Clare Robinson - 10785991 - EART60372 MSc Project

by Brian Bukenya

Assessment of the relationship between sanitary risk inspection and shallow groundwater contamination in Kisumu, Kenya

Brian Bukenya
Student I.D Number: 10785991
School of Earth and Environmental Sciences

Supervisor:
Dr Clare Robinson, PhD

A dissertation submitted to The University of Manchester's Faculty of Science and Engineering in partial fulfilment of the requirements for the Award of a Master's Degree of Science in Pollution and Environmental Control

2021

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

TTC	Thermotolerant Coliforms
DEET	N, N-diethyl-meta-toluamide
SSA	Sub-Saharan Africa
UK	United Kingdom
MS	Microsoft
cfu	Colony Forming Units
mg/l	Milligrams per litre
ANOVA	Analysis of Variance
WHO	World Heath Organisation
WASREB	Water Services Regulatory Board
RDWSSP	Rural Domestic Water Supply and Sanitation Programme
KOP	Kenya Open Data
WWCI	World Weather and Climate Information

Declaration

No portion of the work referred to in this dissertation has been submitted in support of an application for another degree or qualification of this or any other university or institute of higher learning.

Intellectual Property Statement

Final Word Count = 8114

(Does not include title page, table of contents, list of tables, list of figures, list of acronyms and abbreviations, declaration, intellectual property statement, abstract, figure legends, table legends, acknowledgement and references)

Abstract

Nearly half of Sub-Saharan Africa's fast rising population is exposed to contaminated drinking water. The health implications of using contaminated water are undeniable. Sanitary risk inspections and water quality analyses have been recommended by the World Health Organisation as the most suitable approach for assessing the risk of water source contamination. Various experts have produced contradictory conclusions addressing the relationship between these two techniques. With thermotolerant coliforms as the microbiological and (nitrate, chloride, sulphate, fluoride, and phosphate) as the chemical parameters, this research project focused on assessing the relationship between sanitary risk inspection and shallow groundwater contamination. This study used secondary data which were downloaded from the Environment Information Data Centre website. Sanitary risk inspection and water quality datasets from 40 shallow groundwater wells in two peri-urban regions of Kisumu, namely Manyatta and Migosi, were used to characterise the secondary data. The sanitary risk inspection data and TTC data were translated into % sanitary risk scores and Log10 TTC, respectively, prior to statistical analysis. Sanitary risk scores (%) varied from medium (50%) to very high (100%) with a high average of 79.38%. The presence of scattered waste within 30m of the well and the unsanitary well covers posed the very highest risk (95%) to groundwater contamination. All of the samples surpassed the WHO recommendation threshold for TTC, however none of the samples exceeded the guideline values for nitrate, chloride, sulphate, or phosphate. Only three of the forty fluoride samples surpassed the WHO recommended limit. These findings suggest that sanitary risk factors contribute to microbiological and chemical pollution of shallow groundwater to a higher and minor extent, respectively. The weak positive associations between the % sanitary risk scores and: Log10 TTC (r^2=0.006, p=0.624), nitrate (r^2=0.043, p=0.198), chloride (r^2=0.002, p=0.764), and sulphate (r^2=0.027, p=0.315) were not statistically significant, according to the research. The moderate inverse relationship between the percentage sanitary risk scores and phosphate (r^2=0.091, p=0=089) was also not statistically significant. On the other hand, the moderate inverse relationship between the percentage sanitary risk scores and fluoride was found to be statistically significant (r^2=0.121, p=0.028). Future studies should address the study's potential limitations by: 1) taking multiple measurements and expanding the sampling space to reduce variability; 2) incorporating more factors such as: location for spatial analysis, seasonality, rainfall events, and well depth, among others; and 3) weighting risk factors using appropriate models for better representation. It is therefore very critical to include more appropriate datasets in research attempting to demonstrate the relationship between sanitary risk inspection and shallow groundwater contamination in peri-urban areas in Sub-Saharan Africa, as well as elsewhere.

1 Introduction

1.1 Background

The global urban population is expected to increase from 54 percent in 2014 to 66 percent by 2050, resulting in an increase of 2.5 billion people, with 90 percent of them concentrated in Africa and Asia (*United Nations, 2015*). For the same timelines, this report further highlights that the increase will be 37% to 55% for the Sub-Saharan Africa (SSA) urban population where 55% of this population lives in slums (*UN-Habitat*, 2014). While access to water and sanitation is a human right (United Nations, 2016), the growing urbanisation in developing nations has sparked water security concerns due to unequal infrastructure expansion, putting the most vulnerable slum residents at risk. According to WHO (2017), 42 percent of people in Sub-Saharan Africa do not have access to a basic water supply, which is defined as a better water source that can be accessed in less than 30 minutes (Kelly., *et al* 2021). This population is obliged to obtain water from surface resources, illegal main distribution system connections, vendors, and ground water in order to meet its domestic water needs (Grönwall et al., 2010).

Groundwater has been cited to be the largest and most important resource in Africa (MacDonald et al., 2012) following its reliability in terms of proximity to users, vulnerability to pollution and resilience to climate irregularities compared to surface water resources (MacDonald *et al*., 2011; Lapworth *et al*., 2013). Much as ground water is very affordable and available for domestic use by the urban poor, its quantity and quality may be strained due to the rapid population growth, unsustainable land use and climate change (Okotto *et al*., 2015). Following their low cost of establishment where hand digging is employed, shallow wells happen to be the most commonly used sources of groundwater in low income urban communities which face difficulties in accessing reliable piped water supplies and sewerage systems (Pedley *et al*., 1997; Cronin *et al*., 2007).

