Modelling spatio-temporal heterogeneities in groundwater quality in Ghana: a multivariate chemometric approach

Frederick Ato Armah, Arnold Paintsil, David Oscar Yawson,
Michael Osei Adu and Justice O. Odoi

ABSTRACT

Chemometric techniques were applied to evaluate the spatial and temporal heterogeneities in groundwater quality data for approximately 740 goldmining and agriculture-intensive locations in Ghana. The strongest linear and monotonic relationships occurred between Mn and Fe. Sixty-nine per cent of total variance in the dataset was explained by four variance factors: physicochemical properties, bacteriological quality, natural geologic attributes and anthropogenic factors (artisanal goldmining). There was evidence of significant differences in means of all trace metals and physicochemical parameters ($p < 0.001$) between goldmining and non-goldmining locations. Arsenic and turbidity produced very high value F's demonstrating that 'physical properties and chalcophilic elements' was the function that most discriminated between non-goldmining and goldmining locations. Variations in *Escherichia coli* and total coliforms were observed between the dry and wet seasons. The overall predictive accuracy of the discriminant function showed that non-goldmining locations were classified with slightly better accuracy (89%) than goldmining areas (69.6%). There were significant differences between the underlying distributions of Cd, Mn and Pb in the wet and dry seasons. This study emphasizes the practicality of chemometrics in the assessment and elucidation of complex water quality datasets to promote effective management of groundwater resources for sustaining human health.

Key words | bacteriological, discriminant analysis, groundwater, negative log-log, physicochemical, regression

Frederick Ato Armah (corresponding author)
Department of Environmental Science, School of
    Biological Sciences, College of Agriculture &
    Natural Sciences,
University of Cape Coast,
Cape Coast, Ghana
E-mail: farmah@ucc.edu.gh

Arnold Paintsil
Department of Civil and Environmental
    Engineering, Faculty of Engineering, Spencer
    Engineering Building, Western University,
London,
Ontario N6A 5B9, Canada

David Oscar Yawson
Department of Soil Science, School of Agriculture,
    College of Agriculture & Natural Sciences,
University of Cape Coast,
Cape Coast, Ghana

Michael Osei Adu
Department of Crop Science, School of Agriculture,
    College of Agriculture & Natural Sciences,
University of Cape Coast,
Cape Coast, Ghana

Justice O. Odoi
Nature Today Ghana,
P.O. Box OS 1455,
Osu,
Accra, Ghana

INTRODUCTION

Although groundwater is generally less susceptible to contamination than surface waters, it is usually more highly mineralized in its natural state (Chapman 1996). As water moves slowly through the ground it can remain for extended periods of time in contact with minerals present in the soil and bedrock and become saturated with dissolved solids from these minerals (Chapman 1996). This dissolution process continues until chemical equilibrium is reached between the water and the minerals with which it is in contact. The types and relative concentrations of the chemical constituents in groundwater provide information on the evolution of groundwaters, age (residence time), solubility, rates of movement, flow history, and sources of recharge and pollution (Chapman 1996; Edmunds *et al.* 2003; Purtschert 2009). According to Morris *et al.* (2003), an estimated two billion people across the world depend on aquifers for a drinking water supply. In a rural context, groundwater provides the mainstay for agricultural irrigation and will be the key to providing additional resources for food security (Foster & Chilton 2003; Morris *et al.*

2003; Arias-Estévez et al. 2008). In urban centres, groundwater supplies are important as a source of relatively low cost and generally high quality municipal and private domestic water supply (Foster & Chilton 2003; Morris et al. 2003; Arias-Estévez et al. 2008). The importance of groundwater quality has become increasingly recognized as development of groundwater continues to expand in developing countries such as Ghana. Monitoring of groundwater quality is concomitantly becoming more important because of natural and anthropogenic contamination concerns and development of new equipment and techniques for measuring contaminants in minute concentrations (Jordana & Piera 2004; Pinder & Celia 2006; Armah 2014).

The need to expand research and monitoring efforts and develop a comprehensive, consistent, and reliable database from which to better understand and characterize existing conditions in groundwater, identify existing and potential problems, establish priorities, and develop viable groundwater policies and strategies is at the very least compelling but it is a complex challenge. In this context, Gazzaz et al. (2012) mention that water analyses usually generate bulky and complex datasets constituting large numbers of samples and water quality parameters whose analysis and interpretation using univariate and bivariate statistical methods can be far from complete. This brings into sharp focus the importance of chemometric approaches. Previous research has demonstrated that chemometric techniques are useful tools for the extraction of considerable, meaningful information from environmental data (Zhou et al. 2007; Giridharan et al. 2009; Wu et al. 2010; Gazzaz et al. 2012; Kumari et al. 2013). Such techniques help in identifying possible factors that affect water quality and also help in determining solutions to these problems (Varol et al. 2012).

Notwithstanding the utility of chemometrics in the assessment of water quality data at the multivariate level, currently our understanding of groundwater quality is partial with a substantial assortment of missing pieces. Existing programs of acquiring and managing groundwater monitoring data in developing countries such as Ghana are inadequate to meet groundwater quality challenges. Furthermore, with few exceptions (see Armah 2014), only a few studies have assessed groundwater quality in Ghana using a combination of multivariate statistical approaches to elicit a comprehensive picture of the spatial and temporal distributions and the complex interplay of physicochemical factors, trace metals, and bacteriological parameters. This is a fundamental motivation for this study. Three broad objectives were formulated to guide this study. The first was to assess zero-order relationships among physicochemical and trace metals in the original and transformed groundwater data from 738 locations in Ghana. This was achieved via Spearman and Pearson product moment correlations, respectively. Second, the study evaluated the spatial and temporal variability of groundwater quality parameters. This objective was met using principal component analysis/ factor analysis (PCA/FA) and Wilcoxon rank-sum (Mann–Whitney) test, respectively. Third, the pollution status of each location was investigated using negative log-log regression and discriminant analysis (DA).

