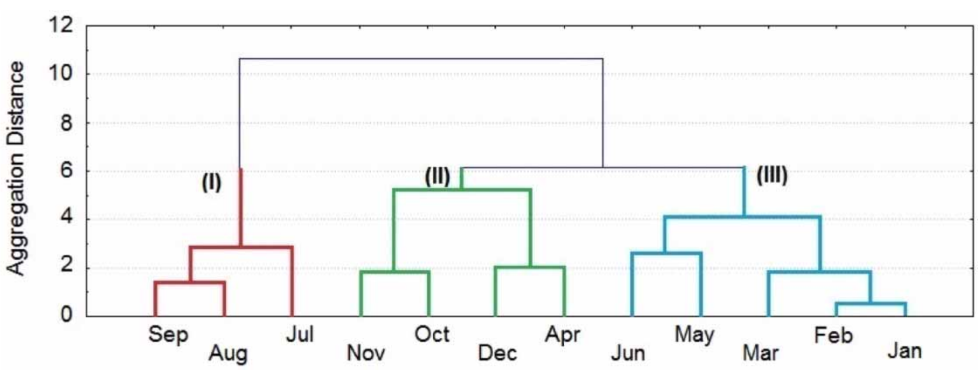

Figure 8 | Hierarchical classification dendrogram of observations (months).

Second, employ statistical methods to identify periods of varying pollution intensity, focusing on concentrations of organic matter indicators like COD, BOD_5, and TSS. This allows for operational adjustments based on seasonal fluctuations.

Finally, manage seasonal dynamics by recognizing and adapting to variations in the performance of the activated sludge treatment system. This includes accounting for changes in water temperature, microbiological biomass, and oxygen demand, thereby optimizing performance throughout the year.

Period-specific management strategies

The optimized wastewater management plan for the station should incorporate strategies tailored to three distinct operational periods.

1. The environmental stress period: during this period, closely monitor system performance and adjust operations to minimize the effects of organic overload. This may include optimizing sludge return ratios, increasing aeration, or adding coagulants to aid in the flocculation of suspended matter.
2. The optimal operation period: take advantage of this period to maximize treatment efficiency, ensuring ideal conditions for microbial growth and organic matter degradation. Optimize sludge control strategies to maintain high efficiency and consistent purified water quality.
3. The sludge settling period: during this period, focus on the effective management of excess sludge produced by the activated sludge treatment process. Optimize settling operations to maximize recovery of activated sludge and minimize losses of microbiological biomass.

CONCLUSION

The aim of this study was to develop a practical method for optimizing wastewater treatment and management processes. The data were efficiently processed using multivariate statistical approaches such as PCA and HAC. Our results reveal similarities among certain months throughout the year in terms of levels of organic pollution. Furthermore, the combined PCA and HAC analysis identified three distinct periods within the wastewater treatment process, each characterized by specific operational conditions based on the level of organic pollution.

The first period (July, August and September) is characterized by decreased organic load and increased DO which reflects optimal system performance and efficient operation. This period represents the optimal phase of system operation. The second period (April, October, November, and December) exhibits simultaneous reductions in organic load and DO, corresponding to the sludge settling phase, where treated water is clarified. The third period (January, February, March, May, and June) shows high levels of TSS, BOD_5, and COD, along with low DO, indicating a potential overload or dysfunction in the system and representing an environmental stress period. Understanding these periods is crucial for implementing effective management strategies to optimize treatment performance and address environmental and operational challenges.

Overall, the study successfully identifies key cyclical periods within the wastewater treatment process, highlighting their operational characteristics and implications for treatment efficiency and management.

ACKNOWLEDGEMENTS

The authors express their gratitude to the National Sanitation Office for making the data available. The authors thank Dr Ismahene Mesbah from Aix-Marseille University, France, for her assistance in reviewing the grammar and syntax of this manuscript.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abbaa, S., Elkiranb, G. & Nouranic, V. 2021 Improving novel extreme learning machine using PCA algorithms for multi-parametric modeling of the municipal wastewater treatment plant. *Desalination and Water Treatment* **215**, 414–426. doi:10.5004/dwt.2021.26903.

Aguado Garcia, D. & Rosen, C. 2008 Multivariate statistical monitoring of continuous wastewater treatment plants. *Engineering Applications of Artificial Intelligence* **21** (7), 1080–1091. doi:10.1016/j.engappai.2007.08.004.

Arcand, Y., Paris, J., Rouleau, D. & Mayer, R. 1989 Effects of operating parameters on biological siudges decantation of an oil refinery waste water plant. *Journal of Water Science* **2** (1), 43–69. https://doi.org/10.7202/705023ar.

