Evaluation of physicochemical and heavy metals characteristics in surface water under anthropogenic activities using multivariate statistical methods, Garra River, Ganges Basin, India

Mohd Yawar Ali Khan¹,²,³*, Jie Wen²,⁴

¹Department of Hydrogeology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
²Department of Hydraulic Engineering, Tsinghua University, Beijing- 100084, China
³Department of Earth Sciences, Indian Institute of Technology Roorkee, Roorkee – 247667, India
⁴China Institute of Water Resources & Hydropower Research (IWHR), Beijing 100044, China

ABSTRACT
Water samples from twenty-six locations were collected during March 2014, to evaluate the spatial variation in the water quality of Garra River, Ganges Basin. The physicochemical and heavy metals concentration along the river water were significantly different between the natural and urbanized areas. In the case of the mainstream, the chemical oxygen demand (COD) and the biological oxygen demand (BOD₅) which indicate the levels of organic pollution of the river water, were in the range of 8.1 - 51.4 mg/L and 5.8 - 26.9 mg/L, respectively. On the other hand, in the case of tributaries, these values were in the range of 19 – 36 mg/L and 11.2 – 20 mg/L, respectively. Multivariate statistical analysis was also carried out to discriminate sources of spatial variation of water quality. The results of principal component analysis (PCA) show the relationship between physicochemical parameters suggesting that the quality of water is mostly controlled by carbonate weathering. Dolomite is most likely to be present, which is a common source of Ca⁺ and Mg⁺. Spatial distribution of PO₄³⁻ with a high concentration in the downstream area of river (average 0.35 mg/L) suggests runoff from agricultural fields in this region as the main source.

Keywords: Garra River, Ganges Basin, Heavy metals, Multivariate statistical analysis, Water quality

1. Introduction
Surface water quality is one of the most critical environmental concerns in many parts of the world. Manahan [1] stated that surface water consists of all the water that flows through landmasses like creeks, springs, lakes, rivers and oceans. A stream is a system involving mainstream rivers and tributaries that transports a one-way flow of large amounts of particulate and dissolved load from both anthropogenic and natural sources. At any point, the chemistry of surface water of a river reflects several noteworthy influences, including atmospheric inputs, the lithology of the basin, anthropogenic inputs and climatic conditions [2]. Natural (erosion) and anthropogenic (agricultural discharge) activities affect the quality of regional surface water [3-10]. According to Finnevveden et al. [11], crop waste, degradable solid waste and food waste contribute to the organic content of surface water. Nutrients such as nitrogen (N) and phosphorus (P) cause the growth of algae and phytoplankton, leading to eutrophication. River water quality research helps to develop water management and pollution prevention strategies. Examining the water quality of the river by considering its organic and physicochemical properties helps to reduce the implication in the clarification of water quality due to spatial variations [12].

Recent research has increasingly focused on planning new monitoring strategies [13, 14], in which multivariate statistical techniques such as PCA and hierarchical cluster analysis (HCA) have emerged. These methods have been used to assess the effectiveness of surface water quality monitoring networks, plans to increase the number of examining stations and select the basic parameters of water quality. Along these lines, it is conceivable to identify and evacuate stations as well as repetitive parameters to reduce the economic cost of monitoring plan without sacrific-
ing fluctuation in water quality data. A typical drawback of such techniques is that they can only focus on identifying monitoring locations to be discontinued. This is a consequence of the evaluation being centered on the data of water quality already acquired, ignoring the attributes of the river basin under investigation [13]. Despite this, PCA and HCA have been greatly recognized and widely used in monitoring plans, optimizing spatial sampling points and determining the most appropriate water quality factors [15-24].

The Ganges is an indispensable sacred river in India, especially for the widely practised Hindu religion in the country. In the past few decades, the water quality of the river has deteriorated due to anthropogenic factors, which has led to many national and international organizations to analyze and develop plans to clean the river. In 2009, the National Ganga River Basin Authority (NGRBA) was established under the Environment (Protection) Act of 1986 to clean up the Ganges [25]. The main tributaries are also important contributors to pollution; therefore, conversation strategies should also apply to tributaries.

The tributaries of the Ganges that have been well studied are Gomti, Hindon, Yamuna and Ramganga [26-35]. However, the important tributary in terms of pollution, the Garra River, lacks such an assessment. This river acts as a discharge point for the treated wastewater effluents in adjacent cities. Due to the diversity of the catchment area, anthropogenic and non-anthropogenic pollution sources were discovered along the river.

Therefore, this study investigated the assessment of physicochemical parameters and heavy metals in the water quality of the Garra River and its tributaries and identified the contaminated stretches of the river. This survey was designed to assess the productivity of the Garra River water quality monitoring network using PCA, HCA, Pair sample t-test, and one way ANOVA. Subsequently, it is recommended to identify sampling locations that are critical to the monitoring plan; select the most detailed water quality parameters and assess potential sources, either anthropogenic or natural-based, which will affect the water quality of each sampling location. Therefore, there is a need to contribute to changes in the monitoring plan and to guide executives in decision-making and priorities.

2. Study Area

The Garra River originates from the eastern part of the Nainital district of the Kumaon Himalayas, Uttarakhand state (Fig. 1). The total length of the river is 382 km of which the initial 41 km flowing in the Kumaon Himalayas. The river enters in the Ganga Flood Plains (GFP) at Nanakmatta town, where the famous Nanaksagar dam has been built. After the confluence of the Ramganga, another major tributary of the Ganges, the river is connected to the Ganges River in the Kannauj area of Uttar Pradesh [36-39]. The average elevation of the Garra river is 530 m above sea level (a.s.l) with a total catchment area of 6,832 km².

In GFP, river flow through densely populated districts and highly industrial and agricultural areas of Uttar Pradesh state, such as Pilibhit, Shahjahanpur, Hardoi and Kannauj. The river serves as a source of water for domestic and agricultural purpose. The diversity of the river has implications on the pollution sources as both natural and anthropogenic pollution sources are being present. This makes its study interesting and challenging in identifying various sources of pollution.