MATERIALS AND METHOD

Study area

The exact location from which groundwater tests were collected and a thorough exposition of the study site as well as an exhaustive representation of the study zone are very much archived by Armah (2014). Briefly, the monitoring locales for this study were of the following distribution: Ayanfuri (16), Savelugu (173), Tolon (47), Damang (6), Teberebie (85), Bogoso (40), Prestea (166), and Tarkwa (205). Ayanfuri together with Bogoso, Damang, Tarkwa, Prestea, and Teberebie constitute the localities in southwestern Ghana, as shown in Figure 1. The geographical physiognomies of this zone is exemplified by greenstone belts made out of mafic volcanic rocks and mediating basins commonly comprising fine-grained, subterranean marine silt transformed at greenschist facies. Three fundamental gold deposits, including placer or alluvial deposits, non-sulphidic pale placer or free milling mineral ore, and oxidized mineral ore are found in the region (Kortatsi 2004; Bhattacharya et al. 2012). Alluvial deposits are found in stream-depleting regions with gold-containing deposits occurring in places especially in Biriminan rock zones where the bedrock has marginally been transformed and influenced by Dixcove sandstone (Kortatsi 2004).
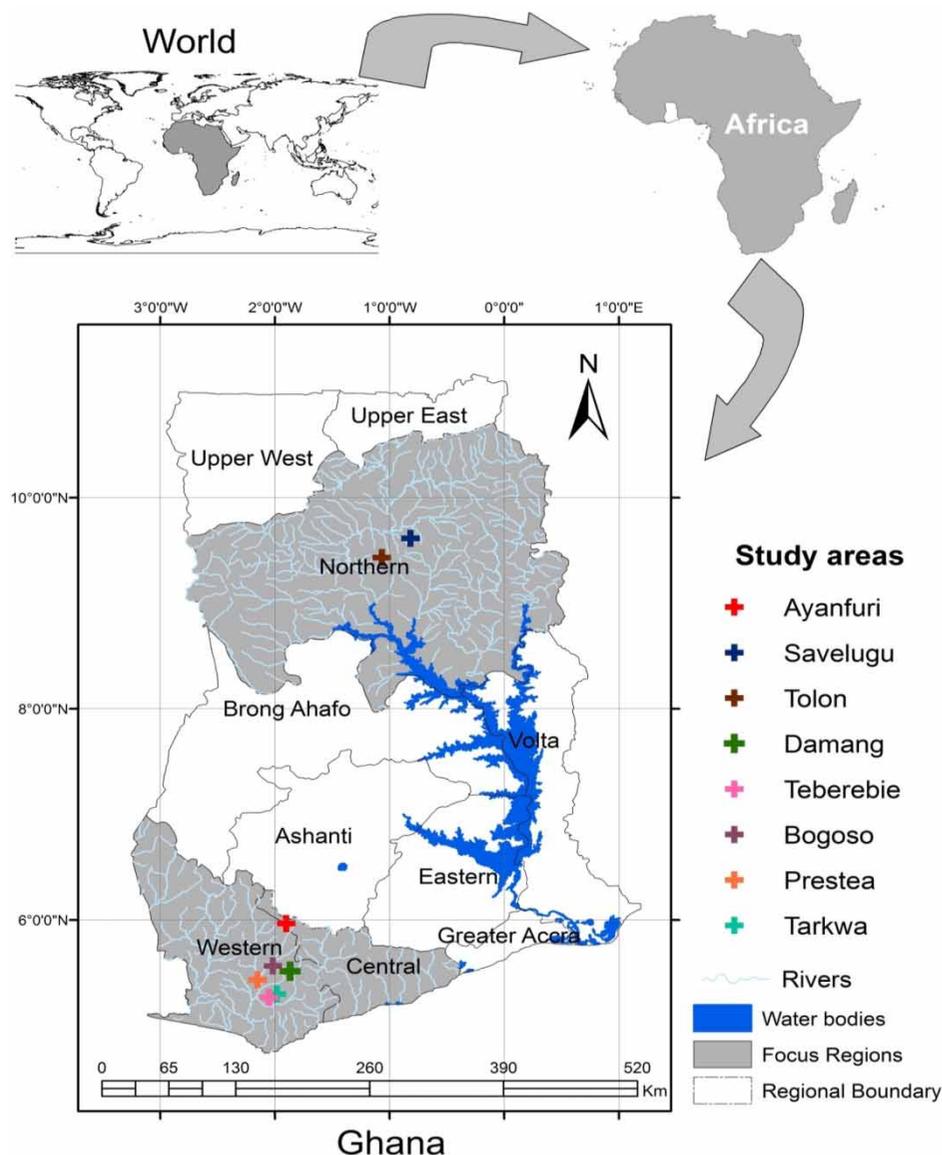
Figure 1 | Map of water quality sampling areas.

Non-sulphidic pale placer ore predominantly occurs in hard rock, and mostly is associated with series conglomerate within the Tarkwaian sedimentary formation (Kortatsi 2004). Oxidized ores, whose stem includes sulphides, arsenopyrite, realgar (AsS), and opiment (As_2S_3) pyrites, occur in weathered rocks (Kortatsi 2004). The northern zone, made up of Savelugu and Tolon, is characterized by conglomerates which range in age from Precambrian to Paleozoic and thus has two unambiguous geologic attributes: Precambrian basement rocks and Palaeozoic rocks

of the Voltaian sedimentary basin. The Voltaian Formation is an asymmetrical epicontinental interior sag basin comprising mostly shallow marine to continental sediments with sub-horizontal beds of sandstones, shale, mudstone, sand, and pebbles (Dapaah-Siakwan & Gyau-Boakye 2000). The sediments of this basin are for the most part protected by a dainty (<30 m) weathered layer that fundamentally influences the hydrogeologic attributes of the aquifer. The area within the Voltaian basin used for this study (i.e., Tolon and Savelugu zone), a transcendently

rural agricultural zone, is astoundingly level, shaping a delicately moving low-lying relief with topographic rises running from 120 m to 180 m above mean sea level.

Sampling and laboratory analysis of bacteriological and chemical parameters

The procedures of water sampling and laboratory analysis of groundwater quality adopted for this study has extensively been described by Armah (2014). Geo-satellite positioning of the exact locations of all the groundwater zones were resolved with a Garmin Etrex GPS (Garmin International, Inc., Olathe, Kansas, USA). Acceptable sampling protocol standards (APHA 1989, 2002) were followed during sampling of all the 738 samples that were collected. Before sampling, bottles were washed twice with detergent and rinsed with 10% hydrochloric acid and de-ionized (DI) water. Furthermore, at each of the sampling sites, the bottles were flushed with DI water to diminish or totally dispose of any pollution that might be present. At each location, the water was left running for approximately 3 minutes to cleanse the system before it was collected. One millilitre (1 mL) of concentrated nitric acid (HNO_3) was introduced to samples being prepped for metal analyses to acidify the same. The samples were promptly put into coolers containing ice (around 4 °C) and sent to the laboratory for analysis. This was necessary to prevent growth of microbes, flocculation and lessen any adsorption on container surfaces, processes which could confound the results of the analysis. Globally acknowledged and standard laboratory techniques were followed in the analysis of the samples (APHA 2002). At every sampling position, physicochemical water quality parameters, such as pH, electrical conductivity (EC), temperature, and turbidity were measured *in situ* utilizing the AQuanta multi-parameter water quality meter (Hydrolab Corporation, USA). Standard methods prescribed for the analysis of various elements and parameters (APHA 2002) were adopted for the laboratory analysis. Two 500 mL of water samples were collected at each location into two labelled bottles and were sent to two independent laboratories for analysis. To guarantee quality control and reproducibility of the results, each of the laboratories were sent a full complement of tests to examine. The samples were evaluated for nutrients (nitrates) and other water quality parameters, such as coliform microbes, pH, electrical conductivity, dissolved

solids, and turbidity. Flame atomic absorption spectrophotometry, as described by Armah *et al.* (2010a, 2010b) and Armah (2014), was employed to determine elemental concentrations of As, Cd, Fe, Mn, and Pb. In the present study, standard reference materials were utilized to check the precision of metal analysis in total concentration and the sequential extractions. Heavy metals recovery rates in the standard reference material were around 85–110%. Furthermore, the total concentrations of metals in sediments of successive extraction were equivalent to the independent aggregate concentrations, with recovery rates of 82–104%. Blanks for reagents were likewise utilized for error calculation and background corrections. No less than one duplicate was run for every six specimens to affirm the accuracy of the successive extraction technique. In all the analyses conducted, the precision and biasness were below 10%.