Ayers, R. S. & Westcot, D. W. 1985 *Water Quality for Agriculture, FAO Irrigation and Drainage, Paper 29*. Food and Agriculture Organization, Rome, Italy.

Bahlaoui, M.-A., Frouji, M.-A., Drakides, C. H. & Bondon, D. 1998 The impact of the aging of a wastewater treatment facility using lagoon systems on the receiving environment. *Actes Inst. Agron. Veto., Maroc* **18** (4), 227–238. https://www.agrimaroc.org/index.php/Actes_IAVH2/article/download/63/43.

Becerra-Castro, C., Lopes, A. R., Vaz-Moreira, I., Silva, E. F., Manaia, C. M. & Nunes, O. C. 2015 Wastewater reuse in irrigation: A microbiological perspective on implications in soil fertility and human and environmental health. *Environment International* **75**, 117–135. https://doi.org/10.1016/j.envint.2014.11.001.

Bhaduri, A., Bogardi, J., Siddiqi, A., Voigt, H., Vörösmarty, C., Pahl-Wostl, C., Bunn, S. E., Shrivastava, P., Lawford, R. & Foster, S. 2016 Achieving sustainable development goals from a water perspective. *Frontiers in Environmental Science* **4** (64). https://doi.org/10.3389/fenvs.2016.00064.

Chachuat, B. 1998 *Traitement D'effluents Concentrés par Cultures Fixées sur Gravier. Rapport de DEA*. ENGEES Cemagref, France.

Chadli, A., Benhasseini, H., Fertout Mouri, N. & Aguiar, H. 2021 Wastewater reuse, a sustainable alternative for the management of water resources in the city of Sidi Bel Abbes (Algeria). *Geo-Eco-Trop* **46** (2), 301–313. http://www.geoecotrop.be/index.php?page=numero-46.

Chojnacka, K., WItek-Krowiak, A., Moustakas, K., Skrzypczak, D., Mikula, K. & Loizidou, M. 2020 A transition from conventional irrigation to fertigation with reclaimed wastewater: Prospects and challenges. *Renewable and Sustainable Energy Reviews* **130**, 109959. https://doi.org/10.1016/j.rser.2020.109959.

Clescerl, L. S., Greenberg, A. E. & Eaton, A. D. 1998 *Standard Methods for Examination of Water & Wastewater*, 20th ed. American Public Health Association, Washington, DC, USA.

Derdour, A., Bouanani, A., Kaid, N., Mukdasai, K., Algelany, A., Ahmad, H., Menni, Y. & Ameur, H. 2022 Groundwater potentiality assessment of Ain Sefra region in UpperWadi Namous basin, Algeria using integrated geospatial approaches. *Sustainability* **14** (8), 4450. https://doi.org/10.3390/su14084450.

Drouiche, N., Ghaffour, N., Naceur, M. W., Lounici, H. & Drouiche, M. 2012 Towards sustainable water management in Algeria. *Desalination and Water Treatment* **50** (1–3), 272–284. doi:10.1080/19443994.2012.719477.

El Aatik, A., Navarro, J. M., Martínez, R. & & Vela, N. 2023 Estimation of global water quality in four municipal wastewater treatment plants over time based on statistical methods. *Water* **15**, 1520. https://doi.org/10.3390/w15081520.

Garry, B. 2007 *Étude des Processus D'écoulements de la Zone non Saturée Pour la Modélisation des Aquifères Karstiques: Expérimentation Hydrodynamique et Hydrochimique sur les Sites du Laboratoire Souterrain à Bas Bruit (LSBB) de Rustrel et de Fontaine de Vaucluse. Thèse de Doctorat*, Université d'Avignon et des pays de Vaucluse, France.

Ghernaout, D. & Elboughdiri, N. 2020 On the treatment trains for municipal wastewater reuse for irrigation. *Open Access Library Journal* **7** (2), 1–15. https://doi.org/10.4236/oalib.1106088.

Guergueb, M. & Ferhat, A. 2021 Water management in Algeria: Towards a new paradigm. *Journal of Advanced Economic Research, El Oued University, Eloued, Algeria* **06** (01). https://www.asjp.cerist.dz/en/downArticle/494/6/1/153655.

Haidara, R., Abdelbaki, C. & Badr, N. 2022 Feasibility of Water Reuse for Agriculture – Case Study of Ain Temouchent (Algeria). In: *Sustainable Energy-Water-Environment Nexus in Deserts, Proceeding of the First International Conference on Sustainable Energy-Water-Environment Nexus in Desert Climates*, pp. 279–285. doi:10.1007/978-3-030-76081-6_34.

