

## Relationship of Trace Metal Covariates and pH Distribution in Groundwater within Gold mining and Non-Gold mining Areas in Ghana

Frederick Ato Armah<sup>1\*</sup>, Arnold Paintsil<sup>2</sup>, Michael Osei Adu<sup>3</sup>, David Oscar Yawson<sup>4</sup>, Justice Odoi<sup>5</sup>

<sup>1</sup>Frederick Ato Armah  
Department of Environmental Science  
School of Biological Sciences  
College of Agricultural & Natural Sciences  
University of Cape Coast, Ghana  
[farmah@ucc.edu.gh](mailto:farmah@ucc.edu.gh)

<sup>2</sup>Arnold Paintsil  
Department of Civil and Environmental Engineering  
Faculty of Engineering  
Spencer Engineering Building,  
Western University  
London, Ontario, Canada, N6A 5B9  
[apaintsi@uwo.ca](mailto:apaintsi@uwo.ca)

<sup>3</sup>Michael Osei Adu  
Department of Crop Science  
School of Agriculture  
College of Agricultural & Natural Sciences  
University of Cape Coast  
Cape Coast, Ghana  
[michael.adu@ucc.edu.gh](mailto:michael.adu@ucc.edu.gh)

<sup>4</sup>David Oscar Yawson  
CERMES  
University of West Indies  
Cavehill Campus  
Barbados  
[david.yawson@cavehill.uwi.edu](mailto:david.yawson@cavehill.uwi.edu)

<sup>5</sup>Justice Odoiquaye Odoi  
Environmental Specialist  
P.O. Box OS 1455,  
Osu, Accra, Ghana  
[odoi18676@alumni.itc.nl](mailto:odoi18676@alumni.itc.nl)

Author to whom correspondence should be addressed:

## Abstract

One of the most important defining characteristics of groundwater quality is pH as it fundamentally controls the amount and chemical form of many organic and inorganic solutes in groundwater. Groundwater data are frequently characterized by a wide degree of variability of the factors which possibly influence pH distribution. For this reason, it is challenging to link the spatio-temporal dynamics of pH to a single environmental factor by the ordinary least squares regression technique of the conditional mean. In this study, quantile regression was used to estimate the response of pH to nine environmental factors (As, Cd, Fe, Mn, Pb, turbidity, electrical conductivity, total dissolved solids and nitrates). Results of 25%, 50%, 75% quantile regression and ordinary least squares (OLS) regression were compared. The standard regression of the conditional means (OLS) underestimated the rates of change of pH due to the selected factors in comparison with the regression quantiles. The effect of arsenic increased for sampling locations with higher pH values (higher quantiles) likewise the influence of Pb and Mn. However, the effects of Cd and Fe decreased for sampling locations in higher quantiles. It can be concluded that these detected heterogeneities would be missed if this study had focused exclusively on the conditional means of the pH values. Consequently, quantile regression provides a more comprehensive account of possible spatio-temporal relationships between environmental covariates in groundwater. This study is one of the first to apply this technique on groundwater systems in sub-Saharan Africa. The approach is useful and interesting and has broad application for other mining environments especially tropical low-income countries where climatic conditions can drive rapid cycling or transformations of pollutants. It is also pertinent to geopolitical contexts where regulatory; monitoring and management capacities are weak and where mining pollution of groundwater largely occur.

Key words: quantile regression; groundwater; environmental; multivariate; metals; health

## 1. Introduction

Contamination of and pressures on surface water in Africa has directed attention to groundwater as an alternative source of safe drinking water (Adelana and MacDonald, 2008; Armah et al., 2010a, 2010b; MacDonald et al., 2012). It is estimated that approximately half of the nearly one billion people in Africa rely on groundwater for their daily water supply (Taylor et al., 2008; 2009). However, groundwater might also not be entirely safe as it can be contaminated with heavy metals, and bacteria may be introduced by faulty septic installations or contaminated systems. Owing to this, it is pertinent that groundwater be monitored recurrently, which is costly and requires technical abilities that may be systematically lacking in rural areas in Africa (Armah, 2014; Awuah, et al., 2009). Hence, groundwater data is frequently incomplete and relationships between health and contaminants are complex and characterized by uncertainty (Emmanuel et al., 2009; Jarsjö et al., 2005; Kapaj et al., 2006; Siirila et al., 2012). This suggests the need for tools and methods that gives rapid and effective assessment of the extent of groundwater contamination and its cascading repercussions on human health and the environment.

Statistical techniques have been used extensively in the past to study seasonal and temporal variability of foregoing factors and also to elicit a deeper understanding of the relationships between bacteriological and chemical contaminants in groundwater (Armah et al., 2017). In this context, research on the relationships between covariates (e.g., trace metal concentrations, coliform counts) and response variables (e.g., pH levels, pollution status) in groundwater have almost exclusively used least squares modeling, which focuses on modeling the conditional mean of the response variable without taking into consideration the full conditional distribution of the response variable (Gardner and Vogel, 2005; Goovaerts et al., 2005; Kaown et al., 2007).

The statistical distributions of groundwater data frequently exhibit unequal variation due to complex interactions between the factors affecting physicochemical and bacteriological constituents that cannot all be measured and accounted for in statistical models (Cressie, 2015; Gibbons et al., 2009; Wendland et al., 2005). Uneven variation indicates multiple slopes (rate of change) describing the relationship between a response and predictor variables measured on a subset of these factors. While models of the conditional mean (linear regression) can address the question “are trace metal and bacteriological contaminants important determinants of pH and vice versa?”, it cannot answer an important question: “does trace metal and bacteriological contaminants influence pH differently for groundwater monitoring locations with low pH than for those with average or higher pH values?” This question warrants the need for alternative statistical techniques that can account for multiple rates of change, and brings into sharp focus the usefulness of quantile regression. According to Cade and Noon (2003), quantile regression approximates several orders of change (slopes) from the minimum to maximum response, giving a more comprehensive picture of the relationships between variables missed by other regression methods.

In this study, quantile regression was applied to explore the effect of trace metal contaminants on the rate of change of pH in groundwater. It was hypothesized that slope estimates obtained using ordinary least squares (OLS) methods would not fully describe the relationship between trace metal contaminants and pH values in groundwater at all points of the distribution of pH scores. This approach is necessary to guide groundwater monitoring initiatives and target specific

groundwater quality indicators for policy attention. This study is one of the first to apply this technique to groundwater systems in sub-Saharan Africa. This study is pertinent and interesting and has extensive application to other mining environments in especially tropical, low-income countries where climatic conditions can drive rapid cycling or transformations of pollutants. It is also useful in settings where regulatory; monitoring and management capacities are weak and where mining pollution of groundwater largely occur.