Data treatment and multivariate statistics

All statistical analyses were performed in STATA 13 (StataCorp, College Station, TX, USA). The standardized skewness and standardized kurtosis were determined to assess whether the samples originate from a normal distribution. Values of these statistics outside the range of −2 to +2 indicated significant departures from normality. The statistical analyses of data were carried out on all variables in the original or transformed dataset depending on the specific objectives. Multivariate analysis of the groundwater quality data was carried out using Spearman rho (original data), Pearson product moment correlation (normalized data), PCA/FA (normalized data), and discriminant analyis (DA) (original data) (see Simeonov *et al.* 2002, 2003; Shrestha & Kazama 2007; Armah *et al.* 2010b; Armah & Gyeabour 2013). In addition, negative log-log regression model was used to assess the pollution status of each groundwater location based on the magnitude of each of the water quality indicators.

Spearman rho and Pearson product moment correlation

Spearman's correlation coefficient (r_s) is a statistical measure of the strength of a *monotonic* relationship between paired data. This effect size measure was used to determine the nature of the relationships between the physicochemical parameters and trace metals in the original data without

imposing the assumption of normality (non-parametric). A value of zero implies no (monotonic) correlation, however it does not suggest there is no relationship between the variables. A perfect quadratic relationship may exist even though $r_s = 0$. Pearson's correlation coefficient is a statistical measure of the strength of a *linear* relationship between paired data. This effect size indicator was used for variables that were interval or ratio level, linearly related, and bivariate normally distributed. Variables that did not fully satisfy the assumption of normality were transformed prior to the determination of Pearson's correlation.

PCA/FA

Prior to the use of PCA/FA, correlation analysis was carried out. This was utilized to find an internal structure and assist in the identification of pollutant sources not accessible at first glance. Similarly, to determine the factorability of inter-correlation matrix or suitability of the data for PCA/FA, Bartlett's test of sphericity and Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy were performed on subsets and the entire variables. The KMO ranged between 0.60 and 0.76 indicating that the degree of common variance among the 12 variables is 'mediocre' bordering on 'middling' (see Varol *et al.* 2012). If a factor analysis is conducted, the factors extracted will account for a fair amount of variance but not a substantial amount. PCA and FA were applied to the altered data standardized through z-scale transformation to avoid misclassification due to wide differences in data dimensionality (see Shrestha & Kazama 2007; Armah *et al.* 2010a; Mustapha & Aris 2012). Furthermore, the standardization procedure eliminated the influence of different units of measurements and rendered the data dimensionless. PCA reduces the dimensionality of a dataset consisting of a large number of interrelated variables, while retaining as much of the variability present in a dataset as possible (Filik Iscen *et al.* 2008). This reduction is achieved by transforming the dataset into a new set of variables, the principal components (PCs), which are orthogonal (non-correlated) and arranged in decreasing order of importance.

DA

DA, a multivariate statistical technique, is commonly used to build a predictive or descriptive model of group discrimination based on observed predictor variables and to classify each observation into one of the groups as shown in Equation (1):

$$f(Gi) = Ki + \sum_{j=1}^{n}(Wij)(Pij) \qquad (1)$$

where i is the number of groups (G), k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, w_j is the weight coefficient, assigned by DA to a given selected parameters (p_j). In DA, multiple quantitative attributes are used to discriminate a single classification variable (Mustapha & Aris 2012). DA is different from the cluster analysis because prior knowledge of the classes, usually in the form of a sample from each class, is required. In this study, there were two classes: goldmining and non-goldmining locations. The common objectives of DA are: (i) to investigate differences between groups; (ii) to discriminate groups effectively; (iii) to identify important discriminating variables; (iv) to perform hypothesis testing on the differences between the expected groupings; and (v) to classify new observations into pre-existing groups. In order to determine the most parsimonious way to distinguish between goldmining and non-goldmining locations on the basis of the physicochemical properties of the study locations, DA was applied. This descriptive technique indicates which water quality parameters contribute most to group separation and successively identifies the linear combination of attributes known as canonical discriminant functions (equations) which contribute maximally to group separation (see Singh *et al.* 2005; Mustapha & Aris 2012; Devic *et al.* 2014). DA was performed based on the standard, forward stepwise and backward stepwise modes to evaluate the spatial variations in groundwater quality. The best discriminant functions for each mode were constructed considering the quality of the classification matrix and the number of parameters.

Wilcoxon rank-sum (Mann–Whitney) test

In this study, two seasons (rainy and dry) were used as proxies of temporal distribution of physicochemical parameters and trace metals. The main wet season is typically

from March to July whereas the dry season lasts from December to March (see McSweeney et al. 2010). Two-sample Wilcoxon rank-sum (Mann–Whitney) test was used to assess whether the levels of the physicochemical parameters and trace metals differ based on seasonality. It tests the null hypothesis that data in the wet season and dry season are samples from continuous distributions with equal medians, against the alternative that they are not. The test assumes that the two samples are independent.

Negative log-log regression

The outcome variable used for the analysis of pollution status was dichotomous. Sampling locations that complied with all the drinking water quality standards of the World Health Organization and the Ghana Water Company (see Armah et al. 2010a) were assigned a value of 0 indicating that they were not polluted, otherwise a value of 1 (polluted) was allocated. Based on this criterion, the sampling locations were unevenly or asymmetrically distributed. Approximately 27% of sampling locations were coded as 1 whereas 73% were coded as 0. A generalized linear model was then fitted to the binary response variable using the physicochemical parameters and trace metals as covariates. For the generalized linear model in which the dependent variable is binary, a probit or logit link function that assumes a symmetrical distribution could produce biased parameter estimates (see Armah 2014). Diagnosis of three candidate regression models (logistic, negative log-log, and complementary log-log) revealed that the negative log-log model fits much better than the other two models, as evaluated by Akaike information criterion (AIC) or Bayesian information criterion (BIC) statistics. Hence, the negative log-log model was the most parsimonious.

RESULTS AND DISCUSSION

Summary statistics of groundwater quality parameters

Table 1 shows the measures of central tendency (mean), measures of dispersion (minimum and maximum values, standard deviation) and measures of distribution (skewness and kurtosis) of the concentration of groundwater quality parameters. Skewness indicated that most of the variables exhibited asymmetry and deviation from a normal distribution. All the water quality parameters had skewness greater than zero (right skewed) indicating that most values are concentrated on the left of the mean, with extreme values to the right. Kurtosis signifies flattening or 'peakedness' of the distribution of each of the groundwater quality indicators. Kurtosis of pH is less than 3 and signifies a platykurtic distribution, flatter than a normal distribution with a wider peak. The probability for extreme values is less than for a normal distribution, and the values are wider spread around the mean. Kurtosis of all other variables in Table 1 is greater than 3 and suggests leptokurtic distributions characterized by sharper than a normal distribution, with values concentrated around the mean and thicker tails. This means high probability for extreme values. None of the variables had kurtosis of 3, implying that none of the distribution was mesokurtic.