Using N,N-diethyl-meta-toluamide (DEET) insect repellent as a tracer, Sorensen *et al*., (2015) disclosed that emerging contaminants in the SSA were more prevalent in shallow wells of informal settlements due to the inadequate well protection, sanitation and waste disposal. However, these researchers emphasised the lack of personal care items, lifestyle chemicals, and medications due to their limited use in the developing SSA and their widespread use in the industrialised world. Solid waste management for instance is quite a driver for groundwater contamination following the rapid increase in waste generation due to population growth coupled with under-funding which consequently leads to insufficient collection, transportation, treatment and disposal of the waste (Sibanda *et al*., 2017; Shamim *et al*., 2019). The microbiological quality of shallow groundwater sources has been found to change periodically, with a notable degradation at the start of the wet season (Barrett *et al*., 2000; Wright, 1986).

The pollution from onsite sanitation facilities such as pit latrines has been blamed for the contamination (Lewis *et al*., 1984). Wright *et al*., (2013) found that there was no correlation between the levels of thermotolerant coliform (TTC), a microbiological biomarker, and daily rainfall patterns in their study. Bain *et al*., (2014) estimated that 52 percent of the SSA population was exposed to faecally polluted drinking water, while Kumpel *et al*., (2016) found that 31% of groundwater sources in the same region (SSA) exceeded the WHO limit of no identifiable faecal indicator bacteria per 100-mL sample. Wright *et al*., (2013) also discovered that the population density and proximity of pit latrines had an impact on the chemical contamination of ground water in Kisumu's peri-urban districts. In their (Wright *et al*., 2013) study, they found that the density of pit latrines within a 100-meter radius significantly raised nitrate and chloride levels. Over the study period (2004 and 2014), sanitary concerns at the wells grew, whereas nitrate levels and thermotolerant coliform counts decreased, according to Okotto-Okotto *et al*., (2015). These researchers theorised that the high nitrate levels at the start of the study period were due to the high rainfall in 2004. The public health consequences of using contaminated ground water are apparent, as demonstrated by Mondal *et al*., (2014), who found that end-user practises far outnumbered the microbiological quality of supplied water when assessing diarrheal disease burdens. As if that was not enough, Okotto *et al*., (2015) discovered that while a significant volume of groundwater was used by slum dwellers for purposes other than drinking and cooking, a small percentage of them drank it untreated.

According to the WHO (2004), the most effective way to continuously ensure the safety of a drinking water supply is to utilise a complete risk assessment and risk management methodology. Cronin *et al*., (2006) regard sanitary inspection to be an effective and low-cost risk assessment method since it visually assesses risk elements that may contribute to the chance of pollution in groundwater and other water systems. This study focused on exploring the relationship between sanitary risk inspection and shallow groundwater contamination, as this is crucial when determining the scope of water supply and sanitation improvements.

1.2 Aim

This project was therefore aimed at examining the relationship between sanitary risk inspection and shallow groundwater contamination in the peri-urban areas of Kisumu, Kenya.

1.3 Objectives

In order to achieve the project aim, it was necessary to:

1. Quantify the potential sanitary risks at shallow ground water wells;
2. Assess the relationship between sanitary risk scores and contamination parameters of shallow groundwater wells. This objective was further broken down into two subobjectives as follows:
   a) Assessed the relationship between sanitary risk scores and microbial contamination of shallow groundwater wells. TTC was considered as the microbial parameter for this study.
   b) Assessed the relationship between sanitary risk scores and chemical contamination of shallow groundwater wells. The study focussed on nitrate, chloride, sulphate, fluoride and phosphate as the chemical parameters.

1.4 Research Questions

This research project therefore addressed the following questions:

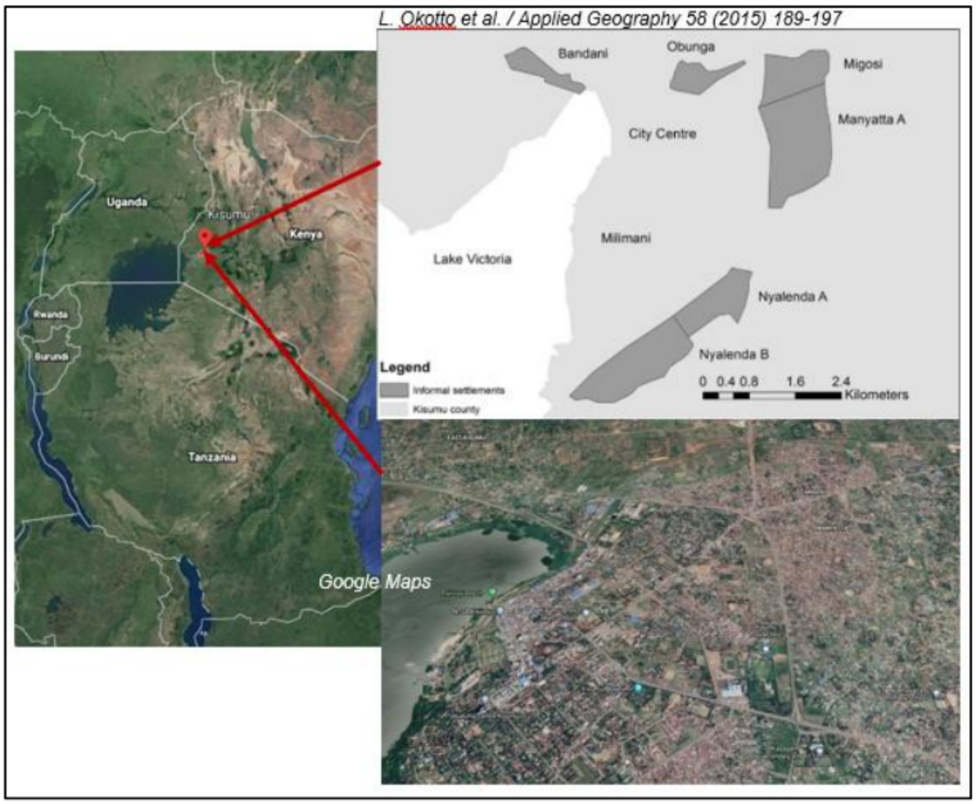
1. What is the quantification of the sanitary risk scores at the shallow ground water wells?
2. What is the nature of the relationship between sanitary risk scores and:
   a) thermotolerant counts at shallow groundwater wells?
   b) nitrate levels at shallow groundwater wells?
   c) chloride levels at shallow groundwater wells?
   d) sulphate levels at shallow groundwater wells?
   e) fluoride levels at shallow groundwater wells?
   f) phosphate levels at shallow groundwater wells?

2 Methodology

This section describes the geographical extent of the study region, the materials used, how they were employed, and the procedures used to gather, present, analyse, and interpret secondary data during the research project. The study was aimed at assessing the relationship between sanitary risk inspection and the contamination of shallow groundwater wells in Kisumu, Kenya.

2.1 Study Area

Figure 1: Geographical illustration of the study area



The study area is situated in Kisumu, a city in western Kenya on the eastern beaches of Lake Victoria, on the Winam Gulf. Kisumu is located at a height of roughly 1174 metres above sea level, between latitude 00°02′N; 00°11′S and longitude 34°35′W; 34°55′E. The city is approximately 417 km^2 in size, with Lake Victoria waters accounting for 35.5 percent of the total area. Having been named by the UN-Habitat in 2006 as the first UN Millennium city in the world, Kisumu city is also the third largest city in Kenya and the second largest in the lake Victoria basin with an estimated core urban population of 259,258 (KOD, 2016).

The city's higher-income neighbourhoods are surrounded by informal settlement regions that lack sewered sanitation and reliable piped water (UN-Habitat., 2005). Bandani, Manyatta (A & B), Migosi, Nyalenda (A & B), Nyamasaria, and Obunga are the names of the informal settlements. This research project focused on two of the aforementioned informal settlements, namely Migosi and Manyatta A to the east of the Kisumu city centre (Figure 1). Manyatta A and Migosi cover 2.36 km^2 and 1.93 km^2 of land, respectively (Okotto-Okotto *et al*., 2015). The population densities of Manyatta A and Migosi were 203 and 103 people per hectare respectively according to the Kenya Census of 2009. Groundwater was used by 39 percent of Manyatta A homes and 24 percent of Migosi households to meet domestic needs, with piped water being used by 31 percent and 23 percent, respectively (Okotto-Okotto *et al*., 2015). Pit latrines were used by 91 percent of Manyatta A families and 38 percent of Migosi homes, respectively, with septic tanks being used by 4 percent and 29 percent (Okotto-Okotto *et al*., 2015). The research area includes a gently sloping environment with a tertiary phonolitic lavas underlay that exposes partially decomposed phonolites and a few porphyritic granite intrusions in some spots (RDWSSP, 1987). This geology may be traced back to the post-Nyanzian and Bukoban periods, when bouldered rock formations of various sizes with widespread cracks and joints were discovered primarily on the route to the surface. Wells are sunk into the fractured phonolitic basalt and pyroclastics in the west and east, respectively, while the pyroclastic deposit on top of the phonolites grows in thickness towards the southeast of the study sites. A thin layer of reddish-brown dirt covers the majority of the rock surface (Okotto-Okotto *et al*., 2015). In most places, the study area's soil associations are built on lateritic and degraded rock material with medium to low flexibility and good drainage (Andriesse *et al.,* 1985). However, there is a small piece of Migosi's low lying eastern sections with restricted drainage, resulting in waterlogging on occasion, as evidenced by the inundation of roughly 2% of the wells during wet seasons. As a result of the short well depth, fracture, and lack of heavy soil cover in the research area, the aquifer system is extremely fragile (Okotto-Okotto *et al*., 2015).

2.2 Secondary data collection

The secondary data was taken between February 23rd and March 4th, 2014, which coincided with the commencement of the rainy season (WWCI, 2021). A cooperation between the University of Southampton (as co-ordinating institution), VIRED International, Kenya's Jaramogi Oginga Odinga University of Science and Technology (JOOUST), and the University of Surrey gathered this secondary data. At each of the shallow wells sampled, single measurements and observations were made. The secondary data for this study were obtained by downloading them from the Environment Information Data Centre website (Okotto-Okotto *et al*., 2014; Pedley *et al*., 2014).