## 2. Materials and Methods

### 2.1 Study area

The locations from which (underground water that supplies wells) samples were collected (Figure 1) and a complete and thorough written work of the study area are well recorded/supported by Armah (2014) and Armah et al. (2017). The distribution of the locations in the study areas is as follows: Ayanfuri (16) Savelugu (173), Tolon (47), Damang (6), Teberebie (85), Bogoso (40), Prestea (166) and Tarkwa (205). A complete and thorough representation of the study area is given by Armah (2014).

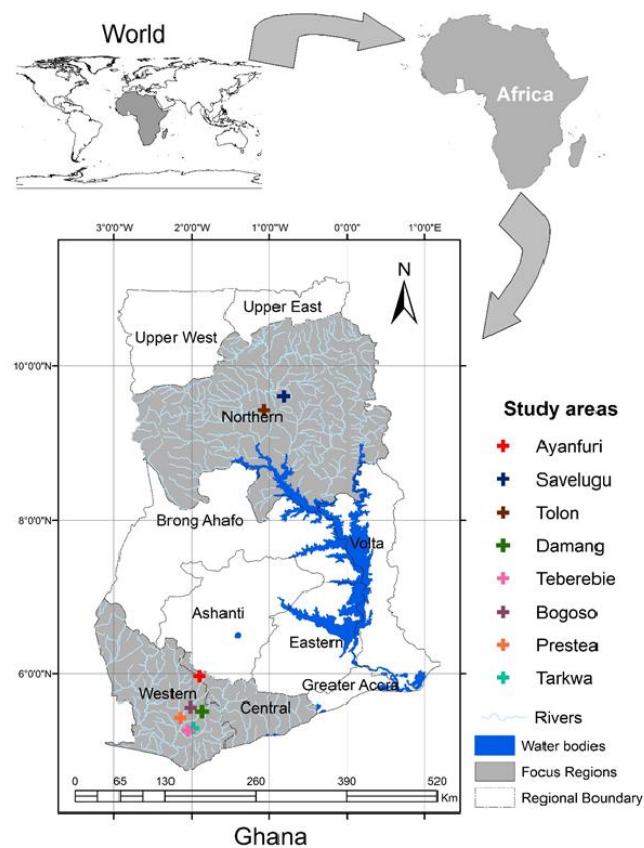


Figure 1: Study area

The study of rocks of southwestern Ghana (Ayanfuri, Bogoso, Damang, Tarkwa, Prestea, Teberebie) shows greenstone belts composed of mafic volcano-related rocks and in between basins drained by rivers usually made up of fine-grained, deep marine sediments changed at greenschist facies. The study area reveals three main gold deposits. Placer or (coming from floodwaters) deposit, non-sulphidic pale placer or free milling ore and oxidized ore (Kortatsi, 2004; Bhattacharya et al., 2012). The alluvial deposits originate from streams draining areas with auriferous deposits where the underground rock-related material is only (a) little changed and intruded by Dixcove (very hard rock) especially in Birimian rock areas (Kortatsi, 2004). Non-sulphidic pale placer ore emanates mainly from hard rock. It is especially connected with Banket conglomerates of Tarkwa formation (Kortatsi, 2004). Oxidized ore manifest in weathered rocks and is taken from sulphides, arsenopyrite, realgar (AsS), and opiment ( $As_2S_3$ ) pyrites (Kortatsi, 2004).

The northern segment (Savelugu and Tolon) show two distinct areas based on the history and behaviour of the physical Earth in terms of features that is the Precambrian basement rocks and Palaeozoic rocks from the Voltaian material that sinks in liquid and is drained by a river. These rocks are usually covered by a thin weathered layer less than thirty metres that significantly affect the hydrogeologic features of the underground water-storage related system. The Voltaian (creation and construction/ group of objects) consists mainly of sandstones, shale, mudstone, sandy and pebbly beds and limestones (Dapaah-Siakwan & Gyau-Boakye, 2000). The Tolon and Savelugu area, a mostly-farming based area, is (in a surprising and interesting way) flat, forming a gently rolling low-lying relief with topographic elevations ranging from 120 to 180m above mean sea level.

## 2.2 Water Sampling, Handling and Laboratory Analysis

A thorough record of groundwater quality parameters in these study areas was originally reported by Armah (2014) and Armah et al. (2017). Geo-satellite situating of all the groundwater areas were resolved with a Garmin Etrex GPS. Seven hundred and thirty-eight (738) samples were assembled and analysed. Testing took after conventional benchmarks (APHA, 1989 and 2002). Inspecting bottles were washed many times with cleanser and flushed with 10% hydrochloric acid and twofold distilled water preceding examining. At each of the testing areas, the bottles were washed with distilled water to decrease or totally dispose of any impurities that may be presented. At every area the water was permitted to keep running for quite a while to cleanse the framework before being examined. Tests for metal examination were acidified with 1 ml of concentrated nitric acid. The gathered samples were promptly put into coolers containing ice blocks (around 4°C) and passed on to the research facility for examination. This method turns away microbial development, flocculation and diminish any adsorption on compartment surfaces, interferences which could influence the outcomes. Globally acknowledged and standard research facility systems were followed in the examination (APHA, 2002). At each examining area, physicochemical water quality parameters (pH, electrical conductivity, temperature, and turbidity) were measured *in situ* utilizing the AQUANTA multi-parameter water quality meter.

The lab investigation was contingent on standard strategies for examination recommended for the different components and parameters (APHA, 2002). Water samples were gathered from each area in two plainly marked plastic containers (500ml each). The samples were sent to two autonomous research facilities for examination. Every research center had a total complement of tests to investigate. This was done to guarantee quality control and reproducibility of the outcomes. Nitrates and other water quality parameters including coliform microscopic organisms, pH, electrical conductivity, dissolved solids, and turbidity were determined. As, Cd, Fe, Mn, and Pb levels were determined utilizing flame Atomic Absorption Spectrophotometry (AAS) (Armah et al., 2010a; Armah, 2014). In this examination, standard reference materials were utilized to confirm the exactness of metal investigation in complete focus and the consecutive extraction strategy. The recovery rates of substantial metals in the standard reference material were around 85–110%. Likewise, the aggregate groupings of metals in consecutive extractions were tantamount to the free aggregate focuses, with recovery rates of 82–104%. Reagent blanks were likewise utilized for foundation adjustment and arbitrary mistake figuring. No less than one copy was kept running for each six samples to affirm the accuracy of the consecutive extraction technique. The accuracy and predisposition to bias were under 10% in all investigations.