2.1. Physiography and Relief

Due to The contrast between the landform, slope and elevation, the entire catchment area shows a large variation in weather and climate (Fig. S1). The annual average precipitation in the catchment area is about 1,000 mm, controlled by Indian monsoon (http://www.indiawaterportal.org/met_data/). The altitude of the river ranges from less than 150m a.s.l for GFP to more than 2,000m a.s.l for the Himalayas (Fig. S1(a) and S1(b)). The minimum slope of the river is shown in GFP, which is about 0-1%, and in the Himalayas, it is greater than 15% (Fig. S1(b)). The minimum slope of the river is shown in GFP, which is about 0-1%, and in the Himalayas, it is greater than 15% (Fig. S1(b)). From Fig. S1(d), it is evident that the population density of the catchment area ranges from below 1,000,000 in the Himalayas to more than 3,000,000 in GFP.

2.2. Geology

In the mountain, the catchment constitutes the major lithotectonic...
zones, namely Sub-Himalayas consisting of siltstone, clays, sandstones and boulders. It shows the characteristics of the molasse sediments of Mid-Miocene to Pleistocene [40]. In GCP, which is closely linked with the extension of the Himalayan orogenic belt, the basin area represents the quaternary lithostratigraphic sequence, including (1) Varanasi Older Alluvium with two facies i.e. sandy facies and silt clay facies, (2) Ganga/Ramganga Terrace Alluvium and (3) Ganga/Ramganga Recent Alluvium. The latter two constitute the Newer Alluvium [41].

3. Materials and Methods

3.1. Description of the Sampling Sites and Analysis

Fig. 1 shows 26 sampling sites that were consciously identified and observed along the 382 km stretch of the Garra River. Five liters of the river water were collected in the month of March, 2014 from each sampling location. According to the standard protocols [42], all samples have been preserved and transferred to the laboratory. Digital meters (HACH Instruments) were used to measure the pH and electrical conductivity (EC) of each water sample. Total dissolved solids (TDS) and temperature were measured using an automatic TDS meter (HACH Instruments) and a mercury thermometer (HACH Instruments), respectively. The Winkler Azide method and the molybdate ascorbic acid method were used to determine chemical oxygen demand (COD) and phosphate (PO4$^{3-}$), respectively (Spectrophotometer, HACH). On the other hand, the dichromate reflux method was used to determine the biological oxygen demand (BOD5). Ion chromatographer (IC) (Metrohm 782 Basic IC) was used to determine the concentration of potassium (K$^+$), magnesium (Mg$^{2+}$), sodium (Na$^+$), calcium (Ca$^{2+}$), lithium (Li$^+$), chloride (Cl$^-$), fluoride (F$^-$), sulfate (SO$_4^{2-}$), nitrate (NO$_3^-$) and bicarbonate (HCO$_3^-$). The ion instrument was calibrated by using a standard solution of cations and anion eluents. The IC detection limit of various cations and anions was 0.1 mg/L.

The concentrations of heavy metals such as iron (Fe), zinc (Zn), and manganese (Mn) were measured by an atomic absorption spectrometer (AAS) (THERMO FISHER MODEL iCE3000) according to the procedure recommended by Environmental Protection Agency (EPA, Method 3005A). All samples were filtered using a 0.45 μm cellulose nitrate membrane and digested with aquaregia (HNO$_3$ 67%: HCl 37% = 3:1). The samples were covered with ribbed watch glass and heated on a hot plate at 90 to 95°C until the volume was reduced to 15-20 mL. All lamps used in AAS for heavy metal analysis were a hollow cathode. The detection limit were 0.05 mg/L, 0.001 mg/L, 0.01 mg/L for Fe, Zn, and Mn, respectively. To maintain the quality of the data of each processed sample, a blank was carried through the analysis and sample preparation. These blanks may help determine if the sample is contaminated.

3.2. Software Used for Statistical Analysis

Excel 2003, a part of the Microsoft Office Suite, SPSS 10 (IBM Corporation, Windows version) and STATISTICA 6 (StatSoft, Inc.) were used. Statistical analysis such as HCA, paired sample t-test, PCA, and analysis of variance (ANOVA) with a confidence level of p < 0.05 were used.

4. Results and Discussion

The examination of the physicochemical and heavy metals characteristics in this study was subjected to statistical analysis and evaluated to their recommended values prescribed in the Bureau of Indian Standard [43] and the World Health Organization [44] (as shown in Table S1).

4.1. Physicochemical Characteristics

Salt solubility, biodegradation, physicochemical parameters and dissolved oxygen rate are exaggerated by temperature [45]; therefore it plays an important role in a riverine system. The temperature of the Garra River and its tributaries range between 24.9°C to 31.8°C with an average of 28.9°C, which is higher than the recommended values by the BIS [43] and WHO [44] (Fig. S2(a)). Generally, the upstream temperature values are lower compared to the middle and downstream. The increase in solar radiation from the morning to noon may be the reason for variation in temperature values (Fig. S2(a)) because the samples were collected at different times between morning and noon.

The pH variation in mainstream water and its tributaries were 6.6 to 8.7 and 7.4 to 8.0 respectively. This confirms that it is slightly alkaline in nature. This is attributable to the HCO$_3^-$ ion concentration, which is relatively high in the tributaries (148.9 mg/L) and the mainstream (151.1 mg/L) (Table S1). Higher pH values of 7.8 (GR2) and 7.6 to 8.7 (GR9-GR20) were recorded in the main river stream, while pH values of 8.1 (T3) and 8.0 (T5) were observed in the tributaries (Fig. S2(b)). Most samples of the Garra River and its tributaries have lower pH values than those set by BIS [43] and WHO [44].