2.3 Sampling

Groundwater sources were chosen by selecting the groundwater source that was nearest to a random position in both neighbourhoods (Migosi and Manyatta) (Okotto-Okotto *et al.*, 2014; Pedley *et al.*, 2014).

2.4 Procedure: Sanitary risk inspections

Secondary data collection used sanitary risk inspection forms adopted from WHO, (1997) where each form comprised 12 yes/no questions representing the presence/absence of sanitary risk factors. By direct observation, these sanitary risk inspection forms/ observation checklists were used to identify sanitary risk factors (Table 2) at well heads and immediate surroundings by single selection of the yes/no questions under the response column at the sampled wells (Okotto-Okotto *et al.*, (2015). Each shallow well sampled was subjected to a single observation. Furthermore, well owners were notified when the sanitary risk inspection revealed an obvious contamination hazard/sanitary risk factor that could be corrected relatively easily, such as if a well lacked a cover (Okotto-Okotto *et al.*, (2015).

2.5 Procedure: Groundwater quality

2.5.1 Sample collection and fieldwork:

An approximate of 500ml groundwater sample was collected from each surveyed well using stainless steel cups that were rinsed several times in well-water. Having been stored in iced sterile plastic bottles and a darkened environment, all samples were then safely transferred to the laboratory and processed within seven hours of collection starting with the microbial samples to avoid the possibility of cross contamination (Okotto-Okotto *et al.*, 2015).

2.5.2 Laboratory instrumentation and methods

Using the membrane filtering method and Membrane Lauryl Sulphate Broth (MLSB, Oxoid, UK) as the selective medium, the thermotolerant coliforms (TTC) were extracted and quantified. A 0.45m nitrocellulose membrane filter was used to filter 100ml of the sample, or an adequate dilution, under vacuum (Pall-Gelman, UK). The membranes were placed on MLSB-soaked pads, pre-incubated for one hour at room temperature, and then moved to a $44 \pm 0.5^{o}C$ incubator (DelAgua Water Test Kit) for a total incubation duration of 20±2 hours. After 15 minutes of incubation, the membranes were taken from the incubator and checked for characteristic coliform colonies (cream to yellow colonies with a diameter higher than 1mm). The number of colonies was counted, and the volume of the original sample was converted to 100ml (Okotto-Okotto *et al.*, 2015). Nitrate, sulphate, phosphate, fluoride, and chloride concentrations were tested using a Palintest photometer, model 7100, using specialised reagents (Palintest Ltd, UK) and following the manufacturer's instructions (Okotto-Okotto *et al.*, 2015).

2.5.3 Calibration steps

Before each sampling operation, the temperature of the DelAgua Water Test Kit incubator was checked and adjusted (if necessary), and then every two days during the work with a liquid thermometer.

2.5.4 Quality control

With the lab's facilities, quality control of analytical results was difficult. For each analytical procedure, including the microbiological tests, sample blanks were conducted. When the number of samples allowed (the DelAgua Water Test kit may incubate a maximum of 16 samples) was reached, double analysis was performed on the first and final samples of the batch to ensure analytical consistency. Individual tests were performed by the same analyst for each sampling work to reduce fluctuation over the course of the work (Okotto-Okotto *et al.*, 2015).

2.6 Data analysis and statistical methods

For this study, secondary data analysis was carried out using MS 365 Excel version 2106 and the versatile RStudio version 4.1.0 which provides free and open-source tools for R and enterprise-ready professional software. Prior to any statistical analysis, the sanitary risk inspection data and TTC data were transformed into percentage sanitary risk scores and Log10 TTC respectively. Multivariate statistical analysis was carried out by descriptive statistical analysis, box plots, correlation analysis and regression analysis. Histograms were used to represent the frequency distribution of continuous variables. Pie charts and bar graphs were used to visualize qualitative data. Descriptive statistical analysis, histograms and box plots were used to clearly illustrate variations in the sanitary risk scores (%), the Log10 transformed microbial contamination (Log10 TTC) and the chemical contamination (nitrate, chloride, sulphate, fluoride and phosphate). A pie chart was used to illustrate the fraction of samples collected from the sample areas. Bar graphs were used to illustration the quantification of sanitary risk scores at sample wells. Correlation analysis was conducted to quantify the relation degree of two individual variables (sanitary risk scores, Log10 TTC, nitrates, chlorides, sulphates, fluorides and phosphates). This study used regression analysis to examine the relationship between sanitary risk scores and the contamination parameters as supported by Okotto-Okotto *et al.*, (2015). Six model scenarios were generated with simple linear regression modelling using R where data for sanitary risk scores and one contamination parameter at a time were used as the predictor and response variables respectively.

3 Results

This part covers the presentation and interpretation of findings as perceived and based on the study's aims, using secondary data. This section also includes comparisons of the findings to earlier studies of a similar nature. This research project focused on assessing the relationship between sanitary risk inspection and shallow groundwater contamination.

3.1 Descriptive analysis

The percentage sanitary risk scores and contamination metrics (Log10 TTC (cfu/100ml), nitrates (mg/l), chlorides (mg/l), sulphates (mg/l), fluorides (mg/l), and phosphates (mg/l)) are summarised in Table 1.