### 2.3 Data Analysis

This study utilized full sample data, including 518 groundwater monitoring wells in gold-mining areas and 220 sites in non-gold-mining locations. Information obtained from the examination of Groundwater and Human Health in Ghana was used to illustrate the application of quantile regression. An exhaustive record of the physicochemical and trace metal quality parameters in the dataset can be found in Armah (2014). The present paper studies the relationships between pH and conductivity, dissolved solids, turbidity, nitrates, As, Cd, Fe, Mn and Pb. Based on the assertion that the concentration of trace metals in groundwater may be predominantly controlled by pH values and vice versa (Lei et al., 2010), pH was set as a response variable in the present study and the other physicochemical factors and trace metals as covariates to assess whether their effects on pH were uniform across the pH distribution.

The pooled descriptive statistics of the groundwater data indicated that trace metals and physicochemical parameters exhibit a wide degree of variability in terms of the measures of central tendency, dispersion and distribution (Table 1).

Table 1: Summary statistics of physicochemical parameters and trace metals in groundwater (n=738)

Variable (units)	Mean	Std. Dev.	Min	Max
pH	7.095	1.195	4.060	11.330
Conductivity ( $\mu\text{Scm}^{-1}$ )	675.134	1353.768	23.9	14920
Total Dissolved Solids (ppm)	268.546	397.709	12.1	8206
Turbidity (NTU)	78.910	253.699	0.000	2940

Nitrates (ppm)	13.340	104.718	0.000	2120
Arsenic (ppm)	0.238	3.627	0.000	78
Cadmium (ppm)	0.004	0.012	0.000	0.276
Iron (ppm)	1.033	2.864	0.001	37.628
Manganese (ppm)	0.715	1.981	0.000	26.8
Lead (ppm)	0.024	0.054	0.000	0.497

#### 2.4 Statistical techniques

Quantile regression (QR) was the key statistical procedure for the present study. Quantile is an identical term to percentile, where the 50th quantile is the median value. In like manner, the 25th and 75th quantiles are identical to the first and third quartiles. Quantiles are examples of fundamental non-parametric statistics (specifically, order-statistics) and are moderately impervious to outliers (Petscher and Logan, 2014). Given normally distributed errors, results of linear regression and QR will lie at the median. In instances where errors were not normally distributed, or in non-homoscedasticity situations, QR gives a more effective and precise gauge of parameters (Chen and Chalhoub-Deville, 2014; Petscher and Logan, 2014). In this study, heteroscedasticity tests were conducted before statistical analysis in order to validate the utilization of QR. The Breusch-Pagan and Cook-Weisberg test statistic was found to be significantly different than zero, suggesting that there was heteroscedasticity in the groundwater parameters and hence the utilization of QR was justifiable. There were no data transformation in any parameter before the analysis so as to: (i) preserve the natural variability of the physicochemical factors and trace metals in the dataset and (ii) highlight the complex relationship that varies rather than stays constant along the pH distribution.

The least squares mean regression is more common in literature for examining the relationship between and among variables and to assess how independent variables influence alterations in a dependent variable (Chen and Chalhoub-Deville, 2014; Petscher and Logan, 2014). By conducting linear square mean regression, it is expected that the means fall on a line (Koenker, 2005). Schröder et al. (2005) however suggested that it is quite problematic to relate field data to explanatory variables, because that in principle, leads to the analysis of only a small subset of all possible influencing factors. Assumptions of normality and homoscedasticity that are fundamental for least square regression, in majority of cases, cannot be satisfied for groundwater field data. Reason being that, there are uneven variations and consequently more than a single slope describing the relationship between a response variable and predictor variables measured on a subset of these factors (Cade and Noon, 2003; Hao and Naiman, 2007). For such dataset, the median therefore becomes a more suitable measure of central tendency than the mean because the relationships between variables change at different points in the distribution. The advantage of QR is that it does not only provide estimates of the relationships between the predictor(s) and outcome, but does so across multiple points of the outcome's distribution (Petscher and Logan, 2014). Quantile regression provides an alternative statistical approach in which the effects of independent covariates on the outcome can be indicated at any location along the distribution of the outcome (Magzamen et al., 2015). More so, quantiles are relatively resistant to outliers and

offer a more efficient and accurate estimate of parameters than OLS (Chen and Chalhoub-Deville, 2014).

The relationship between pH and physicochemical factors and trace metals in groundwater was calculated using Eq.(1):

$$pH_i = g(P_i, T_i) = \alpha_i + \beta_i P_i + \gamma_i T_i + \varepsilon_i \quad (1)$$

Where  $pH_i$  is the pH of groundwater at location  $i$ ;  $T_i$  is a vector of independent variables capturing trace metal characteristics that can affect pH distribution;  $P_i$  is the average of each of the physicochemical factors at location  $i$ ;  $\beta_i$  and  $\gamma_i$  are a set of parameters to be estimated; and  $\varepsilon_i$  is the correspondent error term. Also,  $\alpha_i$  is the level of pH when all the predictors have a value of zero.

The QR model is:

$$\ln pH_i = \beta_\theta T_i + \mu_\alpha \text{ with } Quant_\theta(\ln pH_i) = \beta_\theta T_i \quad (2)$$

where  $T_i$  denotes the vector of exogenous variables and  $\beta_\theta$  is the vector of estimated parameters at quantile  $\theta$ .  $Quant_\theta(\ln pH_i)$  denotes the  $\theta$ th conditional quantile of the  $\ln$  pH given  $T$ . The  $\theta$ th regression quantile,  $0 < \theta < 1$ , is defined as a solution to the problem:

$$\min_{\beta \in R^k} \left\{ \sum_{i: \ln pH_i \geq \beta T_i} \theta |\ln pH_i - \beta_\theta T_i| + \sum_{i: \ln pH_i < \beta T_i} (1 - \theta) |\ln pH_i - \beta_\theta T_i| \right\} \quad (3)$$

Conventionally, the QR model is written as:

$$\min \sum_i \rho_\theta(\ln pH_i - \beta_\theta T_i),$$

where  $\rho_\theta(\varepsilon)$  is the check function defined as  $\rho_\theta(\varepsilon) = \theta\varepsilon$  if  $\varepsilon \geq 0$  or  $\rho_\theta(\varepsilon) = (\theta-1)\varepsilon$  if  $\varepsilon < 0$ .

In addition, Pearson's product moment correlation coefficients and Spearman's rank order correlation were calculated to ascertain the degree to which the variance in a parameter was determined by the variability in another parameter.



## 2.5 Data analysis

Data were processed in IBM SPSS version 20 whereas quantile and OLS regressions were done in STATA version 13 (StataCorp, TX, 2013). QREG add-on in STATA was used to determine the regression coefficients, standard errors and confidence intervals. Statistical significance was set to  $\alpha \leq 0.05$ .