The turbidity in the stream implies the ambiguity of the water due to the presence of suspended matter of clay, silt, waste and other particulate materials. Along the Garra River, the water at the sampling stations GR9 - GR20 was found to be the most turbid, between 178 NTU – 1,355 NTU. GR10 and GR14 exhibited the lowest and highest turbidity values, respectively. In the case of tributaries, the turbidity range is between 111 NTU – 336 NTU, with T1 being the lowest and T6 has the highest turbidity value. Peat water from highlands due to weathering and erosion may be the cause of upstream turbidity. In the middle and downstream of the river, the contribution to turbidity is mainly due to the disturbance of the topsoil by agriculture and other processes such as sand mining, as rivers usually move downwards from hilly areas [46]. Contaminated runoff from industry, agriculture, markets and roads is the main reason for the high turbidity values of these stations (Fig. S2(b)), which are located around urbanized cities such as Pilibhit, Shahjahanpur and Hardoi.

The temperature influences the electrical conductivity (EC) of water, which is an assessment of dissolved solids in water. The average EC values in the water samples from the Garra River and its tributaries (Fig. S2(c)) were 242.6 μS/cm and 222.5 μS/cm, respectively. As can be seen from Fig. S2(c), in the case of River Garra, the highest EC values are 490 μS/cm and 421 μS/cm at GR12 and GR20, respectively, while in the tributaries, the highest values is 340 μS/cm at T4. Fig. S2(c) shows that the change in EC is almost the same as TDS. In the case of the Garra River
and its tributaries, the values of EC shows an increasing pattern from upstream to downstream (Fig. S2(c)). This which may be due to the presence of inorganic dissolved solids such as Cl-, NO3-, SO4²⁻, and PO4³⁻ (Fig. 3(d) and 3(e)), resulting from industrial effluent discharge and agricultural runoff. The EC value of the entire basin is less than 1,000 μS/cm, which is the threshold value for water to be described as unpolluted and fresh.

COD and BOD5 determine the level of organic pollution in lakes and rivers. COD estimates biodegradable and non-biodegradable contamination, while BOD5 defines the biodegradable fractions. The COD and BOD5 values in the water collected from the Garra River ranged from 8 mg/L to 51 mg/L and 6 mg/L to 27 mg/L, respectively. However, in the tributaries, these values ranged from 19 mg/L to 36 mg/L and 11 mg/L to 20 mg/L, respectively (Fig. 2(a)). It can be seen from Fig. 2(a) that the COD and BOD5 values in the middle and downstream of the river increase significantly, which may be due to the high effluent discharge from storm drainage, industries and flushing of sewage directly into the river by urban settlements. The sampling sites GR12 and GR17 show a high increase in values of COD and BOD5, which may be due to the presence of pulp industries, sugar industries and thermal power plants in the vicinity of these sampling sites. The overall values meets the environment protection rules established by the Central Pollution Control Board [47], which describes COD and BOD5 limits of inland waters as well [48]. In surface waters, the assessment of Li⁺ concentration is often overlooked because it naturally exists at low concentration (usually < 0.04 mg/L [49, 50]). The concentration of Li⁺ ranged from 0.111 mg/L to 0.54 mg/L, with an average of 0.19 mg/L in the Garra River. However, in the case of tributaries, the range was between 0.114 mg/L to 0.32 mg/L, with an average of 0.16 mg/L. The higher variation in the values of Li⁺ concentration from upstream to downstream (Fig. S3(a)) suggests that the difference in sampling location has a major control over the Li⁺ content. Weathering of rocks such as sandstone and limestone [51] may be responsible for the concentration of Li⁺ in the upper reaches of the Garra River. In the middle and downstream of the river, anthropogenic sources may be the main cause of Li⁺ concentration, including waste disposal and industrial wastewater, as the use of Li⁺ includes pharmaceuticals, synthetic rubber, chemical manufacturing, lubricants and batteries [52].

Through the weathering processes, Na⁺ and K⁺ are washed away from the soil and rocks in the catchment area. Rivers generally contain about 9 mg/L and 2-3 mg/L of Na⁺ and K⁺, respectively; however, these concentrations may fluctuate due to geological conditions and wastewater contamination (https://www.lenntech.com/periodic/water/). The concentration of Na⁺ and K⁺ in the Garra River ranged from 3.3 mg/L to 47.4 mg/L and 0.5 mg/L to 11.0 mg/L with an average of 21.2 mg/L and 5.8 mg/L, respectively. As seen from the Fig. S3(b), the concentration of Na⁺ and K⁺ in the upper reaches of the river is not very high and may be dissolved through weathering processes. The middle and lower reaches of the river show a large increase in Na⁺ and K⁺ values, which is the result of human activities, because these regions are densely populated and have more industrial and agricultural activities. Wastewater discharged from industries and agricultural runoff is a possible cause of elevated concentrations of Na⁺ and K⁺. Since, Na⁺ is used in various industrial purposes, such as the chemical industries, it is converted to sodium hydroxide, chlorine gas or sodium carbonate, food industries as a flavouring agent or a preservative, leather and textile industries and in cleanser and soap production. K⁺ is commonly used in synthetic fertilizers, glass production, fluid soap production, medicines, photography tanning, etc.

4.1.1. Concentration of major ions and its possible sources
Li⁺ elements are found in nature mostly in the clay fraction of the soil, potable water, surface water, groundwater and marine waters as well [48]. In surface waters, the assessment of Li⁺ concentration is often overlooked because it naturally exists at low concentration (usually < 0.04 mg/L [49, 50]). The concentration of Li⁺ ranged from 0.111 mg/L to 0.54 mg/L, with an average of 0.19 mg/L in the Garra River. However, in the case of tributaries, the range was between 0.114 mg/L to 0.32 mg/L, with an average of 0.16 mg/L. The higher variation in the values of Li⁺ concentration from upstream to downstream (Fig. S3(a)) suggests that the difference in sampling location has a major control over the Li⁺ content. Weathering of rocks such as sandstone and limestone [51] may be responsible for the concentration of Li⁺ in the upper reaches of the Garra River. In the middle and downstream of the river, anthropogenic sources may be the main cause of Li⁺ concentration, including waste disposal and industrial wastewater, as the use of Li⁺ includes pharmaceuticals, synthetic rubber, chemical manufacturing, lubricants and batteries [52].