Table 1: Descriptive statistical analysis table showing the measures of central tendency and dispersion for the data variables

	Sanitary Risk Score (%)	*Log10 TTC (cfu/100ml)*	*Nitrate (mg/l)*	*Chloride (mg/l)*	*Sulphate (mg/l)*	*Flouride (mg/l)*	*Phosphate (mg/l)*
Sample Size (n)	40	40	40	40	40	40	33
Mean	79.38	3.02	3.94	67.22	44.58	1.37	2.33
Standard Deviation	11.48	1.02	3.94	39.57	15.57	1.31	2.86
Median	79.15	3.05	2.37	67.5	43.5	1.02	1.65
Mode	75	4.48	0.36	87	40	0.98	1.2
Trimmed Mean	79.43	3.12	3.41	65.34	43.41	1.06	1.81
Mean Absolute Deviation (MAD)	6.15	0.82	3.25	45.96	14.08	0.26	0.67
Minimum	50	-1	0.08	1	14	0.57	0.43
Maximum	100	4.5	20.2	190	94	7.9	17.5
Range	50	5.5	20.12	189	80	7.33	17.07
Skewness	-0.16	-1.55	1.81	0.58	0.76	3.71	4.5
Kurtosis	-0.36	4.13	4.82	0.5	0.94	14.05	20.96
Standard Error	1.81	0.16	0.62	6.26	2.46	0.21	0.5

Figure 2: A pie chart showing the fraction of sample collection from the composition of sample area and size (n=40)

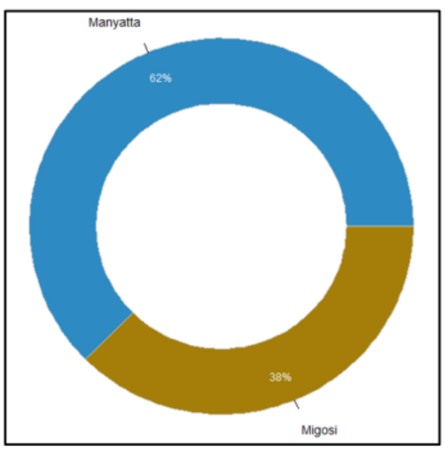
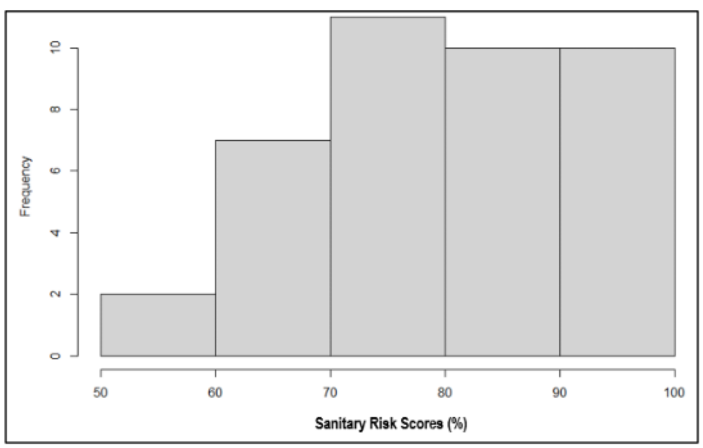


Figure 3: A histogram showing the distribution of the percentage sanitary risk scores data throughout the sample area



Within the two areas that formed the scope of sample collection, 62% and 38% of the samples were collected from Manyatta and Migosi respectively (Figure 2).

3.1.1 Variations in the sanitary risk scores data at sample wells

The number of negative elements ("Yes" response) as a percentage of the total number of risk factors (Table 2) was calculated to transform sanitary risk inspection data from secondary data into percentage sanitary risk scores. These calculations were carried out using the Microsoft Excel application. The percentage sanitary risk scores were classed as: low risk (0-20%), medium risk (30-50%), high risk (60-80%) and very high risk (90-100%) similar to those used in other studies (Lloyd *et al.*, 1991; Patrick *et al*, 2011; Snoad *et al.*, 2017). The data for the percentage sanitary risk ratings ranged from 50% to 100%, with a median of 79.15 percent (Table 1) and one outlier (Figure 13). The data was fairly symmetrical and somewhat skewed to the left, meaning that the majority of the scores fell between 70 and 100 percent (Figure 3).