Hamiche, A. M., Stambouli, A. B. & Flazi, S. 2015 A review on the water and energy sectors in Algeria: current forecasts, scenario and sustainability issues. *Renewable. and Sustainable Energy Reviews* **41**, 261–276. https://doi.org/10.1016/j.rser.2014.08.024.

Hashem, M. S. & Qi, X. B. 2021 Treated wastewater irrigation, *A review*. *Water* **13** (1527). https://doi.org/10.3390/w13111527.

Hotelling, H. 1933 Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24** (6), 417–441. http://dx.doi.org/10.1037/h0071325.

Hurley, T., Sadiq, R. & Mazumder, A. 2012 Adaptation and evaluation of the Canadian council of ministers of the environment water quality index (CCME WQI) for use as an effective tool to characterize drinking source water quality. *Water Research* **46** (11), 3544–3552. https://doi.org/10.1016/j.watres.2012.03.061.

Hussain, M. T. 2004 Hydrochemical evaluation of groundwater in the Blue Nile Basin, eastern Sudan, using conventional and multivariate techniques. *Hydrogeology Journal* **12** (2), 144–158. https://doi.org/10.1007/s10040-003-0265-5.

Hussain, M. I. & Qureshi, A. S. 2020 Health risks of heavy metal exposure and microbial contamination through consumption of vegetables irrigated with treated wastewater at Dubai, UAE. *Environment Science Pollution Research* **27** (10), 11213–11226. https://doi.org/10.1007/s11356-019-07522-8.

Hussain, M. I., Muscolo, A., Farooq, M. & Ahmad, W. 2019 Sustainable use and management of non-conventional water resources for rehabilitation of marginal lands in arid and semiarid environments. *Agricultural Water Management* **221**, 462–476. https://doi.org/10.1016/j.agwat.2019.04.014.

Ilori, B. A., Adewumi, J. R., Lasisi, K. H. & Ajibade, F. O. 2019 Qualitative assessment of some available water resources in Efon-Alaaye, Ekiti State Nigeria. *Journal of Applied Sciences and Environmental Management* **23** (1), 35–40. https://dx.doi.org/10.4314/jasem.v23i1.5.

JORA 2012 *Interministerial Decree of 2 January 2012 Setting the Specifications for Treated Wastewater Used for Irrigation Purposes*. Available from: https://www.joradp.dz/FTP/jo-francais/2012/F2012041.pdf (accessed 17 January 2023).

Kadouche, S., Hammoum, H., Ghedamsi, H. & Si Tahar, L. 2018 Assessment of purifying performance of a wastewater filtration basin – case study. *Journal of Water Science* **31** (4), 387–398. https://doi.org/10.7202/1055596ar.

Ketrouci, K., Meddi, M., Bouregba, N., Zatouta, I. & Rezki, A. 2023 Classification of bottled waters marketed and consumed in Algeria through statistical approaches. *Journal of Water and Health* **21** (10), 1489. doi:10.2166/wh.2023.142.

Lefkir, A., Maachou, R., Bermad, A. & Khouider, A. 2016 Factorization of physico chemical parameters of activated sludge process using the principal component analysis. *Desalination Water Treatement* **57** (43), 20292–20297. https://doi.org/10.1080/19443994.2015.1108425.

Libutti, A., Gatta, G., Gagliardi, A., Vergine, P., Pollice, A., Beneduce, L., Disciglio, G. & Tarantino, E. 2018 Agro-industrial wastewater reuse for irrigation of a vegetable crop succession under Mediterranean conditions. *Agricultural Water Management* **196**, 1–14. https://doi.org/10.1016/j.agwat.2017.10.01.

Masmoudi, R., Kettab, A. & Bremond, B. 2016 Drinking water consumption and loss in Algeria the case of networks with low level counting. *Journal of Urban and Environmental Engineering* **10** (2), 162–168. doi:10.4090/juee.2016.v10n2.

Mouhtady, O., Obeid, E., Abu-Samha, M., Younes, K. & Murshid, N. 2022 Evaluation of the adsorption efficiency of graphene oxide hydrogels in wastewater Dye removal: application of principal component analysis. *Gels* **8** (7), 447. https://doi.org/10.3390/gels8070447.

Moussaoui, T., Derdour, A., Hosni, A., Ballesta-de los Santos, M., Legua, P. & Pardo-Picazo, M. Á. 2023 Assessing the quality of treated wastewater for irrigation: A case study of Ain SefraWastewater treatment plant. *Sustainability* **15**, 11133. https://doi.org/10.3390/su151411133.