## 3. Results and Discussion

In this study, the effects of other physicochemical factors and trace metals covariates on pH throughout the pH distribution were investigated. Overall, the results reveal that the quantile regression was significantly more sensitive to sub-groups than the ordinary least squares regression. The analysis of trace metal concentrations and physicochemical parameters indicated that there were marked variations in the magnitudes of the changes of the physicochemical parameters and trace metals across quantiles (Table 2).

Table 2: Descriptive statistics by quantiles

25th Quantile (n=185)				
Variable	Mean	Std. Dev.	Min	Max
pH	5.673	0.554	4.060	6.310
Conductivity	183.957	123.799	30.000	685.000
Total Dissolved Solids	95.816	64.877	16.900	324.000
Turbidity	27.880	56.827	0.000	435.000
Nitrates	3.885	12.368	0.000	140.000
Arsenic	0.045	0.094	0.000	0.800
Cadmium	0.004	0.005	0.000	0.027
Iron	1.007	2.270	0.001	19.000
Manganese	1.228	2.366	0.000	12.280
Lead	0.024	0.045	0.000	0.497
50th Quantile (n=185)				
pH	6.655	0.206	6.320	7.000
Conductivity	291.538	202.476	37.900	1453.000
Total Dissolved Solids	155.685	105.575	21.500	650.000
Turbidity	79.828	240.459	0.000	2940.000
Nitrates	10.938	37.434	0.000	250.000
Arsenic	0.054	0.133	0.000	0.900
Cadmium	0.004	0.007	0.000	0.070
Iron	1.313	3.519	0.001	37.628
Manganese	1.240	2.931	0.000	26.800
Lead	0.026	0.063	0.000	0.462

	75th Quantile (n=184)			
pH	7.419	0.254	7.010	7.830
Conductivity	913.787	1946.909	23.900	14920.000
Total Dissolved Solids	371.088	686.991	12.100	8206.000
Turbidity	51.465	137.972	0.000	986.000
Nitrates	4.628	13.074	0.000	132.000
Arsenic	0.032	0.119	0.000	1.100
Cadmium	0.004	0.007	0.000	0.080
Iron	0.667	1.439	0.001	10.100
Manganese	0.183	0.389	0.000	2.170
Lead	0.014	0.025	0.001	0.211
	99th Quantile (n=184)			
pH	8.643	0.764	7.840	11.330
Conductivity	1316.010	1634.276	152.700	14180.000
Total Dissolved Solids	453.146	248.478	70.800	1865.000
Turbidity	156.737	411.474	0.000	2540.000
Nitrates	33.973	204.521	0.001	2120.000
Arsenic	0.823	7.244	0.001	78.000
Cadmium	0.005	0.021	0.000	0.276
Iron	1.145	3.619	0.001	27.600
Manganese	0.203	0.543	0.000	6.300
Lead	0.032	0.070	0.001	0.416

For example, large changes were observed in conductivity values across the various quantiles - i.e.: 25th to 50th (58%), 50th to 75th (213%) and 75th to 99th quantile (44%) (Table 2). It was also observed that increases in As from the 25th to 50th quantile (20%) and from the 50th to 75th quantile (40%) were relatively smaller increments compared to the increase recorded for the 75th to 99th quantile (2500%). There were increases in concentrations and levels of the physicochemical parameters from the 25th to 50th quantile in all trace metals, except Cd (Table 2). The transition from the 75th to 99th quantile is unique given that this is where the concentration of Cd did not only change but also increased by 25%. The percentage change ranged from <1% (for Mn) to 180% (for turbidity and nitrates). However, from the 50th to 75th quantile, about six groundwater quality indicators including nitrates, turbidity, As, Fe, Mn and Pb reduced in magnitude (Table 2).

The results in this study suggest that the levels and the rates of increase or decrease of these covariates in groundwater are not uniform across their respective distributions. This finding is consistent with previous studies using QR in ecological settings (Cade and Noon, 2003; Cade, 2011). It also demonstrates the usefulness of QR in the analysis of physicochemical and bacteriological quality of groundwater. Yet, we did not find a single previous study on groundwater quality in the sub-Saharan African context that used quantile regression.

It was also seen in Table 2 that, pH increased unevenly across quantiles. Percentages increased from the 25<sup>th</sup> to 50<sup>th</sup>, 50<sup>th</sup> to 75<sup>th</sup> and from 75<sup>th</sup> to 99<sup>th</sup> were 17%, 11% and 16%, respectively. Similar observation is seen in Figure 2, where it is evident that the rate of increase of pH across the full range of the distribution is uneven, with sharp increases at the tails and gradual increases between the 20<sup>th</sup> and 80<sup>th</sup> quantiles.

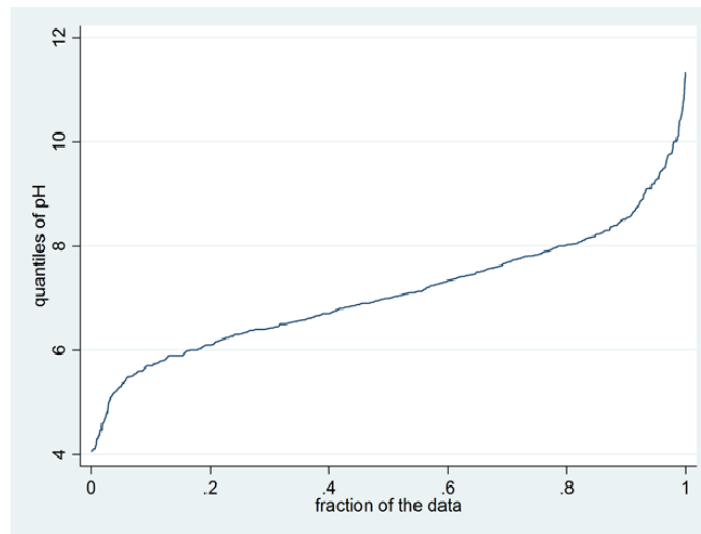


Figure 2: Distribution of pH by quantiles

This reflects extreme pH values for the present groundwater data and suggests that modeling the conditional mean value of pH may not be supported by this set of data. The mobility of arsenic in groundwater is contingent on its valence state and chemical form, which in turn, is controlled by pH. Between pH of 4 and 10, As(V) species are negatively charged in water, and the predominant As(III) species is neutral in charge (Panagiotaras et al., 2012). According to Bissen & Frimmel (2003), inorganic arsenic species arsenite (As(III)) and arsenate (As(V)) typically predominate in the water environment. The mobility of these forms is determined by the occurrence of adsorbing species such as hydroxides and oxides of Al(III), Mn(III/IV), Fe(III), humic substances, and clay minerals as well as pH value, and the redox potential (Bissen & Frimmel, 2003; Panagiotaras et al., 2012; Smedley et al., 1996).