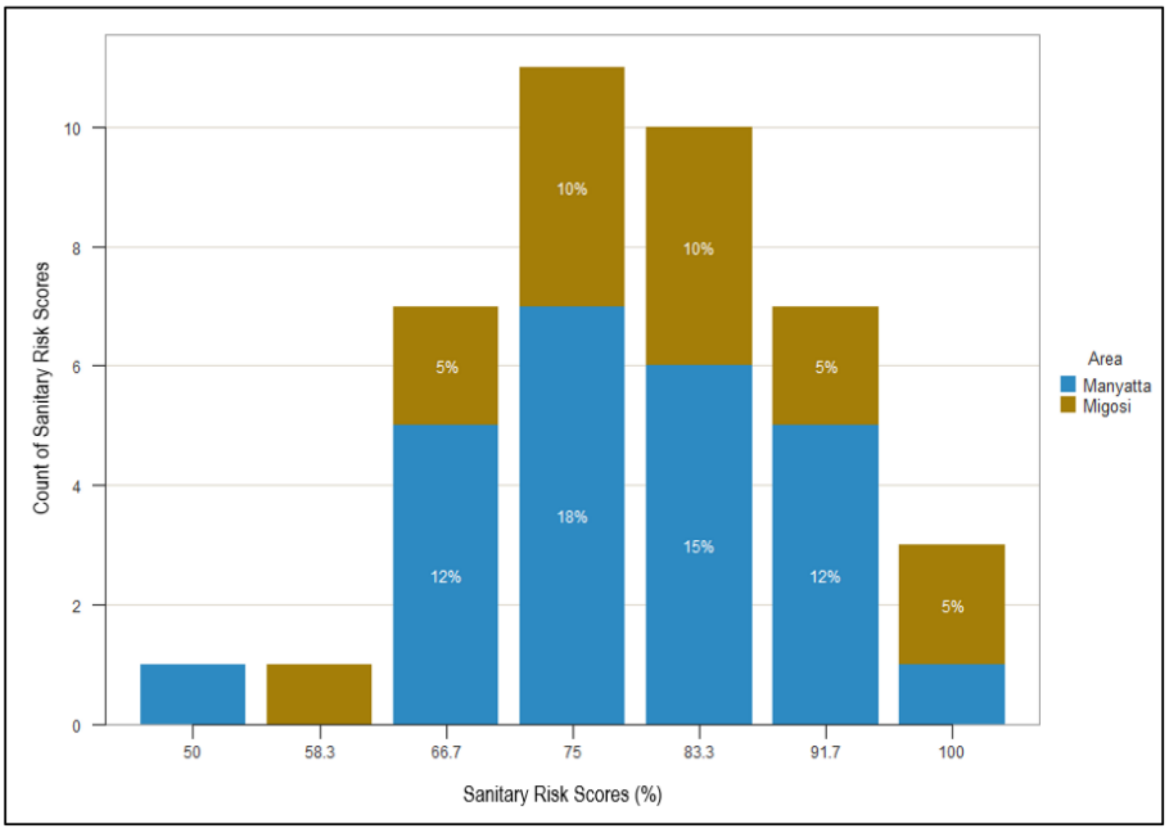
Table 2: Observation checklist used in the identification of sanitary risk factors

Sanitary Risk factors	**Response** **(**0 = No (Positive factor); 1 = Yes (Negative factor)**)**
Does the cement floor extend less than 1.5m from the well? (NarrowCementFloor)	No/Yes
Is there ponding of water on the cement floor? (WaterPondingFloor)	No/Yes
Are there cracks in the cement floor? (CementCracks)	No/Yes
Is there ponding of water beyond the cement floor? (WaterPondingOther)	No/Yes
Is the drainage channel cracked or needs cleaning? (DrainChannelProb)	No/Yes
Do animals have access within 10m of well? (AnimalAccess)	No/Yes
Are there any latrines within 10m of well? (Latrines10m)	No/Yes
Are there any additiional latrines within 30m of well? (Latrines30m)	No/Yes
Is there any scattered waste within 30m of the well? (WasteNearby)	No/Yes
Are there any waste dumps within 30m of the well? (DumpsNearby)	No/Yes
Is the cover of the well unsanitary? (UnsanWellCover)	No/Yes
Are the well walls inadequately sealed at any point below ground level? (WallsUnsealed)	No/Yes

3.1.2 Quantification of sanitary risk scores at sample wells

The computations of the percentage sanitary risk scores at the sampled wells were expounded in 3.11. According to Figure 4, the percentage sanitary risk scores at sample wells ranged from 50% (medium) to 100% (very high). In their study, Omara *et al.*, (2019), reported that the percentage sanitary risk scores ranged from high (60%) to very high (90%). In another study conducted in Bangladesh by Ercumen *et al.*, (2017), 31%, 45% and 22% of the groundwater sources were at low, medium and high risk based on the percentage sanitary risk scores. The percentage scores with the highest counts ranged from 66.7% to 91.7% with 75% as the score with the highest count (Figure 4). The highest-count percentage scores varied from 66.7 percent to 91.7 percent, with 75 percent being the highest-count percentage score (Figure 4). Meanwhile, the counts of the scores were attributed to the areas that made up the sample size (Manyatta and Migosi). Manyatta had the highest counts of the percentage sanitary risk ratings at sampled wells, with the exception of 58.3 percent and 100 percent. This could be explained by the fact that Manyatta (62 percent) had more samples gathered than Migosi (42 percent) as referenced by Figure 2.

Figure 4: A stacked bar graph showing sanitary risk scores for sample wells against their count in the composition of the sample area (Manyatta and Migosi)



Specific sanitary risk factor scores

The specific risk factor scores (percent) for the sample region, on the other hand, were calculated by dividing the number of "Yes" replies (Table 2) by the sample size (n=40) and then shown with a scatter bar chart (Figure 5). In reference to Figure 5, the specific sanitary risk factor scores ranged from 50% (medium) to 95% (very high) with a median of 86.25%. The narrow extension (less than 1.5m) of the cement floor from the well (NarrowCementFloor) posed the least risk of contamination (50%) while the presence of scattered waste within 30m of the well (WasteNearby) and the unsanitary well covers (UnsanWellCover) both at 95% scores posed the highest risk of groundwater contamination (Figure 5). The increased solid waste generation coupled with its inadequate disposal were also reported by Sibanda *et al.*, (2017) and Shamim *et al.*, (2019) as major risk factors towards the contamination of shallow groundwater wells. Generally, the risk factor scores were high (average = 79.4%), thus their potential towards groundwater contamination was presumed to be significant.