Monotonic and linear relationships between groundwater quality parameters

The purpose of this section is to determine strong correlation between two parameters. This is useful in source apportionment and the identification of the origin of the bacteriological and physicochemical parameters. From Table 2, it can be

Table 1 | Descriptive statistics of the groundwater quality parameters

	pH	EC	TDS	Turbidity	Nitrates	As	Cd	Fe	Mn	Pb	Total coliform	E. coli
Mean	7.095	675.134	268.55	78.91	13.3399	0.238	0.0045	1.0332	0.7148	0.0242	63.142	21.1954
Std. dev.	1.195	1,353.768	397.71	253.699	104.718	3.627	0.012	2.864	1.981	0.054	334.96	199.945
Skewness	0.367	7.069	12.56	6.418	16.953	19.506	18.358	6.805	6.024	4.99	6.72	10.852
Kurtosis	0.507	59.986	228.91	51.698	312.116	387.01	416.527	61.497	52.645	30.342	44.263	121.394
Minimum	4.06	23.9	12.1	0	0	0	0	0.001	0	0	0	0
Maximum	11.33	14,920	8,206	2,940	2,120	78	0.276	37.628	26.8	0.497	2,419.6	2,419.6

Table 2 | Pearson's product moment correlation coefficients (parametric) and Spearman's rho (non-parametric)

	pH	Conductivity	TDS	Turbidity	Nitrate	As	Cd	Fe	Mn	Pb	Total coliform	E. coli
Pearson's product moment correlation coefficients (parametric)												
pH	1	0.624**	0.633**	−0.01	0.033	−0.052	0.056	0.015	−0.304**	0.038	0.209**	−0.138
Conductivity		1	0.882**	−0.220**	−0.105**	−0.179**	0.005	−0.118**	−0.136**	−0.072	0.008	−0.008
TDS			1	−0.208**	−0.053	−0.166**	−0.021	−0.109**	−0.130**	−0.064	0.063	0.116
Turbidity				1	0.431**	0.343**	−0.024	0.574**	0.383**	0.169**	0.236**	−0.176
Nitrate					1	0.338**	0.122**	0.207**	0.039	0.096**	0.105	−0.148
As						1	0.205**	0.128**	0.196**	0.256**	−0.088	0.185
Cd							1	0.003	−0.015	0.224**	0.118	−0.145
Fe								1	0.459**	0.127**	0.219**	−0.055
Mn									1	0.097**	−0.178*	−0.034
Pb										1	0.214**	0.462**
Total coliform											1	0.870**
E. coli												1
Spearman's rho (non-parametric)												
pH	1	0.732**	0.721**	−0.116**	−0.014	−0.132**	0.132**	−0.064	−0.346**	0.025	−0.105**	−0.051
Conductivity		1	0.900**	−0.227**	−0.116**	−0.204**	0.069	−0.128**	−0.152**	−0.04	−0.245**	−0.106**
TDS			1	−0.223**	−0.073*	−0.215**	0.051	−0.110**	−0.154**	−0.046	−0.229**	−0.083*
Turbidity				1	0.376**	0.292**	−0.097**	0.594**	0.376**	0.099**	0.473**	0.185**
Nitrate					1	0.266**	0.117**	0.183**	−0.01	0.099**	0.249**	0.083*
As						1	0.072	0.144**	0.180**	0.310**	0.393**	0.015
Cd							1	−0.105**	−0.191**	0.257**	−0.232**	−0.125**
Fe								1	0.485**	0.05	0.346**	0.128**
Mn									1	0.029	0.274**	0.062
Pb										1	0.120**	0.002
Total coliform											1	0.413**
E. coli												1

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).

deduced that when analysing Pearson's and Spearman's coefficients simultaneously, it is logical to expect that the significance of one would imply the significance of the other. However, a reverse implication does not necessarily seem to be logically true. It is also possible to encounter a situation where Pearson's coefficient is negative whereas Spearman's coefficient is positive, as evidenced in the relationship between Mn and total coliform bacteria.

In all, the set of data contains nine correlations above 0.50, including four generated by Spearman's method (Table 2). The five Pearson's correlations with values higher than 0.50 were also above 0.50 when the Spearman's method was used. The only inconsistency is the relationship between *Escherichia coli* and total coliform bacteria. Therefore, a relationship appears to exist between the following parameters: pH/electrical conductivity, pH/total dissolved solids (TDS), electrical conductivity/total dissolved solids, and Fe/turbidity, although to varying degrees of strength and statistical significance.

Fifty-one locations had pH values greater than 9.0. Twenty-three of the locations were in Tarkwa and 15 were found in Prestea. The remainder were distributed as follows: Bogoso (4), Savelugu (3), and Teberebie (6). Eight locations simultaneously recorded high pH values and high levels of turbidity. These were distributed as follows: Tarkwa (3), Teberebie (3), and Prestea (2). Turbidity increased with increasing iron concentration. This finding is consistent with the literature. It has been suggested that turbidity is attributable to the colloidal organic and inorganic matter from water, which does not settle in time (Lefebvre & Legube 1990; Likens 2009; Khraisheh *et al.* 2010). Furthermore, in the marginally alkaline pH conditions observed in natural groundwater, coagulation and sedimentation of iron compounds does not occur. They remain in the waterbody culminating in increased turbidity. Consequently, increased concentrations of iron in water at high pH contributes to increased water turbidity (Lefebvre & Legube 1990; Likens 2009; Khraisheh *et al.* 2010). Primary contributors to turbidity include clay, silt, finely divided organic and inorganic matter, soluble coloured organic compounds, plankton, and microscopic organisms (APHA 2002).

Regardless of the technique used, the strongest correlation was observed between total dissolved solids, a good indicator of the mineralized character of the water and conductivity suggesting a common origin. Conductivity of groundwater, defined as the ratio of the current density (electrolytic) to the electric field strength, is particularly influenced by the quality of the soil. The average conductivity of groundwater in areas that are covered with clay was higher ($1,400\ \mu S\ cm^{-1}$) than areas covered with gravel or sand ($360\ \mu S\ cm^{-1}$).

It is remarkable that most of the correlations between trace metal concentrations were statistically significant; however, none was higher than 0.5 notwithstanding the technique used. In both Pearson's and Spearman's correlations, the strongest linear and monotonic relationships occurred between Mn and Fe. According to Thurman (2012), in oxidized natural waters, the largest amount of iron is found as a complex compound bound to humus, a colloidal precipitate or bound to solid matter (such as micro-organisms). Only in conditions that are devoid of oxygen or in highly acidic waters is all iron found in an ionic form.