Some scholars argue that low pH can restrict chemical reduction/oxidation processes in groundwater that supplies wells (Adriano, 2001; Adriano et al., 2004). Extreme pH levels on the other hand, can limit microbial activity and interfere with or slow down the use of both onsite and above ground applications of treatment of biological systems (Brown & Calas, 2011; Smith & Huych, 1999). The effects of pH on groundwater that supplies wells, its surrounding conditions and the soil matrix are well researched. For example, pH affect steadiness/balance

between the solid and water-based phases of many metals and other chemicals in the underground water that supplies wells and its surrounding conditions (Deutsch & Siegel, 1997).

It also controls the electric charge on the surface of very small patterned particles, especially clays (Merkel et al., 2005), such that at low pH, particles tend to have positively charged surfaces, whereas at high pH, negatively charged surfaces develop (Alloway, 2013). Soils generally have pH values within the range of 4-8.5 (Alloway, 2013). High pH conditions tend to limit the concentrations and ability to move around of metals in underground water that supplies wells by improving the tendency for metal ions (cations) to attach to clay particles (Edmunds & Smedley, 1996).

In groundwater, bioavailability of trace metals is a complicated function of numerous factors (chemical, biological, and environmental processes) including total concentration and speciation (physical-chemical forms) of metals, mineralogy, pH, redox potential, temperature, total organic content (both particulate and dissolved fractions), and suspended particulate content, along with volume of water, water velocity, and duration of water availability, particularly in arid and semi-arid environments (Adriano, 2001; Adriano et al., 2004; Bradl, 2005; Brown and Calas, 2011; Smith and Huyck, 1999). Several factors vary temporally, and most factors are interrelated. The outputs of the conditional mean and quartiles are shown in Table 3.

Table 3: Quantile and ordinary least squares regression coefficients

pH levels	Ordinary Least Squares				
	Coef.	Std. Err.	P-Value	[95% Conf.	Interval]
Conductivity	0.000111	3.45E-05	0.001	0.000043	0.000179
Total Dissolved Solids	0.000752	0.000118	0.000	0.0005197	0.000983
Turbidity	0.000985	0.000182	0.000	0.0006275	0.001343
Nitrates	0.001058	0.001055	0.316	-0.0010127	0.003129
Arsenic	-0.02666	0.029731	0.370	-0.0850316	0.031705
Cadmium	4.007802	3.423801	0.242	-2.7139	10.7295
Iron	0.013319	0.014465	0.357	-0.0150792	0.041718
Manganese	-0.1273	0.019616	0.000	-0.1658068	-0.08878
Lead	1.360053	0.752709	0.071	-0.1176868	2.837793
Intercept	6.759263	0.054988	0.000	6.651308	6.867218
25th Quantile					
Conductivity	0.000111	3.45E-05	0.001	0.0000	0.0002
Total Dissolved Solids	0.000752	0.000118	0.000	0.0005	0.0010
Turbidity	0.000985	0.000182	0.000	0.0006	0.0013
Nitrates	0.001058	0.001055	0.316	-0.0010	0.0031
Arsenic	-0.02666	0.029731	0.370	-0.0850	0.0317
Cadmium	4.007802	3.423801	0.242	-2.7139	10.7295
Iron	0.013319	0.014465	0.357	-0.0151	0.0417
Manganese	-0.1273	0.019616	0.000	-0.1658	-0.0888

Lead	1.360053	0.752709	0.071	-0.1177	2.8378
Intercept	6.759263	0.054988	0.000	6.6513	6.8672
50th Quantile					
Conductivity	6.05E-05	0.000033	0.067	-4.19E-06	0.000125
Total Dissolved Solids	0.002914	0.000113	0.000	0.0026931	0.003135
Turbidity	0.000878	0.000174	0.000	0.0005371	0.00122
Nitrates	-0.00057	0.001007	0.570	-0.0025486	0.001403
Arsenic	0.006026	0.028369	0.832	-0.0496693	0.061721
Cadmium	3.759343	3.267007	0.250	-2.654536	10.17322
Iron	-0.00666	0.013803	0.630	-0.0337595	0.020436
Manganese	-0.07328	0.018718	0.000	-0.1100259	-0.03653
Lead	1.054366	0.718239	0.143	-0.3557009	2.464432
Intercept	6.281422	0.05247	0.000	6.178411	6.384433
75th Quantile					
Conductivity	0.000159	4.06E-05	0.000	0.0000795	0.000239
Total Dissolved Solids	0.003058	0.000139	0.000	0.0027853	0.00333
Turbidity	0.000548	0.000214	0.011	0.0001278	0.000968
Nitrates	-0.00043	0.001239	0.730	-0.0028601	0.002006
Arsenic	0.001745	0.034931	0.960	-0.0668325	0.070322
Cadmium	2.076471	4.022652	0.606	-5.820912	9.973853
Iron	-0.0079	0.016995	0.642	-0.0412632	0.025468
Manganese	-0.09201	0.023048	0.000	-0.1372588	-0.04676
Lead	2.79373	0.884364	0.002	1.057522	4.529939
Intercept	6.690933	0.064606	0.000	6.564096	6.81777

The results presented in Table 3 indicate an increasing effect of certain trace metals while decreasing of others by increasing pH. For example, the independent effects of As, Pb, and Mn increase for sampling locations with higher pH values (higher quantiles). Contrary, the effects of Cd and Fe decreases for sampling locations in higher quantiles. This observation implies that the detected heterogeneities and nuances would be missed if this study had focused exclusively on the conditional means of the pH values.

The parameters estimated from the OLS regression and the 25<sup>th</sup> quantile regressions were almost identical. Nitrates, cadmium, iron, arsenic, and lead were not significant predictors of the pH distribution for the OLS regression and the 25<sup>th</sup> quantile regression unlike conductivity, TDS, turbidity and manganese. The lack of statistical significance for the lowest quantile and OLS pair could probably be interpreted as the inability to differentiate among pH levels for geographical locations with low physicochemical and trace metal concentrations. Only TDS, turbidity and manganese were significant predictors of the pH distribution for the 50<sup>th</sup> quantile regression unlike the 75<sup>th</sup> quantile where conductivity, TDS, turbidity, manganese and lead were significant predictors. The statistically significant pairwise comparisons observed with the other quantile pairs indicate that the relationship is not uniform at the conditional pH percentiles investigated. It also suggests that the models are capturing variability that would otherwise have been missed if only the OLS technique had been used in this study.