Figure 5: A scatter bar graph illustrating specific risk factor scores throughout the sample size (n=40)

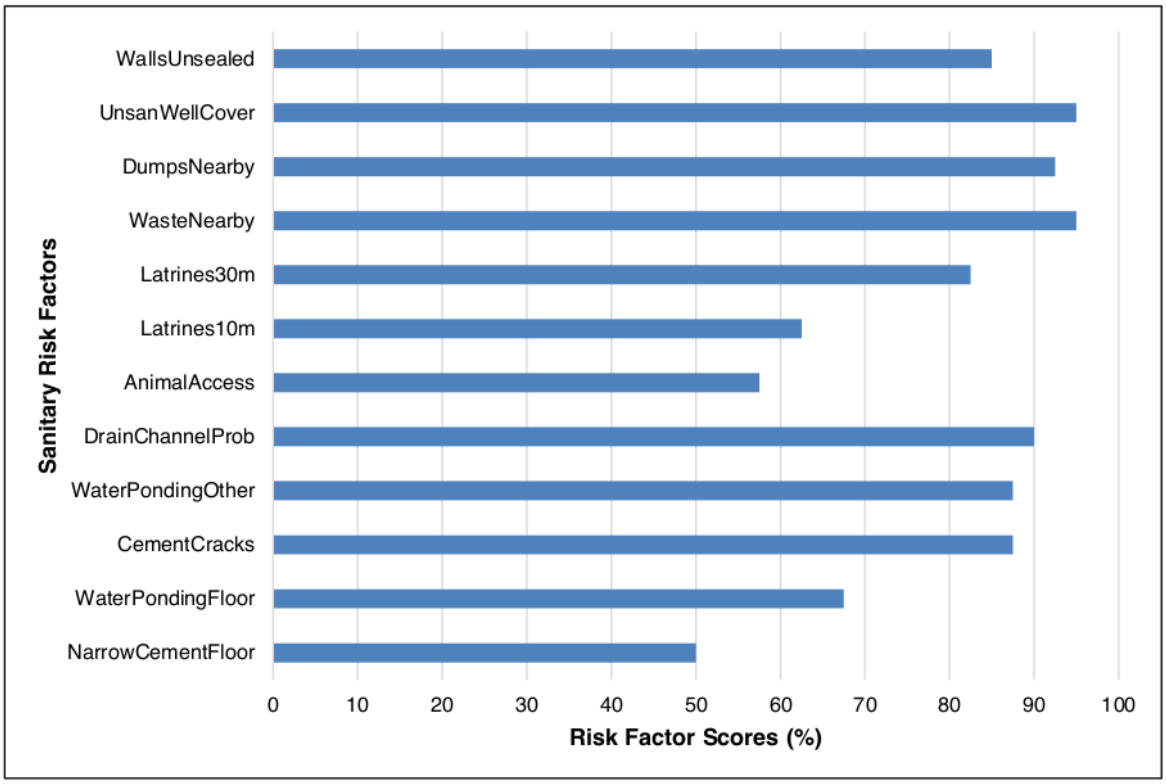


Figure 6: A histogram showing the distribution of the Log10 TTC data throughout the sample area

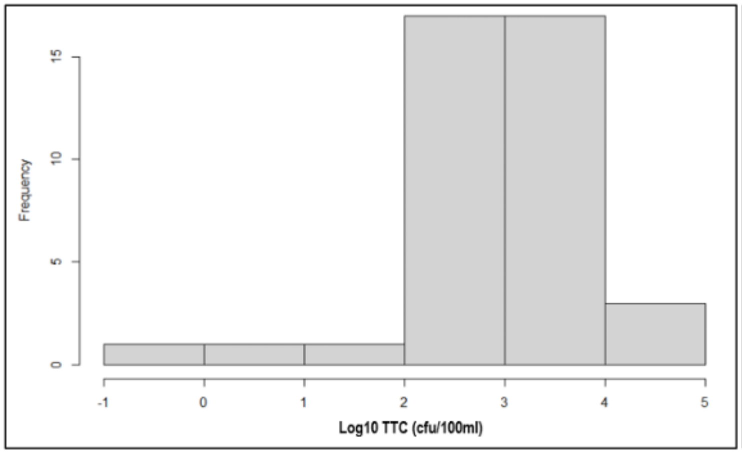


Figure 7: A histogram showing the distribution of the nitrates data throughout the sample area

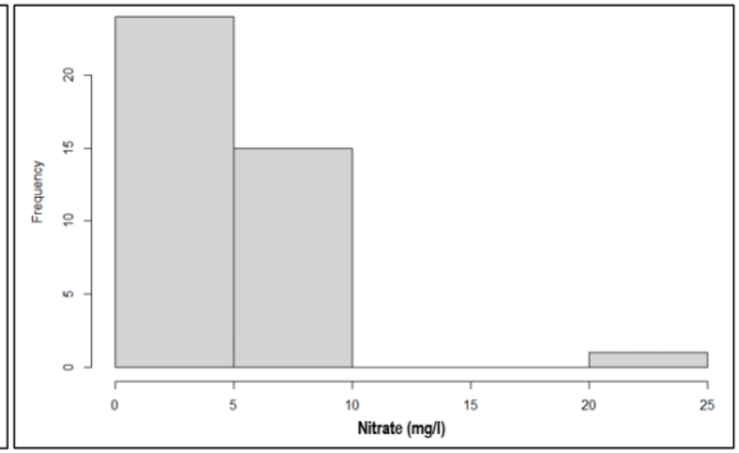


Figure 8: A histogram showing the distribution of the chlorides data throughout the sample area