PCA/FA

PCA/FA was performed on the normalized data to compare the compositional pattern between the groundwater samples and to identify the factors influencing each one. PCA of the complete dataset (Table 3) exposed four PCs with eigenvalues >1 that explained approximately 69% of the total variance in the groundwater quality dataset. The scree plot of eigenvalues (not shown) confirms this cut-off point. The first PC, which accounted for 22.3% of the total variance was correlated (loading >0.75) with electrical conductivity, total dissolved solids, and pH. The second PC accounting for 20.8% of total variance was correlated with *E. coli* and total coliform bacteria. The third PC, which accounted for 13.4% of total variance, was correlated with Fe, whereas the fourth PC accounted for 12.3% of the total variance and was correlated with As.

Based on the factor loadings after varimax rotation, variance factor 1 reflects the natural variability of physicochemical properties in groundwater. In the groundwater dataset, the physicochemical properties varied markedly and reflect both natural and anthropogenic influences. Previous research has established that these physicochemical properties are influenced by solubility, adsorption, degradation, volatility, soil texture, organic matter content, soil permeability, depth to

Table 3 | PCs, total variance and rotated component matrix

Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of variance	Cumulative %
1	2.796	23.302	23.302	2.796	23.302	23.302	2.682	22.346	22.346
2	2.539	21.16	44.462	2.539	21.16	44.462	2.5	20.836	43.183
3	1.564	13.037	57.499	1.564	13.037	57.499	1.611	13.425	56.608
4	1.375	11.462	68.961	1.375	11.462	68.961	1.482	12.353	68.961
5	0.924	7.703	76.665						
6	0.821	6.839	83.504						
7	0.66	5.504	89.007						
8	0.533	4.438	93.446						
9	0.467	3.895	97.341						
10	0.231	1.922	99.263						
11	0.087	0.721	99.984						
12	0.002	0.016	100						

Varimax rotated component matrix[a]

	Component			
	1	2	3	4
pH	0.77	−0.013	0.147	0.074
Conductivity	0.958	0.019	−0.052	0.018
TDS	0.946	0.151	−0.065	0.02
Turbidity	0.39	−0.311	0.233	−0.213
Nitrate	−0.044	−0.367	−0.004	0.668
As	0.213	0.179	0.201	0.765
Cd	−0.029	−0.178	0.66	0.125
Fe	0.04	0.034	0.847	0.037
Mn	0.193	0.039	0.562	−0.539
Pb	0.18	0.687	−0.135	−0.265
Total coliform	−0.008	0.945	−0.011	−0.068
E. coli	−0.055	0.903	−0.022	0.132

Component transformation matrix

1	0.868	0.481	0.08	−0.091
2	0.444	−0.826	0.317	0.142
3	−0.183	0.088	0.761	−0.616
4	−0.127	0.28	0.56	0.769

[a]Rotation converged in five iterations.

Extraction method: principal component analysis. Rotation method: Varimax with Kaiser normalization.

groundwater, rainfall, and geologic conditions (Chapman 1996; Nosrati & Van Den Eeckhaut 2012; Barringer & Reilly 2013; Devic *et al.* 2014).

Based on the strong positive factor loadings on *E. coli* and total coliform, variance factor 2 reflects the microbiological aspect of groundwater. It is well documented that a fundamental microbiological problem in groundwater, especially where many dispersed, shallow dug wells or boreholes provide protected but untreated domestic water supplies, is the health hazard emanating from faecal matter contamination. It has been argued that of the four types of pathogens contained in human excreta, only bacteria and viruses are likely to be minute enough to be transported through the soil and aquifer matrix to groundwater bodies (see Chapman 1996; Graham & Polizzotto 2013). In this context, the presence or absence of oxygen is, consequently, one of the most significant attributes affecting microbial activity. Previous studies suggest that microbe populations are largest in the nutrient-rich humic upper parts of the soil, and decline with decreasing nutrient supply and oxygen availability at greater depths (see Chapman 1996; Barringer & Reilly 2013). In the presence of organic material, anaerobic microbial activity can occur far below the soil and has been observed at depths of hundreds and even thousands of metres. The depth to which such activity is likely is contingent on the nutrient supply, pH, salt content, groundwater temperature, and the permeability of the aquifer (Chapman 1996; Barringer & Reilly 2013; Devic *et al.* 2014).

Variance factor 3, which has strong factor loadings on Fe, epitomizes the natural geologic attributes of the groundwater. This finding is consistent with the previous work by Armah *et al.* (2010b). The distribution and origin of Fe in groundwater has been studied in various contexts across the world. Iron is abundant in most of the study areas in the form of pyrite and arsenopyrite ores (see Armah *et al.* 2011). According to Chapman (1996), high iron levels in groundwater are widely reported from developing countries, where they are often an important water quality issue. The magnitude and distribution of Fe in groundwater is influenced by a plethora of factors including the redox condition, pH values, the components of aquifer media, runoff conditions, and characteristics of overlying soils (see Chapman 1996; Appelo & Postma 2005).

Variance factor 4 had a strong positive factor loading on As and reflects the goldmining activities in some of the study locations. This is consistent with the findings of Armah *et al.* (2010b). It has been suggested that in instances where the spatial extent of As contamination of groundwater is wide, the sources of the As are likely to be geogenic (see Smedley & Kinniburgh 2002; Barringer & Reilly 2013). The speciation of arsenic is an important determining factor of its mobility, reactivity, and potential bioavailability in arsenic- and goldmine impacted regions. Arsenic speciation in these complex natural systems is, moreover, influenced by a number of physical, geological, and anthropogenic variables (see Bowell *et al.* 1994; Palumbo-Roe *et al.* 2007; Cancès *et al.* 2008). Previous research identified mining of arsenic and metal ores and natural geologic sources of arsenic as dominant environmental conditions that trigger inputs of As to groundwater (Nordstrom 2002; Eisler 2004; Barringer & Reilly 2013). Several biochemical mechanisms determine the mobility of As in groundwater. These include oxidation of pyrite, arsenopyrite, and arsenite (Morin & Calas 2006), reduction of sulphate and formation of sulphide (Nriagu *et al.* 2007), and microbially mediated precipitation of orpiment (Ehrlich & Newman 2009).

DA

A DA was conducted to predict whether a particular groundwater sample was obtained from a goldmining area or not. Predictor variables were conductivity, total dissolved solids, turbidity, nitrate, heavy metal concentrations (arsenic, cadmium, iron, manganese, lead), and pH. Significant mean differences were observed for all the predictors on the dependent variable. While the log determinants were quite similar, Box's M indicated that the assumption of equality of covariance matrices was violated. However, given the large sample, this problem is not regarded as serious. The discriminate function revealed a significant association between groups and all predictors, accounting for 41.09% of between group variability, although closer analysis of the structure matrix revealed two significant predictors, namely turbidity (0.650) and arsenic (0.560). The cross-validated classification showed that overall 82.0% were correctly classified.