The intensity of non-compound, free oxygen present in the river water is called dissolved oxygen (DO). It is an important parameter for investigating water quality because of its impact on organisms living in water. Too low or too high level of DO can affect water quality and endanger aquatic life. The DO values did not show large spacial variations in the Garra River and its tributaries, ranging from 7.38 mg/L to 8.92 mg/L and 7.3 mg/L to 8.67 mg/L, respectively (Fig. 2(b)). This indicates that water quality is not harmful to the aquatic environment.

4.1.1. Concentration of major ions and its possible sources
Li⁺ elements are found in nature mostly in the clay fraction of the soil, potable water, surface water, groundwater and marine waters as well [48]. In surface waters, the assessment of Li⁺ concentration is often overlooked because it naturally exists at low concentration (usually < 0.04 mg/L [49, 50]). The concentration of Li⁺ ranged from 0.111 mg/L to 0.54 mg/L, with an average of 0.19 mg/L in the Garra River. However, in the case of tributaries, the range was between 0.114 mg/L to 0.32 mg/L, with an average of 0.16 mg/L. The higher variation in the values of Li⁺ concentration from upstream to downstream (Fig. S3(a)) suggests that the difference in sampling location has a major control over the Li⁺ content. Weathering of rocks such as sandstone and limestone [51] may be responsible for the concentration of Li⁺ in the upper reaches of the Garra River. In the middle and downstream of the river, anthropogenic sources may be the main cause of Li⁺ concentration, including waste disposal and industrial wastewater, as the use of Li⁺ includes pharmaceuticals, synthetic rubber, chemical manufacturing, lubricants and batteries [52].

Through the weathering processes, Na⁺ and K⁺ are washed away from the soil and rocks in the catchment area. Rivers generally contain about 9 mg/L and 2-3 mg/L of Na⁺ and K⁺, respectively; however, these concentrations may fluctuate due to geological conditions and wastewater contamination (https://www.lenntech.com/periodic/water/). The concentration of Na⁺ and K⁺ in the Garra River ranged from 3.3 mg/L to 47.4 mg/L and 0.5 mg/L to 11.0 mg/L with an average of 21.2 mg/L and 5.8 mg/L, respectively. As seen from the Fig. S3(b), the concentration of Na⁺ and K⁺ in the upper reaches of the river is not very high and may be dissolved through weathering processes. The middle and lower reaches of the river show a large increase in Na⁺ and K⁺ values, which is the result of human activities, because these regions are densely populated and have more industrial and agricultural activities. Wastewater discharged from industries and agricultural runoff is a possible cause of elevated concentrations of Na⁺ and K⁺. Since, Na⁺ is used in various industrial purposes, such as the chemical industries, it is converted to sodium hydroxide, chlorine gas or sodium carbonate, food industries as a flavouring agent or a preservative, leather and textile industries and in cleanser and soap production. K⁺ is commonly used in synthetic fertilizers, glass production, fluid soap production, medicines, photography tanning, etc.

Multivalent cations and anions are the cause of water hardness, and in most cases, this is due to Mg²⁺ and Ca²⁺. In this study,
the Mg\(^{2+}\) and Ca\(^{2+}\) hardness are expressed as Eq. (1) \[46\] in terms of the equivalent of calcium carbonate.

\[
\text{Hardness (as CaCO}_3\text{)} = \frac{50}{\text{Eq. Wt. of M}^2}\times M^2
\]  

(1)

Among them, M\(^{2+}\) stands for divalent metallic ion. The total hardness is described as an increase in the hardness of Mg\(^{2+}\) and Ca\(^{2+}\), which is 65 to 70 mg/L as CaCO\(_3\) in the upstream (GR1, GR2, GR3, GR4) and 75 to 144 mg/L as CaCO\(_3\) in the middle and downstream (GR5 to GR20 and T1 to T6) (Fig. 3(f)). *Potasznik and Szymczyk \[53\] and Chapra et. al. \[54\] observed that the concentration of Mg\(^{2+}\) has been increased in the surface water due to the presence of sedimentary rocks, mainly dolomite. Weathering of rocks such as sandstone may be responsible for the concentration of Mg\(^{2+}\) in the upper reaches of the Garra River.

WHO \[44\] classifies hardness ranges >300, 150–300, 75–150 and 0–75 into very hard, hard, moderately hard and soft, respectively. Therefore, the water of the mainstream and its tributaries is in soft to moderately hard category.

The amounts of F\(^-\) composition in the upper earth crust is soluble in water; therefore, it is usually present in groundwater and surface waters \[55\]. Fig. 3(a) shows that the F\(^-\) concentrations has a spatial variation ranging from 0.10 mg/L to 0.60 mg/L with an average concentration of 0.27 mg/L. The upstream F\(^-\) concentration is lower than downstream of the river, which may be the result of natural sources, such as geological, chemical, physical and the porosity and acidity of the soil and rock. The effect of anthropogenic sources on F\(^-\) concentration can be seen in the middle and downstream of the river (Fig. 3(a)), mainly from GR14 to GR20. This may be due to the presence of coal-burning from the thermal power plant, brick-making industries, and the use of phosphatic fertilizer in agricultural fields \[56\].

The contents of Cl\(^-\) and SO\(_4^{2-}\) in the Garra River ranged from 0.43 mg/L to 8.33 mg/L and from 0.97 mg/L to 32.71 mg/L, with an average concentration of 3.42 mg/L and 9.06 mg/L, respectively. From Fig. 3b and 3c, it is evident that the concentrations of Cl\(^-\) and SO\(_4^{2-}\) in the upstream are lower compared to the downstream of the river and can be attributed to the same reason as F. Typically, the SO\(_4^{2-}\) and Cl\(^-\) ions in the river water are introduced by effluent from the anaerobic wastewater treatment plant and chloride used in the tertiary treatment, respectively. The absence of these wastewater treatment plants may also be a possible cause of Cl\(^-\) and SO\(_4^{2-}\) concentration throughout entire sampling period. However, the average values of Cl\(^-\) and SO\(_4^{2-}\) in river water are within the limits specified by BIS \[43\] and WHO \[44\].