MRE 2020 *Algeria's Water Resources. Ministry of Water Resources, Synthesis Report (Algeria), Les Ressources en eau D'Algérie*. Ministère des Ressources en Eau, Rapport de synthèse (Algérie), 73.

Newhart, K. B., Klanderman, M. C., Hering, A. S. & Cath, T. Y. 2024 Holistic evaluation of multivariate statistical process monitoring in a biological and membrane treatment system. *ACS EST Water* **4** (3), 913–924. doi:10.1021/acsestwater.3c00058.

Ouali, A., Azri, C., Medhioub, K. & Ghrabi, A. 2009 Descriptive and multivariable analysis of the physico-chemical and biological parameters of Sfax wastewater treatment plant. *Desalination* **246** (1), 496–505. doi:10.1016/j.desal.2008.04.058.

Paliwal, K., Karunaichamy, K. S. T. K. & Ananthavalli, M. 1998 Effect of sewage water irrigation on growth performance, biomass and nutrient accumulation in Hardwickia binata under nursery conditions. *Bioresource Technology* **66** (2), 105–111. https://doi.org/10.1016/S0960-8524(98)00044-3.

Peng, C., Ying, X. & FanChao, M. 2023 Efficient fault monitoring in wastewater treatment processes with time stacked broad learning network. *Expert Systems with Applications* **233**, 120958. https://doi.org/10.1016/j.eswa.2023.120958.

Rahmat, S., Altowayti, W. A. H., Othman, N., Asharuddin, S. M., Saeed, F., Basurra, S., Eisa, T. A. E. & Shahir, S. 2022 Wastewater treatment plant performance using multivariate statistical analysis: A case study of a regional Sewage treatment plant in Melaka, Malaysia. *Water* **14** (20), 3297. https://doi.org/10.3390/w14203297.

Rekrak, Z. A., Chiboub Fellah, A., Abdelbaki, C., Guellil, F. Z. & Bechlaghem, A. 2021 Analysis and reliability of a wastewater treatment plant: The contribution of operational safety at Tlemcen WWTP, Algeria. *Euro-Mediterranean Journal for Environmental Integration* **6**, 1–11. doi:10.1007/s41207-021-00242-x.

Rodier, J. 2009 *L'analyse de L'eau : Eaux Naturelles, Eaux Résiduaires, Eaux de mer*. Dunod, Paris, France.

Roy, D., McEvoy, J., Blonigen, M., Amundson, M. & Khan, E. 2017 Seasonal variation and ex-situ nitrification activity of ammonia oxidizing archaea in biofilm based wastewater treatment processes. *Bioresource Technology* **244** (1), 850–859. https://doi.org/10.1016/j.biortech.2017.08.060.

Schreiber, J. B. 2020 Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social & Administrative Pharmacy* **17** (5), 1004–1011. doi:10.1016/j.sapharm.2020.07.027.

Sharma, G., Lata, R., Thakur, N., Bajala, V., Chandra Kuniyal, J. & Kumar, K. 2021 Application of multivariate statistical analysis and water quality index for quality characterization of Parbati river, Northwestern Himalaya. *Discover Wate* **1** (5). https://doi.org/10.1007/s43832-021-00005-3.

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** (18), 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011.

Terink, W., Immerzeel, W. W. & Droogers, P. 2013 Climate change projections of precipitation and reference evapotranspiration for the Middle East and Northern Africa until 2050. *International Journal of Climatology* **33** (14), 3055–3072. https://doi.org/10.1002/joc.3650.

Vieira, J., Pires, J. C., Martins, F. G., Vilar, V. J., Boaventura, R. A. & Botelho, C. M. 2012 Surface water quality assessment of Lis river using multivariate statistical methods. *Water Air Soil Pollution* **223** (9), 5549–5561. doI:10.1007/s11270-012-1267-5.

WHO 1989 *Health Guidelines for the Use of Wastewater in Agriculture and Aquaculture, Report of A WHO Scientific Group [Meeting Held in Geneva From 18 to 23 November 1987]*. World Health Organization, Geneva, Switzerland.

Zhang, J., Shao, Y., Wang, H., Liu, G., Qi, L., Xu, X. & Liu, S. 2021 Current operation state of wastewater treatment plants in urban China. *Environmental Research* **195**, 110843. https://doi.org/10.1016/j.envres.2021.110843.

First received 11 May 2024; accepted in revised form 31 July 2024. Available online 12 August 2024