We initially examined whether there were any significant differences between goldmining and non-goldmining locations based on the trace metal concentrations and physicochemical parameters using group means and analysis of variance (ANOVA) results data. The results of group statistics and

tests of equality of group means indicated significant group differences. For example, mean differences between pH scores and turbidity scores suggest that these may be good discriminators as the separations are large. Table 4 provides strong statistical evidence of significant differences between means of goldmining and non-goldmining locations for all trace metals and physicochemical parameters ($p < 0.0001$) with arsenic and turbidity producing very high value Fs. The pooled within-group matrices (Table 4) also support use of these independent variables as inter-correlations are low.

The pooled within-group matrices, Wilks' lambda, standardized canonical discriminant functions, and the coefficients of structure matrix are depicted in Table 4. The log determinants appear similar and Box's M was 782.135 with $F = 13.931$ which is significant at $p < 0.0001$ indicating that the null hypothesis that the groups do not differ cannot be retained. The eigenvalues provide information on each of the discriminate functions (equations) produced. The canonical correlation is the multiple correlations between the predictors and the discriminant function, which provides an index of overall model fit, interpreted as the proportion of variance explained (R^2). A canonical correlation of 0.641 was obtained suggesting that the model explains 41.09% of the variation in the grouping variable, i.e., whether a location is in a goldmining area or not.

Wilks' lambda indicates a highly significant function ($p < 0.0001$) and provides the proportion of total variability not explained, i.e., it is the converse of the squared canonical correlation; that is, 58.9% unexplained. The interpretation of the discriminant coefficients (or weights) is similar to multiple regressions. Table 4 provides an index of the importance of each predictor. The sign indicates the direction of the relationship. Arsenic (As) was the strongest predictor while cadmium (Cd) was next in importance as a predictor. These two variables with large coefficients stand out as those that strongly predict allocation to the goldmining or non-goldmining category. All the other factors were less successful as predictors. The structure matrix table (Table 4) shows the correlations of each variable with each discriminate function. These Pearson coefficients are structure coefficients or discriminant loadings. They serve as factor loadings in factor analysis. Based on the largest loadings for turbidity and arsenic, it can be deduced that 'physical properties and chalcophilic elements' was the function that most discriminated between non-goldmining and goldmining

locations. Generally, just like factor loadings, 0.30 is seen as the cut-off between important and less important variables.

From the coefficients of the canonical discriminant function in Table 4:

$$D = (-0.308*Cond) + (-0.842*TDS) + (0.406*Turbidity) \\ + (0.154*Nitrate) + (0.757*As) + (-1.526*Cd) \\ + (0.162*Fe) + (0.131*Mn) + (-0.035*Pb) \\ + (0.182*pH) - 1.064$$

The discriminant function coefficients b or standardized form *beta* both indicate the partial contribution of each variable to the discriminate function controlling for all other variables in the equation. This is useful to assess each variable's unique contribution to the discriminate function and therefore provide information on the relative importance of each variable. The classification results revealed that 83.6% of sampling locations were classified correctly into goldmining or non-goldmining groups. This overall predictive accuracy of the discriminant function, the 'hit ratio', indicated that non-goldmining locations were classified with slightly better accuracy (89%) than goldmining areas (69.6%). This is not entirely surprising given that the human activities (agriculture, ore mining, etc.) in goldmining areas are complex. It is problematic to isolate anthropogenic influences from natural background concentrations of trace metals in groundwater, and effectively characterize the processes controlling trace and toxic heavy metal mobilization and transport in goldmining localities.

Wilcoxon rank-sum (Mann–Whitney) test

The results of the Wilcoxon rank-sum (Mann–Whitney) test suggest that there is a statistically significant difference between the underlying distributions of Cd ($z = -6.358$, $p = 0.0001$), Mn ($z = 2.012$, $p = 0.0442$), and Pb ($z = -3.600$, $p = 0.0003$) in the wet seasons and dry seasons. In all three instances, the sum of the wet season scores was higher than the sum of the dry season scores. There were no observed seasonal variations in As and Fe concentrations in groundwater samples. However, *E. coli* and total coliforms were higher in the dry season than in the wet season. These results are counter-intuitive and inconsistent with the literature. Previous studies indicate that

Table 4 | Standardized canonical discriminant functions and coefficients of structure matrix

	Tests of equality of groups means		Correlation (pooled within-groups)										Standardized canonical discriminant function coefficients Function	Canonical discriminant function coefficients Function	Structure matrix Function
	Wilks' lambda	F	Conductivity	TDS	Turbidity	Nitrate	As	Cd	Fe	Mn	Pb	pH	1	1	1
Conductivity	0.909	68.532	1	0.865	−0.105	−0.027	−0.08	−0.05	−0.049	−0.075	−0.05	0.633	−0.14	−0.308	−0.38
Total dissolved solids	0.912	66.03		1	−0.093	0.028	−0.07	−0.073	−0.042	−0.067	−0.043	0.642	−0.344	−0.842	−0.373
Turbidity	0.772	200.794			1	0.352	0.193	0.07	0.509	0.302	0.162	0.037	0.356	0.406	0.65
Nitrate	0.923	57.203				1	0.26	0.191	0.139	−0.03	0.079	0.065	0.148	0.154	0.347
Arsenic	0.82	149.12					1	0.324	0.022	0.136	0.242	−0.042	0.587	0.757	0.56
Cadmium	0.957	30.316						1	0.073	0.04	0.243	0.003	−0.539	−1.526	−0.253
Iron	0.898	77.644							1	0.408	0.125	0.032	0.155	0.162	0.404
Manganese	0.935	47.273								1	0.086	−0.289	0.121	0.131	0.316
Lead	0.991	6.519									1	0.03	−0.021	−0.035	0.117
pH	0.988	8.089										1	0.214	0.182	−0.131

bacteriological concentrations in shallow groundwater are usually considerably higher during the wet season than during the dry season (see Alemayehu et al. 2006; Shrestha et al. 2014). In this context, Shrestha et al. (2014) attributed the seasonal variations in microbial quality to a high level of faecal material infiltration during the rainy season.

Negative log-log regression

A generalized linear (log-log regression) model was fitted to binary dependent variable (polluted or not polluted) depending on whether the location was compliant for the WHO drinking water quality standards or not. The physicochemical parameters and trace metals were introduced as covariates in the multivariate model.

All physicochemical parameters and trace metals (except Mn) were significant predictors of the pollution status of sampling locations. Monitoring sites with higher pH values were approximately 22% more likely to be polluted compared with locations with lower pH values (Table 5). As predictors of pollution status, monitoring sites with higher conductivity, turbidity, and total dissolved solids were no different from their counterparts where turbidity, conductivity, and total dissolved solids were low. Locations in which As concentrations were higher had significantly far higher odds of being polluted (OR = 27.68, CI: 5.08–150.59) compared with areas where As concentrations were lower. Monitoring sites with higher

Table 5 | Multivariate relationship between pollution status and selected covariates

Pollution status	Odds ratios	Robust std. err.	$P > z$	[95% confidence interval]
pH	1.215	0.0849	**0.005**	1.060–1.394
Conductivity	0.999	0.0004	**0.008**	0.998–1.000
Total dissolved solids	0.998	0.0007	**0.027**	0.997–1.000
Turbidity	1.002	0.0004	**0.000**	1.001–1.002
Nitrates	1.010	0.0038	**0.007**	1.003–1.018
Arsenic	27.680	23.9210	**0.000**	5.088–150.593
Cadmium	0.000	0.0000	**0.000**	0.000–0.000
Iron	1.068	0.0253	**0.005**	1.020–1.119
Manganese	1.036	0.0315	0.240	0.976–1.100
Lead	6.820	8.1872	0.110	0.649–71.711
Constant	0.323	0.1340	**0.006**	0.143–0.728

Statistically significant relationships are shown in bold font.