PO\(_4^{3-}\) in rivers is usually the result of runoff from agricultural fields sprayed with fertilizers and effluent discharge from sewage treatment plants \[3, 57\]. Another anthropogenic: source is water containing dissolved detergents which are discharged into the river and increase PO\(_4^{3-}\) concentration of the river due to PO\(_4^{3-}\) additive (an essential ingredient in detergent). The concentration of PO\(_4^{3-}\) in the river ranged from 0.02 mg/L to 1.47 mg/L, with an average of 0.35 mg/L (Fig. 3(d)). Generally, the PO\(_4^{3-}\) concentration is found to be higher in the downstream region of the river. These high values of PO\(_4^{3-}\) may be the result of agricultural runoff, as the area is dominated by agricultural practices.

The main source of NO\(_3^-\) in river is decaying legumes, but the oxidation of animal and human excreta also contribute a lot. In addition, municipal emissions are also a very high contributor.
especially for facilities that do not have a nitrogen removal method.
In the Garra River, the NO$_3^-$ concentration ranged from 0.09 mg/L to 14.24 mg/L, with an average concentration of 4.99 mg/L (Fig. 3(e)). The NO$_3^-$ concentration in all samples collected during the campaign was recorded below the BIS [43] and WHO [44] limit of 45 mg/L and 50 mg/L, respectively, for methaemoglobinaemia in children. Therefore, from this perspective, it makes water safe.

The concentration of HCO$_3^-$ in the Garra River ranged from 80.8 mg/L to 193.5 mg/L, with an average concentration of 150.64 mg/L (Fig. 3(f)). In the upper reaches of the river, the concentration of HCO$_3^-$ is mainly attributed to the weathering of silicates and carbonates rock. However, in the middle and downstream regions, these concentrations increase and the effects of anthropogenic activity may be a possible factor.

### 4.2. Stretch Wise Classification of the River Water

To understand the overall quality of water in the Garra River catchment, the entire basin is divided into four sections, and the water quality parameters of each section are averaged. The initial stretch of the Garra River basin contains rivers flowing through the lower part of the Kumaon Himalayas. The water quality along this stretch was found to be better than the other stretches. The average concentrations of COD and BOD$_5$ in this section of the river were 15.6 mg/L and 7.4 mg/L, respectively. Compared with other sections, ions like F-, SO$_4^{2-}$ and NO$_3^-$ in the water collected from this stretch were also found to be the lowest (Table 1). These results can be attributed to relatively few anthropogenic, agricultural and industrial activities in this stretch.

In terms of organic content, the third and fourth sections of the Garra River from Bisalpur to Shahjahanpur and Shahjahanpur to Kannauj were found to be the most polluted. For the mainstream, the average concentrations of COD and BOD$_5$ in these areas are > 40 mg/L and > 21 mg/L, respectively. Similarly, for tributaries, these values are > 22 mg/L and > 15 mg/L, respectively. The high values in these sections are mainly due to the human influence, because the catchment areas between these regions are densely populated, and there are several industries, such as sugar, paper, thermal power plant, and direct discharge from sewage treatment plants into river.

### 4.3. Statistical Analysis of the Data

#### 4.3.1. Pair sample t-test

Paired sample t-test, sometimes referred to as dependent sample t-test. This is a statistical procedure used to determine if the average difference between the two sets of perceptions is zero. The results obtained for the t-test are shown in Table S2. Except for Ca$^{2+}$ and HCO$_3^-$, all other parameters (COD, BOD$_5$, F, Cl, SO$_4^{2-}$, PO$_4^{3-}$, NO$_3^-$, Li, Na, K, and Mg$^{2+}$) were subjected to the t-test do not show any critical variation in their means. In the t-test, the p-value was found to be higher than the significant value (p < 0.05) in case of all parameters except HCO$_3^-$. Therefore, we can conclude that there is no direct contributor of organic (COD, BOD$_5$), nutrient (PO$_4^{3-}$, NO$_3^-$), minerals (F, Cl, SO$_4^{2-}$, Li, Na, K, Ca$^{2+}$ and Mg$^{2+}$) content to the mainstream except tributaries water. However, HCO$_3^-$ in the river was more than tributaries (t-static= -3.802 and p<0.05). The difference may
be due to the removal of free and combined carbon dioxide in the river by algae, which produced more carbonate [46].

4.3.2. Hierarchical cluster analysis

In order to finalize the statistical combination of the Garra River sampling sites, HCA was used. The result of this analysis led to a dendrogram (Fig. 4) having a significance at (Dlink/Dmax) x 100 < 70. All 20 sampling locations are ultimately divided into three clusters as there were sampling sites with similar water quality and sources of pollution in these groups, and sufficient water quality information (physical, chemical, nutrient, and organic) was available.