While turbidity, total dissolved solids and Mn were statistically significant across all quantiles, Pb was significant only at the 75<sup>th</sup> percentile whereas conductivity was significant at the 25<sup>th</sup> and 50<sup>th</sup> quantile. This suggest that the relationship between these covariates and pH values is different between monitoring sites whose conditional pH score is at the 25<sup>th</sup> percentile versus those at the 50<sup>th</sup> percentile. Furthermore, there is a significant difference between all upper pairs except turbidity, As, Cd, Fe and nitrates. This significance for the pair-wise comparisons of 25<sup>th</sup> quantile and above provides statistical confirmation of the different relationship between pH and the covariates at various points of the pH distribution. This statistical significance is a robust evidence that analyses of the relationship between pH and physicochemical and trace metal concentrations should be assessed further than the mean of the distribution.

Results in Table 3, also point to two types of significant coefficients: those that are significantly different from zero, and the quantile coefficients that are significantly different from the OLS coefficients (outside of the OLS confidence interval). For example, the coefficient on Cd at the 75% quantile is significantly different from zero but not significantly different from the OLS coefficient. This may also be visualized in Figure 3, where the black solid line with gray areas represents the quantile regression slope estimates at each quantile and the confidence interval for the estimates respectively. The horizontal solid dark line with dotted lines above and below represents the mean OLS regression estimates and confidence band. The horizontal line at zero is the reference line for hypothesis testing against a slope value of 0.

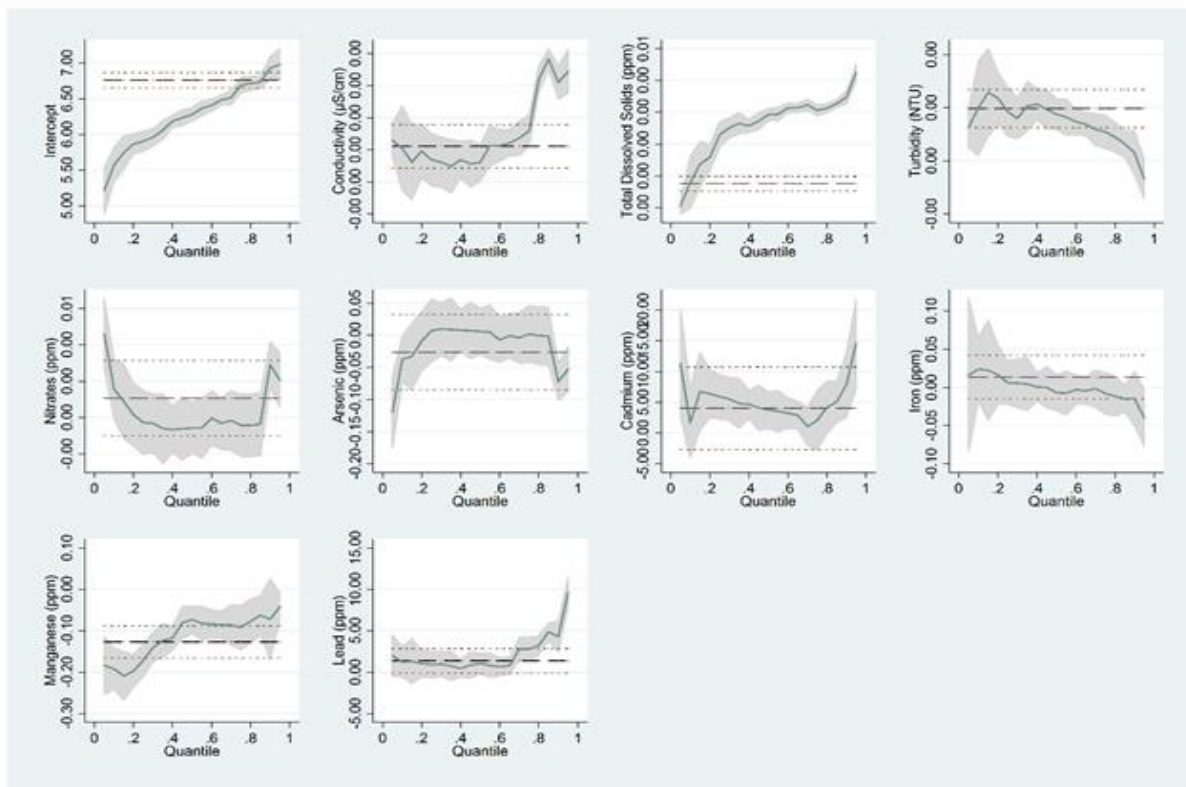


Figure 3: Quantile regression coefficient plots

Comparing the plots of physicochemical and trace metal concentrations in Figure 3, it is evident that the confidence band for the slope coefficients (i.e., the width of the gray area) of conductivity, nitrates, turbidity, Fe, As, Cd, Mn and Pb all reveal less precision of estimation compared to the estimation of total dissolved solids and the intercept. However, some of the coefficients of the covariates (e.g., conductivity, turbidity, Pb) are still statistically significant, especially at higher quantiles, since the confidence band does not cross the 0-reference line (hypothesis of the slope coefficient being 0). Three main discernible patterns emerge in Figure 3. The first pattern is the clear positive relationship between the distribution of pH and quantiles of conductivity, TDS, lead, manganese and lead (higher concentrations of these five variables are associated with higher quantiles of pH). The second pattern is the clear negative relationship between the distribution of pH and quantiles of turbidity and iron (higher concentrations of these two variables are associated with lower quantiles of pH). The third pattern, which is more stable, occurs between the distributions of pH and nitrates, arsenic and cadmium (for higher quantiles of these three variables there is no predictable change in the pH distribution).

Table 4: Pearson product moment and Spearman rank order correlations showing interdependency of water quality parameters

	pH	EC	TDS	Turbidity	Nitrates	As	Cd	Fe	Mn	Lead
<b>Pearson product moment correlation coefficients</b>										
<b>pH</b>	1									
<b>Conductivity</b>	.281**	1								
<b>TDS</b>	.355**	<b>.564**</b>	1							
<b>Turbidity</b>	.274**	.026	.108**	1						
<b>Nitrates</b>	.130**	.026	.064	.420**	1					
<b>Arsenic</b>	.104**	.026	.055	.372**	<b>.934**</b>	1				
<b>Cadmium</b>	.056	-.016	-.015	.042	.029	.029	1			
<b>Iron</b>	.078*	-.038	-.002	.270**	-.004	-.015	.187**	1		
<b>Manganese</b>	-.230**	-.070	-.058	.001	-.021	-.015	.008	.110**	1	
<b>Lead</b>	.108**	-.059	-.030	.249**	.001	-.008	.153**	.161**	.013	1
<b>Spearman rank order correlation coefficients</b>										
<b>pH</b>	1									
<b>Conductivity</b>	<b>.732**</b>	1								
<b>TDS</b>	<b>.721**</b>	<b>.900**</b>	1							
<b>Turbidity</b>	-.116**	-.227**	-.223**	1						
<b>Nitrates</b>	-.014	-.116**	-.073*	.376**	1					
<b>Arsenic</b>	-.132**	-.204**	-.215**	.292**	.266**	1				
<b>Cadmium</b>	.132**	.069	.051	-.097**	.117**	.072	1			
<b>Iron</b>	-.064	-.128**	-.110**	<b>.594**</b>	.183**	.144**	-	1		
									1	
<b>Manganese</b>	-.346**	-.152**	-.154**	.376**	-.010	.180**	-	.105**	.485**	1
<b>Lead</b>	.025	-.040	-.046	.099**	.099**	.310**	.257**	.050	.029	1