Fe concentration were approximately 7% more likely to be polluted compared with sampling sites with lower Fe concentration. Similarly, locations with higher Pb concentrations had higher odds of being polluted (OR = 6.82, CI: 0.64–71.71) compared with areas where Pb concentrations were lower, although this was not statistically significant. Monitoring sites where Cd concentrations were less likely to be polluted compared with areas where Cd levels were lower.

CONCLUSION

Chemometric approaches including PCA/FA and DA as well as generalized linear modelling (negative log-log regression) were used to assess the spatial and temporal variability of groundwater quality parameters in almost 740 locations in Ghana. This was complemented with parametric (Pearson's product moment correlation) and non-parametric measures of association (Spearman's correlation, Wilcoxon rank-sum test) to elicit a nuanced understanding of the complex relationship between physicochemical and bacteriological factors in groundwater. Although to varying degrees of strength and statistical significance, clear and persistent relationships occur between almost all the physicochemical parameters. However, the weak relationships between trace metals were rather weak. The strongest linear and monotonic relationships occurred between Mn and Fe. Four factors representing the natural variability of physicochemical properties, bacteriological quality, natural geologic attributes, and anthropogenic factors (artisanal goldmining) were extracted from the groundwater data. Cross-validated classification in DA showed that overall 82.0% of locations were correctly classified as either goldmining or non-goldmining based on the means of conductivity, total dissolved solids, turbidity, nitrate, trace metal concentrations (arsenic, cadmium, iron, manganese, and lead), and pH. Arsenic (As) was the strongest predictor followed by cadmium (Cd). These two variables had large coefficients and stood out as those that strongly predicted allocation to the goldmining or non-goldmining category. Although there were no observed seasonal variations in As and Fe concentrations in groundwater samples, E. coli and total coliforms were unexpectedly higher in the dry season than in the wet season. Except for Cd, groundwater monitoring sites with higher pH values, As, Fe, and Pb had higher

odds of being polluted. On the whole, goldmining locations were more likely to be polluted compared with non-gold-mining areas.

The foregoing suggests that interactions between surface water and groundwater resources are often intertwined such that active cross-sector dialogue and integrated vision are also needed to stimulate sustainable surface and groundwater management. To be effective policies on groundwater quality must be tailored to local hydrogeological settings, artisanal goldmining activities, and agro-economic realities. Furthermore, the implementation of such policies demands appropriate institutional arrangements with distinct emphasis and statutory power for groundwater management, full involvement of the goldmining community, and more mainstreaming of industrial development goals with groundwater quality.

REFERENCES

Alemayehu, T., Legesse, D., Ayenew, T., Mohammed, N. & Waltenigus, S. 2006 Degree of groundwater vulnerability to pollution in Addis Ababa, Ethiopia. In: *Groundwater Pollution in Africa* (Y. Xu & B. Usher, eds). Taylor & Francis, London, UK, pp. 203–211.

APHA 1989 *Standard Methods for the Examination of Water and Wastewater*, 17th edn. American Public Health Association, Washington DC, USA, p. 1268.

APHA 2002 *Standard Methods for the Examination of Water and Wastewater*, 20th edn. American Public Health Association, Washington, DC, USA.

Appelo, C. A. J. & Postma, D. 2005 *Geochemistry, Groundwater and Pollution*. A.A. Balkema Publishers, Leiden, The Netherlands.

Arias-Estévez, M., López-Periago, E., Martínez-Carballo, E., Simal-Gándara, J., Mejuto, J. C. & García-Río, L. 2008 The mobility and degradation of pesticides in soils and the pollution of groundwater resources. *Agriculture, Ecosystems & Environment* 123 (4), 247–260.

Armah, F. A. 2014 Relationship between coliform bacteria and water chemistry in groundwater within gold mining environments in Ghana. *Water Quality, Exposure and Health* 5 (4), 183–195.

Armah, F. A. & Gyeabour, E. K. 2013 Health risks to children and adults residing in riverine environments where surficial sediments contain metals generated by active gold mining in Ghana. *Toxicological Research* 29 (1), 69–79.

Armah, F. A., Obiri, S., Yawson, D. O., Onumah, E. E., Yengoh, G. T., Afrifa, E. K. & Odoi, J. O. 2010a Anthropogenic sources and environmentally relevant concentrations of heavy metals in surface water of a mining district in Ghana: a multivariate

statistical approach. *Journal of Environmental Science and Health Part A* 45 (13), 1804–1813.

Armah, F. A., Obiri, S., Yawson, D. O., Pappoe, A. N. M. & Akoto, B. 2010b Mining and heavy metal pollution: assessment of aquatic environments in Tarkwa (Ghana) using multivariate statistical analysis. *Journal of Environmental Statistics* 1 (4), 1–13.

Armah, F. A., Obiri, S., Yawson, D. O., Afrifa, E. K., Yengoh, G. T., Olsson, J. A. & Odoi, J. O. 2011 Assessment of legal framework for corporate environmental behaviour and perceptions of residents in mining communities in Ghana. *Journal of Environmental Planning and Management* 54 (2), 193–209.

Barringer, J. L. & Reilly, P. A. 2013 *Arsenic in Groundwater: A Summary of Sources and the Biogeochemical and Hydrogeologic Factors Affecting Arsenic Occurrence and Mobility*. INTECH Open Access Publisher, Rijeka, Croatia, pp. 83–116.

Bhattacharya, P., Sracek, O., Eldvall, B., Asklund, R., Barmen, G., Jacks, G., Koku, J., Gustafsson, J.-E., Singh, N. & Balfors, B. B. 2012 Hydrogeochemical study on the contamination of water resources in a part of Tarkwa mining area, Western Ghana. *Journal of African Earth Sciences* 66–67, 72–84.

Bowell, R. J., Morley, N. H. & Din, V. K. 1994 Arsenic speciation in soil porewaters from the Ashanti Mine, Ghana. *Applied Geochemistry* 9 (1), 15–22.

Cancès, B., Juillot, F., Morin, G., Laperche, V., Polya, D., Vaughan, D. J., Hazemann, J. L., Proux, O., Brown Jr., G. E. & Calas, G. 2008 Changes in arsenic speciation through a contaminated soil profile: a XAS based study. *Science of the Total Environment* 397 (1), 178–189.

Chapman, D. 1996 *Water Quality Assessments: A Guide to the Use of Biota, Sediments, and Water in Environmental Monitoring*. UNESCO, WHO and UNEP, Geneva.