The sampling stations (GR1 to GR8) belong to cluster 1, and pollution level is low. Due to the quiet anthropogenic, agricultural and industrial activities in the region, natural processes such as weathering and rock erosion are the main sources of pollution. Cluster 2 consists of sampling stations (GR9, GR16, GR18, GR19 and GR20) which show high levels of pH, EC, TDS, NO3-, PO4\(^{3-}\), SO4\(^{2-}\), F, Cl, Li\(^+\), Na\(^+\), K\(^+\), Ca\(^{2+}\), Mg\(^{2+}\) and HCO3\(^-\). However, turbidity, COD and BOD\(_5\) showed low concentration in this cluster. The sampling stations (GR1 to GR8) belong to cluster 1, and pollution level is low. Due to the quiet anthropogenic, agricultural and industrial activities in the region, natural processes such as weathering and rock erosion are the main sources of pollution. Cluster 2 consists of sampling stations (GR9, GR16, GR18, GR19 and GR20) which show high levels of pH, EC, TDS, NO3-, PO4\(^{3-}\), SO4\(^{2-}\), F, Cl, Li\(^+\), Na\(^+\), K\(^+\), Ca\(^{2+}\), Mg\(^{2+}\) and HCO3\(^-\). However, turbidity, COD and BOD\(_5\) showed low concentration in this cluster. The sampling points belong to this cluster flow in low-lying areas, with point and non-point sources, such as various industrial emissions near Hardoi and Kannauj, agricultural activities, urban sewage and discharge from wastewater treatment plants. Cluster 3 consisting of sampling stations (GR10, GR11, GR12, GR13, GR15 and GR17) showed a lower pollution level than cluster 2 for all parameters, but the turbidity, COD, and BOD\(_5\) concentrations were high in this cluster. The reason for the high concentration of these parameters in this cluster may be due to the discharge of industrial effluents from industries such as paper and pulp, fertilizer, sugar and distillery. The sampling station GR14 does not belong to any cluster and it shows a very high concentration of all parameters. This may be due to the discharge of wastewater from the thermal power plant [58] at Rosa town; therefore, it should be sampled separately in future sampling. These results indicate that for grouping of sampling stations, cluster analysis is a useful method consistent with its interrelated features, thus saving time, cost-effectiveness, and most favourable river sampling in the future. Kim and Mueller [59] and Kazi et al. [60] also used the same procedure in their study to assess potential observed water quality plans.

4.3.3. Source Identification of Pollution by Principal Component Analysis (PCA) and Correlation Matrix

The spatial correlation matrix of the water quality parameters is obtained by PCA, as shown in Table S3. Most of the parameters in Table S3 show a positive correlation with other parameters, except for a few such as Li\(^+\) and DO. Every value shows the changeability of a parameter for another parameter in the matrix, and the underlined values show a high and positive correlation, pH and EC are significantly positively correlated with Mg\(^{2+}\) and TDS, respectively. TDS is significantly positively correlated with Ca, HCO3\(^-\) and Cl\(^-\). COD and BOD\(_5\) are positively correlated with each other. DO and Li\(^+\) do not show any significant correlation with any parameters. NO3\(^-\) shows a good and positive correlation with PO4\(^{3-}\) and Mg\(^{2+}\).

PO4\(^{3-}\) also showed a good and positive correlation with parameters such as Ca, Mg\(^{2+}\), and HCO3\(^-\). Na\(^+\) is positively correlated with K\(^+\). Ca is significantly positively correlated with Mg\(^{2+}\) and HCO3\(^-\); however, Mg\(^{2+}\) shows a good and positive correlation with HCO3\(^-\), F, and Cl\(^-\). SO4\(^{2-}\) also showed a good and positive correlation with F and Cl\(^-\). The combined insightful connections between physicochemical parameters indicates that water quality is primarily controlled by carbonate weathering. There are likewise chances of the existence of dolomite, a common source of Mg\(^{2+}\) and Ca\(^{2+}\). A significant correlation between F\(^-\) and Ca\(^{2+}\) also indicates the presence of minerals like fluorapatite, apatite, and fluor spar. At point scales, geochemistry can be controlled by anthropogenic sources, such as mass bathing and the use of excess fertilizers, and can be revealed by a significant correlations between Cl\(^-\), SO4\(^{2-}\) and NO3\(^-\) [61]. In this way, it can be seen that anthropogenic, geogenic and natural processes primarily control the hydro-geochemistry of the Garra River.

The scree plot (Fig. S4) was utilized to distinguish the number of principal components (PCs) to understand the basic structure of the data [62]. After the 5\(^{th}\) eigenvalue, the slope indicated by scree plot changes significantly. Cattell and Jaspers [63] recommend using each PC and the 1\(^{st}\) PC after the break, with the goal of retaining four PCs, whose eigenvalues are more noteworthy than solidarity, and clarify 81.5% of the fluctuations or data contained in the original data set.

Loadings refers to the projection of the original variable on the subspace of the PC and correspond to the correlation coefficient between the variable and the PC. There were three PCs with eigenvalues > 1, accounting for about 77.61% of the total data set variation. Table 2 shows the loading of six maintained PCs.

PC1 explained a variance of 53.6%, and most parameters such as F\(^-\), Cl\(^-\), pH, HCO3\(^-\), NO3\(^-\), SO4\(^{2-}\), PO4\(^{3-}\), EC, TDS, Ca\(^{2+}\), Mg\(^{2+}\), and BOD were significantly involved (Table 2). The starting point of these variables may be shared and may form the weathering of marl, gypsum and limestone soils. Considering the large amount of oxygen consumed by the increase in dissolved organic matter,
the positive loading of COD and BOD in PC1 can be elucidated because organic matter is mainly composed of carbohydrates in urban wastewater, and organic acids. DO have negative participation, and Li- shows positive but not important participation in PC1. PC2 explained a variance of 15.8%, including Na^+, Li^- and DO (positive loading), and NO_3^- (negative participation). PC3, PC4 and PC5 explained 8.3%, 5.7% and 5.5% of the total variance, respectively.

4.3.4. Spatial variations in quality of river water by analysis of variance (one-way ANOVA)

The importance of spatial variation in physicochemical parameters between the Garra River and its tributaries and between each sampling location is determined by one-way ANOVA statistical method. When the p value was less than 0.05 (p < 0.05), a significant change was considered. Table 3 shows the water sample variations analyzed by ANOVA between the Garra River and its tributaries. Except for turbidity, K^+, and HCO_3^- all other parameters did not show any significant change. The variation in turbidity, K^+ and HCO_3^- may be due to leaching of the rocks along the path of the Garra river, while the tributaries have local flow. However, in the case of spatial variation between each sampling point, a highly significant variation is exhibited by turbidity, NO_3^-, Na^+, K^+, Mg^2+ and HCO_3^-.