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

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

Table 4 shows the interdependencies among the water quality parameters when the data are assumed to be distribution dependent or parameteric (Pearson correlation) and when the data are not considered to be distribution dependent or non-parameteric (Spearman correlation). Although most of the correlations are statistically significant at  $\alpha$ -level of 95% or 99%, only two pairs of parameteric (Pearson's  $r > 0.5$ ) and four pairs of non-parameteric correlations (Spearman's  $\rho > 0.5$ ) were strong. Regarding Pearson's product moment correlation coefficients, the relationship between turbidity and conductivity was strong and statistically significant ( $r = 0.564$ ,  $p < 0.01$ ) likewise the relationship between arsenic and nitrates which was very strong and the highest ( $r = 0.934$ ,  $p < 0.01$ ). This suggests that based on the coefficient of determination ( $r^2$ ), 31.8% of the variance in turbidity is accounted for by the variation in conductivity. Also, 87.2% of the variance in arsenic is accounted for by the variation in nitrates. The Spearman rank order correlations reveal a different picture. The relationship between pH and conductivity was strong and statistically significant ( $r_s = 0.732$ ,  $p < 0.01$ ) likewise the relationship between pH and total dissolved solids ( $r_s = 0.721$ ,  $p < 0.01$ ). Also, the correlation of TDS and conductivity was very strong and statistically significant ( $r_s = 0.900$ ,  $p < 0.01$ ). Approximately 81% of the variance in conductivity is accounted for by the variation in TDS. The correlation of Fe and turbidity was strong and statistically significant ( $r_s = 0.594$ ,  $p < 0.01$ ). Generally, stronger and statistically significant correlations among any pair of variables indicate that both variables likely emanate from a common origin (Armah et al., 2010a).

In this paper, we illustrate that by slicing the water quality samples into discrete and more homogeneous slices created by quantiles, we can present a more precise account of the nature and scope of the water quality data. By moving away from the conventional approach of using the mean in order to describe a group, we unearth a broad range of opportunities, not only for descriptive or analytic statistics, but also for the practical study of water quality assessment and monitoring. The capability to examine each quantile will make possible the identification of water quality locations that require immediate and thorough attention or remediation. On the other hand, it will also help us discern what factors exist across the water quality location sub-groups and which factor is capable of enhancing all of the sub-groups. One limitation of this study is the fact that it was cross-sectional, thus we could identify association but not causality. Future longitudinal studies, particularly those following an intervention or occurrence that extensively influences the variability of water quality, will shed more light on the causes of change identified in the different quantiles.

#### 4. Conclusion

This study set out to assess the effect of environmental factors on both the conditional mean (OLS regression) and the full conditional distribution of pH (quantile regression). The non-uniform variation of the pH scores in groundwater indicates that there is more than a single slope (rate of change) describing the relationship between pH and As, Cd, Fe, Mn, Pb, turbidity,



electrical conductivity, total dissolved solids and nitrates. At lower pH, there were no statistically significant differences between the ordinary least squares and the regression of the lowest quantile (25<sup>th</sup> percentile). This suggests that in this groundwater dataset it is challenging to differentiate among pH levels for geographical locations with low physicochemical and trace metal concentrations. However, quantile regressions at the 50<sup>th</sup> and 75<sup>th</sup> percentile indicated otherwise. This study underscores the fact that quantile regression has a number of advantages over ordinary least squares regression in the analysis of groundwater quality. In this context, the quantile coefficient plots established that the connection between most of the environmental covariates and pH scores in the groundwater data was not the same magnitude across all points of the pH distribution. Overall, this study also demonstrates that it is possible to detect whether the relationship between any of the covariates and pH levels at a specific quantile of pH remains static over multiple assessment points. The quantile regression results suggest that the relationship between lead and pH is more complex than is indicated by the OLS approach. The capacity of the quantile regression technique to isolate heterogeneities across the full conditional distribution of the response variable make it a powerful tool that is particularly suitable for understanding the complex dynamics among the environmental factors in groundwater data. It was concluded that the use of quantile regression analysis in studies of water quality can engender the creation of customized response plans, adapted to the needs of sub-groups (such as extremely polluted) locations and help enhance the robustness of remediation actions. Future research can integrate both spatial statistics and quantile regression in order to elicit a more nuanced understanding of water quality in groundwater systems.

## References

- Adelana, S.M., Macdonald, A.M., 2008. Groundwater research issues in Africa. *Applied Groundwater Studies in Africa, IAHS Selected Papers on Hydrogeology*, 13, pp.1-7.
- Adriano, D.C., 2001. Arsenic. In Adriano, D.C. (Eds.), *Trace elements in terrestrial environments*. Springer, New York, pp. 219-261.
- Adriano, D.C., Wenzel, W.W., Vangronsveld, J., Bolan, N.S., 2004. Role of assisted natural remediation in environmental cleanup. *Geoderma*, 122(2), 121-142.
- Alloway, B.J., 2013. Sources of heavy metals and metalloids in soils. In Alloway, B.J., (Eds.), *Heavy metals in soils*. Springer, Netherlands, pp. 11-50.
- American Public Health Association-APHA 1989. *Standard Methods for the Examination of Water and Wastewater*, 17th edition. American Public Health Association, Washington D.C.
- American Public Health Association-APHA 2002. *Standard Methods for the Examination of Water and Wastewater*, 20th ed. American Public and Health Association, Washington D.C.
- 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., 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., Paintsil, A., Yawson, D.O., Adu, M.O., Odoi, J.O., 2017. Modelling spatio-temporal heterogeneities in groundwater quality in Ghana: a multivariate chemometric approach. *Journal of Water and Health*, 15 (4), 658-672.

Awuah, E., Nyarko, K.B., Owusu, P.A., Osei-Bonsu, K., 2009. Small town water quality. *Desalination*, 248, 453-459.

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, 72-84.