Dapaah-Siakwan, S. & Gyau-Boakye, P. 2000 Hydrogeologic framework and borehole yields in Ghana. *Hydrogeology Journal* 8 (4), 405–416.

Devic, G., Djordjevic, D. & Sakan, S. 2014 Natural and anthropogenic factors affecting the groundwater quality in Serbia. *Science of the Total Environment* 468, 933–942.

Edmunds, W. M., Shand, P., Hart, P. & Ward, R. S. 2003 The natural (baseline) quality of groundwater: a UK pilot study. *Science of the Total Environment* 310 (1), 25–35.

Eisler, R. 2004 Arsenic hazards to humans, plants, and animals from gold mining. In: *Reviews of Environmental Contamination and Toxicology* (G. Ware, ed.). Springer, New York, USA, pp. 133–165.

Ehrlich, L. & Newman, D. K. 2009 *Geomicrobiology* 5th edn. CRC Press, Taylor & Francis Group, New York.

Filik Iscen, C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V. & Ahiska, S. 2008 Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environmental Monitoring and Assessment* 144 (1), 269–276.

Foster, S. S. D. & Chilton, P. J. 2003 Groundwater: the processes and global significance of aquifer degradation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358 (1440), 1957–1972.

Gazzaz, N. M., Yusoff, M. K., Ramli, M. F., Aris, A. Z. & Juahir, H. 2012 Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Marine Pollution Bulletin* **64**, 688–698.

Giridharan, L., Venugopal, T. & Jayaprakash, M. 2009 Assessment of water quality using chemometric tools: a case study of River Cooum, South India. *Archives of Environmental Contamination and Toxicology* **56**, 654–669.

Graham, J. P. & Polizzotto, M. L. 2013 Pit latrines and their impacts on groundwater quality: a systematic review. *Environmental Health Perspectives* **121**, 521–530.

Jordana, S. & Piera, E. B. 2004 Natural groundwater quality and health. *Geologica Acta* **2** (2), 175.

Khraisheh, M., Al-Ghouti, M. A. & Stanford, C. A. 2010 The application of iron coated activated alumina, ferric oxihydroxide and granular activated carbon in removing humic substances from water and wastewater: column studies. *Chemical Engineering Journal* **161** (1), 114–121.

Kortatsi, B. K. 2004 Hydrochemistry of Groundwater in the Mining Area of Tarkwa–Prestea, Ghana. PhD thesis, University of Ghana, Legon-Accra, Ghana.

Kumari, M., Tripathi, S., Pathak, V. & Tripathi, B. D. 2013 Chemometric characterization of river water quality. *Environmental Monitoring and Assessment* **185**, 3081–3092.

Lefebvre, E. & Legube, B. 1990 Iron (III) coagulation of humic substances extracted from surface waters: effect of pH and humic substances concentration. *Water Research* **24** (5), 591–606.

Likens, G. E. 2009 Biogeochemistry of inland waters. In: *A Derivative of the Encyclopedia of Inland Waters* (G. E. Likens, ed.). Elsevier/Academic Press, Oxford, UK. Volume 1.

McSweeney, C., Lizcano, G., New, M. & Lu, X. 2010 The UNDP climate change country profiles: improving the accessibility of observed and projected climate information for studies of climate change in developing countries. *Bulletin of the American Meteorological Society* **91** (2), 157–166.

Morin, G. & Calas, G. 2006 Arsenic in soils, mine tailings, and former industrial sites. *Elements* **2** (2), 97–101.

Morris, B. L., Lawrence, A. R., Chilton, P. J. C., Adams, B., Calow, R. C. & Klinck, B. A. 2003 Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management. http://www.unep.org/DEWA/ water/groundwater/pdfs/Groundwater_INC_cover.pdf (retrieved 10 January 2016).

Mustapha, A. & Aris, A. Z. 2012 Spatial aspects of surface water quality in the Jakara Basin, Nigeria using chemometric analysis. *Journal of Environmental Science and Health, Part A* **47** (10), 1455–1465.

Nordstrom, D. K. 2002 Worldwide occurrences of arsenic in ground water. *Science* **296** (5576), 2143–2145.

Nosrati, K. & Van Den Eeckhaut, M. 2012 Assessment of groundwater quality using multivariate statistical techniques in Hashtgerd Plain, Iran. *Environmental Earth Sciences* **65** (1), 331–344.

Nriagu, J. O., Bhattacharya, P., Mukherjee, A. B., Bundschuh, J., Zevenhoven, R. & Loeppert, R. H. 2007 Arsenic in soil and groundwater: an overview. *Trace Metals and other Contaminants in the Environment* **9**, 3–60.

Palumbo-Roe, B., Klinck, B. & Cave, M. 2007 Arsenic speciation and mobility in mine wastes from a copper-arsenic mine in Devon, UK: a SEM, XAS, sequential chemical extraction study. In: *Trace Metals and Other Contaminants in the Environment*, Volume 9 (P. Bhattacharya, A. B. Mukherjee, J. Bundschuh, R. Zevenhoven & R. H. Loeppert, eds). Elsevier BV, pp. 441–471.

Pinder, G. F. & Celia, M. A. 2006 *Subsurface Hydrology*. John Wiley & Sons, Hoboken, NJ, USA.

Purtschert, R. 2009 Groundwater timescales and tracers. In: *Natural Groundwater Quality* (W. M. Edmunds & P. Shand, eds). John Wiley & Sons, Chichester, UK, pp. 91–108.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environmental Modelling & Software* **22** (4), 464–475.

Shrestha, S., Nakamura, T., Malla, R. & Nishida, K. 2014 Seasonal variation in the microbial quality of shallow groundwater in the Kathmandu Valley, Nepal. *Water Science & Technology: Water Supply* **14** (3), 390–397.

Simeonov, V., Einax, J. W., Stanimirova, I. & Kraft, J. 2002 Environmetric modeling and interpretation of river water monitoring data. *Analytical and Bioanalytical Chemistry* **374**, 898.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, Th. 2003 Assessment of the surface water quality in Northern Greece. *Water Research* **37**, 4119–4124.

Singh, K. P., Malik, A., Singh, V. K., Mohan, D. & Sinha, S. 2005 Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. *Analytica Chimica Acta* **550** (1), 82–91.

Smedley, P. L. & Kinniburgh, D. G. 2002 A review of the source, behaviour and distribution of arsenic in natural waters. *Applied Geochemistry* **17** (5), 517–568.

Thurman, E. M. 2012 *Organic Geochemistry of Natural Waters*, (Vol. 2). Springer Science & Business Media, Dordrecht, The Netherlands.

Varol, M., Gökot, B., Bekleyen, A. & Şen, B. 2012 Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. *Catena* **92**, 11–21.

Wu, M.-L., Wang, Y.-S., Wang, H., Sun, C.-C., Dong, J.-D., Yin, J.-P. & Han, S.-H. 2010 Identification of coastal water quality by statistical analysis methods in Daya Bay, South China Sea. *Marine Pollution Bulletin* **60**, 852–860.

Zhou, F., Liu, Y. & Guo, H. 2007 Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong. *Environmental Monitoring and Assessment* **132**, 1–13.