### Table 2. Loadings of 17 Experimental Variables on Six Significant Principal Components for 20 River Water Samples

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg^2+</td>
<td><strong>0.912</strong></td>
<td>-0.216</td>
<td>-0.187</td>
<td>0.097</td>
<td>-0.120</td>
<td>-0.104</td>
</tr>
<tr>
<td>Cl^-</td>
<td><strong>0.868</strong></td>
<td>-0.065</td>
<td>0.034</td>
<td>0.119</td>
<td>0.171</td>
<td>-0.067</td>
</tr>
<tr>
<td>TDS</td>
<td><strong>0.866</strong></td>
<td>-0.021</td>
<td>0.430</td>
<td>-0.129</td>
<td>0.157</td>
<td>0.084</td>
</tr>
<tr>
<td>HCO_3^-</td>
<td><strong>0.866</strong></td>
<td>0.041</td>
<td>0.154</td>
<td>0.259</td>
<td>-0.258</td>
<td>-0.124</td>
</tr>
<tr>
<td>Ca^2+</td>
<td><strong>0.861</strong></td>
<td>0.166</td>
<td>0.136</td>
<td>0.335</td>
<td>-0.131</td>
<td>0.085</td>
</tr>
<tr>
<td>EC</td>
<td><strong>0.808</strong></td>
<td>-0.030</td>
<td>0.263</td>
<td>-0.254</td>
<td>0.161</td>
<td>0.402</td>
</tr>
<tr>
<td>PO_4^3-</td>
<td><strong>0.798</strong></td>
<td>-0.209</td>
<td>-0.028</td>
<td>0.344</td>
<td>-0.288</td>
<td>0.283</td>
</tr>
<tr>
<td>SO_4^2-</td>
<td><strong>0.766</strong></td>
<td>0.088</td>
<td>-0.283</td>
<td>0.123</td>
<td>0.507</td>
<td>-0.177</td>
</tr>
<tr>
<td>BOD</td>
<td><strong>0.737</strong></td>
<td>0.374</td>
<td>-0.242</td>
<td>-0.369</td>
<td>-0.081</td>
<td>0.153</td>
</tr>
<tr>
<td>F</td>
<td><strong>0.730</strong></td>
<td>-0.322</td>
<td>-0.246</td>
<td>0.077</td>
<td>0.468</td>
<td>-0.004</td>
</tr>
<tr>
<td>pH</td>
<td><strong>0.726</strong></td>
<td>-0.299</td>
<td>0.028</td>
<td>-0.309</td>
<td>-0.022</td>
<td>-0.437</td>
</tr>
<tr>
<td>NO_3^-</td>
<td><strong>0.721</strong></td>
<td>-0.583</td>
<td>0.067</td>
<td>-0.076</td>
<td>-0.102</td>
<td>0.098</td>
</tr>
<tr>
<td>COD</td>
<td><strong>0.647</strong></td>
<td>0.468</td>
<td>-0.214</td>
<td>-0.497</td>
<td>-0.183</td>
<td>0.062</td>
</tr>
<tr>
<td>K^+</td>
<td><strong>0.629</strong></td>
<td>0.538</td>
<td>0.197</td>
<td>0.017</td>
<td>-0.246</td>
<td>-0.384</td>
</tr>
<tr>
<td>Na^+</td>
<td>0.540</td>
<td><strong>0.742</strong></td>
<td>0.209</td>
<td>0.129</td>
<td>0.124</td>
<td>0.054</td>
</tr>
<tr>
<td>Li^-</td>
<td>0.035</td>
<td><strong>0.686</strong></td>
<td>-0.662</td>
<td>0.195</td>
<td>0.009</td>
<td>0.099</td>
</tr>
<tr>
<td>DO</td>
<td>-0.438</td>
<td><strong>0.583</strong></td>
<td>0.551</td>
<td>0.056</td>
<td>0.267</td>
<td>-0.017</td>
</tr>
</tbody>
</table>

| **Eigenvalue** | **9.110** | **2.675** | **1.408** | **0.966** | **0.935** | **0.706** |
| **% Variance explained** | **53.590** | **15.736** | **8.285** | **5.682** | **5.501** | **4.152** |
| **% Cum. variance** | **53.590** | **69.327** | **77.611** | **83.293** | **88.794** | **92.946** |

#Bold values denote loadings of interest

---

### Table 3. The Result of ANOVA Analysis between the Garra River and Its Tributaries and between Each Sampling Location

<table>
<thead>
<tr>
<th>Variables</th>
<th>Between Garra River and its tributary</th>
<th>Between each point</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>NS(P = 0.207)</td>
<td>NS(P = 0.283)</td>
</tr>
<tr>
<td>Turbidity (NTU)</td>
<td><em><strong>(P = 0.000)</strong></em></td>
<td><em><strong>(P = 0.001)</strong></em></td>
</tr>
<tr>
<td>EC (μS/cm)</td>
<td>NS(P = 0.066)</td>
<td>NS(P = 0.087)</td>
</tr>
<tr>
<td>TDS (mg/L)</td>
<td>NS(P = 0.198)</td>
<td>NS(P = 0.066)</td>
</tr>
<tr>
<td>COD (mg/L)</td>
<td>NS(P = 0.077)</td>
<td>NS(P = 0.178)</td>
</tr>
<tr>
<td>BOD (mg/L)</td>
<td>NS(P = 0.151)</td>
<td>NS(P = 0.058)</td>
</tr>
<tr>
<td>DO (mg/L)</td>
<td>NS(P = 0.436)</td>
<td>NS(P = 0.129)</td>
</tr>
<tr>
<td>NO_3^- (mg/L)</td>
<td>NS(P = 0.066)</td>
<td><em><strong>(P = 0.003)</strong></em></td>
</tr>
<tr>
<td>PO_4^3- (mg/L)</td>
<td>NS(P = 0.690)</td>
<td>NS(P = 0.081)</td>
</tr>
<tr>
<td>Li (mg/L)</td>
<td>NS(P = 0.058)</td>
<td>NS(P = 0.296)</td>
</tr>
<tr>
<td>Na (mg/L)</td>
<td>NS(P = 0.094)</td>
<td>*(P = 0.031)</td>
</tr>
<tr>
<td>K (mg/L)</td>
<td><em><strong>(P = 0.039)</strong></em></td>
<td><em><strong>(P = 0.004)</strong></em></td>
</tr>
<tr>
<td>Ca (mg/L)</td>
<td>NS(P = 0.143)</td>
<td>NS(P = 0.836)</td>
</tr>
<tr>
<td>Mg (mg/L)</td>
<td>NS(P = 0.134)</td>
<td><em><strong>(P = 0.002)</strong></em></td>
</tr>
<tr>
<td>HCO_3^- (mg/L)</td>
<td>*(P = 0.041)</td>
<td><em><strong>(P = 0.004)</strong></em></td>
</tr>
</tbody>
</table>