Bissen, M., & Frimmel, F. H. 2003. Arsenic—a review. Part I: occurrence, toxicity, speciation, mobility. *Acta hydrochimica et hydrobiologica*, 31(1), 9-18.

Bradl, H. (Ed.), 2005. Heavy metals in the environment: origin, interaction and remediation: origin, interaction and remediation. *Interface Science and Technology*, Vol. 6. Academic Press, New York.

Brown, G. E., Calas, G., 2011. Environmental mineralogy—understanding element behavior in ecosystems. *Comptes Rendus Geoscience*, 343(2), 90-112.

Cade, B. S., Noon, B. R., 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8), 412-420.

Cade, B. S., 2011. Estimating equivalence with quantile regression. *Ecological Applications*, 21(1), 281-289.

Chen, F., Chalhoub-Deville, M., 2014. Principles of quantile regression and an application. *Lang. Test.*, 31, 63–87

Cressie, N., 2015. *Statistics for spatial data*. John Wiley & Sons, New Jersey.

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

- Deutsch, W. J., Siegel, R., 1997. Groundwater geochemistry: fundamentals and applications to contamination. CRC Press, Boca Raton.
- Edmunds, W. M., Smedley, P. L., 1996. Groundwater geochemistry and health: an overview. Geological Society, London, Special Publications, 113(1), 91-105.
- Emmanuel, E., Pierre, M. G., Perrodin, Y., 2009. Groundwater contamination by microbiological and chemical substances released from hospital wastewater: Health risk assessment for drinking water consumers. *Environment international*, 35(4), 718-726.
- Gardner, K. K., & Vogel, R. M., 2005. Predicting ground water nitrate concentration from land use. *Groundwater*, 43(3), 343-352.
- Gibbons, R. D., Bhaumik, D. K., Aryal, S., 2009. Statistical methods for groundwater monitoring. *Statistics in Practice* Vol. 59, John Wiley & Sons, New Jersey.
- Goovaerts, P., Avruskin, G., Meliker, J., Slotnick, M., Jacquez, G., Nriagu, J., 2005. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. *Water Resources Research*, 41(7), W07013.
- Hao L. and Naiman D. Q., 2007. *Quantile Regression*, Sage Publications, Thousand Oaks.
- Jarsjö, J., Bayer-Raich, M., Ptak, T., 2005. Monitoring groundwater contamination and delineating source zones at industrial sites: Uncertainty analyses using integral pumping tests. *Journal of contaminant hydrology*, 79(3), 107-134.
- Kaown, D., Hyun, Y., Bae, G. O., Lee, K. K., 2007. Factors affecting the spatial pattern of nitrate contamination in shallow groundwater. *Journal of environmental quality*, 36(5), 1479-1487.
- Kapaj, S., Peterson, H., Liber, K., Bhattacharya, P., 2006. Human health effects from chronic arsenic poisoning—a review. *Journal of Environmental Science and Health Part A*, 41(10), 2399-2428.
- Koenker, R., 2005. *Quantile regression*. Cambridge University Press, Cambridge, New York.
- Kortatsi, B.K., 2004. *Hydrochemistry of Groundwater in the Mining Area of Tarkwa–Prestea, Ghana*, PhD thesis. University of Ghana, Legon-Accra, Ghana.
- Lei, L. Q., Song, C. A., Xie, X. L., Li, Y. H., Fei, W. A. N. G., 2010. Acid mine drainage and heavy metal contamination in groundwater of metal sulfide mine at arid territory (BS mine, Western Australia). *Transactions of Nonferrous Metals Society of China*, 20(8), 1488-1493.
- MacDonald, A. M., Bonsor, H. C., Dochartaigh, B. É. Ó., Taylor, R. G., 2012. Quantitative maps of groundwater resources in Africa. *Environmental Research Letters*, 7(2), 024009.

- Magzamen, S., Amato, M.S., Imm, P., Havlena, J.A., Coons, M.J., Anderson, H.A., Kanarek, M.S., Moore, C.F., 2015. Quantile regression in environmental health: Early life lead exposure and end-of-grade exams. *Environmental Research*, 137, 108–119.
- Merkel, B. J., Planer-Friedrich, B., Nordstrom, D., 2005. *Groundwater geochemistry. A Practical Guide to Modeling of Natural and Contaminated Aquatic Systems*, 2. doi.org/10.1007/b138774.
- Panagiotaras, D., Panagopoulos, G., Papoulis, D., & Avramidis, P. 2012. Arsenic Geochemistry in Groundwater System, *Geochemistry - Earth's System Processes*, Dionisios Panagiotaras (Ed.). InTech, Available from: <http://www.intechopen.com/books/geochemistryearth-s-system-processes/arsenic-geochemistry-in-groundwater-system>
- Petscher, Y., Logan, J.A.R., 2014. Quantile Regression in the Study of Developmental Sciences. *Child Dev.*, 85, 861–881.
- Schröder, H.K., Andersen, H.E., Kiehl, K., 2005. Rejecting the mean: Estimating the response of fen plant species to environmental factors by non-linear quantile regression. *J. Veg. Sci.*, 16, 373–382.
- Siirila, E. R., Navarre-Sitchler, A. K., Maxwell, R. M., McCray, J. E., 2012. A quantitative methodology to assess the risks to human health from CO<sub>2</sub> leakage into groundwater. *Advances in Water Resources*, 36, 146-164.
- Smedley, P.L., Edmunds, W.M., Pelig-Ba, K.B. 1996. Mobility of arsenic in groundwater in the Obuasi gold-mining area Ghana: some implications for human health. In: Appleton, J.D., Fuge, R., McCall, G.J.H. (Eds.), *Environmental Geochemistry and Health*, vol. 113. Geological Society Special Publication, London, pp. 163–181.
- Smith, K. S., & Huyck, H. L., 1999. An overview of the abundance, relative mobility, bioavailability, and human toxicity of metals. *The environmental geochemistry of mineral deposits*, 6, 29-70.
- Taylor, R., Tindimugaya, C., Aureli, A., 2008. Can Groundwater Meet Africa's Demand for Water as Climates Change? *Groundwater and Climate in Africa: An International Conference; Kampala, Uganda, 24–28 June 2008. EOS, Transactions American Geophysical Union*, 89(42), 407-407.
- Taylor, R. G., Koussis, A. D., Tindimugaya, C., 2009. Groundwater and climate in Africa—a review. *Hydrological Sciences Journal*, 54(4), 655-664.
- Wendland, F., Hannappel, S., Kunkel, R., Schenk, R., Voigt, H. J., Wolter, R., 2005. A procedure to define natural groundwater conditions of groundwater bodies in Germany. *Water Science & Technology*, 51(3-4), 249-257.