#NS=non significant; ***= highly significant; *= significant
4.4. Concentration of Heavy Metals

Heavy metal pollution in the river water is one of the most important concerns of rural and urban communities in developing nations. The entry of these toxic heavy metals into the environment may lead to bioaccumulation and biomagnifications [64]. Heavy metals pollution in the riverine environments has received global attention due to its persistence, abundance and environmental toxicity [65, 66]. To observe the concentration of heavy metals in the Garra River, Fe, Zn, and Mn in the collected water samples were analysed in AAS. The analysis showed that all samples contained Fe, Zn and Mn, including mainstream and tributaries.

Fig. 5 (a) and (b) show the concentration of Fe, Zn and Mn in the Garra River and its tributaries. As can be seen from the figure, the concentration of Fe in the mainstream and the tributaries is very high, while the concentrations of Zn and Mn are also high, but lower compared with Fe. The concentration of Fe ranges from 0.205 mg/L to 55.99 mg/L, with an average concentration of 17.31 mg/L in the Garra River. In the case of tributaries, the concentration of Fe is 0.73 mg/L to 51.92 mg/L, with an average concentration of 16.94 mg/L. In case of the Garra River, the concentrations of Zn and Mn ranges from 0.002 mg/L to 0.84 mg/L and 0.01 mg/L to 0.83 mg/L, with an average concentration of 0.14 mg/L and 0.25 mg/L, respectively. However, in the case of tributaries, they are in the range of 0.02 mg/L to 0.18 mg/L and 0.005 mg/L to 0.69 mg/L, with an average concentration of 0.10 mg/L and 0.24 mg/L, respectively.

As can be seen from Fig. 5(a) and (b), the concentrations of Fe, Zn and Mn show an increased pattern from upstream to downstream. The upstream Fe concentration is mainly from natural iron sources, which may be due to the presence of silicate rock weathering [67] in the upper part of the basin. The middle stream and downstream of the Garra River and its tributaries are contaminated with medium to high heavy metals at several locations, which may be the results of the dense population (Fig. S1(d)) and high industrial and agricultural activity in these stretches of the Garra River. The primary anthropogenic sources are the discharge of partially treated and untreated industrial wastewater and sewage containing toxic metals. In addition, it comes from the widespread use of metal chelates in a wide range of industries and pesticides and fertilizers containing heavy metals in the agricultural sector [68, 69]. The average value of Zn in case of the Garra River and its tributaries is much lower than the approved limit of BIS [43]. However, the Fe and Mn concentrations in mainstream and tributaries show high values from the drinking limits recommended by BIS [43] and WHO [44].

5. Conclusions

The study was designed to assess the water quality parameters of the Garra River and its tributaries. The sampling points located downstream of the river show that the concentration of PO\(_4^{3-}\) is as high as 0.4 - 0.5 mg/L. The downstream section of the river in GFP from Bisalpur to Shahjahanpur and from Shahjahanpur to Kannauj contains more SO\(_4^{2-}\) and organic content compared to the upstream section in GFP and Kumaon hills. A paired sample t-test indicating that HCO\(_3^-\) in the river was more than tributaries. Approximately 77% of the total variation can be explained by extracted PC, which has an eigenvalue > 1 and shows the positive and heavy loadings of the river’s mineral portion. One-way ANOVA clarified turbidity, K\(^+\) and HCO\(_3^\) are parameters having considerable variation between the Garra River and its tributaries. However, NO\(_3^-\), Na\(^+\), K\(^+\), Mg\(^+\) and HCO\(_3^-\) are parameters that show significant variation between each location of the Garra River.

Turbidity, Li\(^+\), HCO\(_3^-\), Fe, and Mn are the most important parameters affecting the water quality of the Garra river. The potential sources of in the Garra River is the dissolution of rocks through chemical and physical weathering processes in the upstream of the river. In the middle and lower reaches of the river, pollution is mainly due to the human influence as the catchment area in between is densely populated. The other sources of pollution in the middle and lower reaches of the river is the presence of various industries, such as sugar industries, paper industries, thermal power plant and sewage discharged directly from sewage treatment plants into rivers. The downstream region of the river containing sample locations from GR9 to GR20 shows high levels of contamination in terms of physicochemical parameters and heavy metals.

These methods can help water quality monitoring agencies to improve current monitoring program(s) by reducing the number of monitoring stations to the ones that best represent the spatial pattern of river water quality. Ultimately, this will help reduce the time, effort, and cost of assessing river water quality.
Some of the limitations in the conclusions drawn from this study may be that large data sets that multivariate techniques need to provide meaningful results are not available. However, as the first study on the Garra River Basin, our research can be used as a case study to further strengthen the research findings of river water quality. In addition, results based on descriptive statistics are also presented in our investigation, which will help to assess comparatively future datasets with this study.

**Acknowledgments**

We would like to thank to deanship of scientific research (DSR) at King Abdulaziz University (KAU)-Jeddah, Saudi Arabia for providing funding support (D1441-559-145) in completion of this article.

**Author Contributions**

M.Y.A.K. (Assistant Professor) conducted all the experiments and wrote the manuscript. M.Y.A.K. (Assistant Professor) and J.W. (Research Associate) revised and edited the manuscript.

**References**

27. Singh KP, Malik A, Sinha S. Water quality assessment and