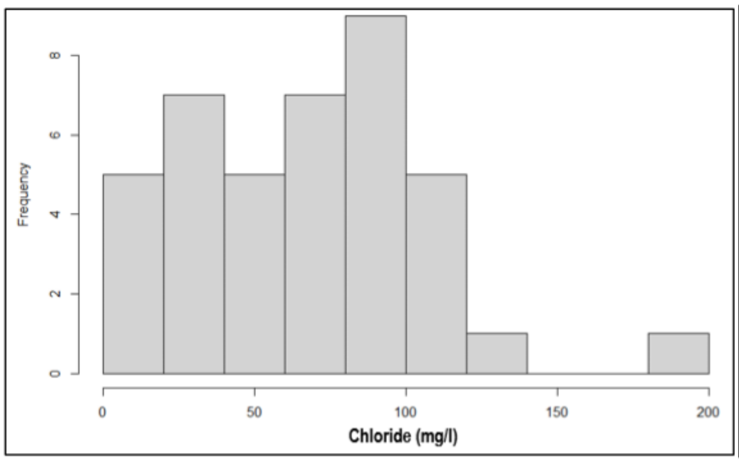


Figure 9: A histogram showing the distribution of the sulphates data throughout the sample area

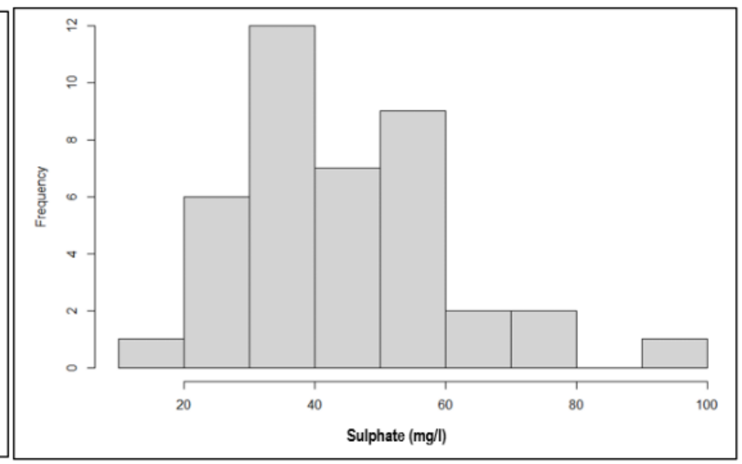


Figure 10: A histogram showing the distribution of the fluorides data throughout the sample area

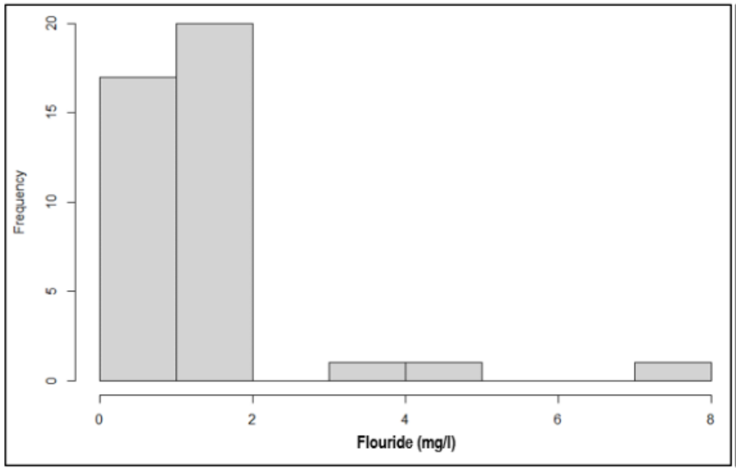


Figure 11: A histogram showing the distribution of the phosphates data throughout the sample area

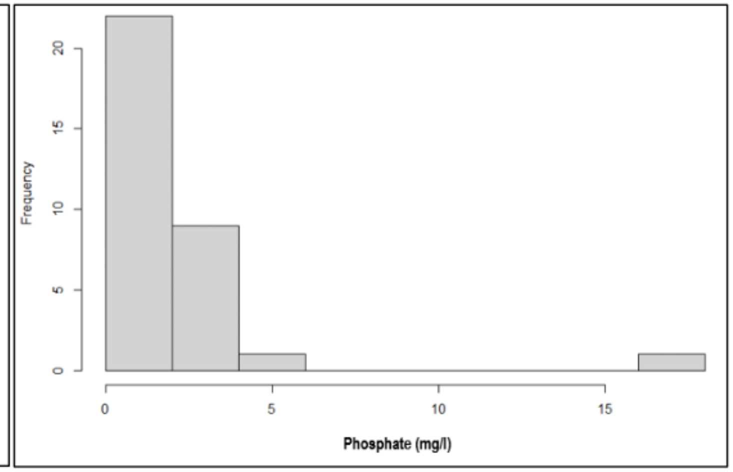


Figure 12: Box plot illustrating the distribution of the Log10 TTC, Nitrates, Fluorides and phosphates data throughout the sample area

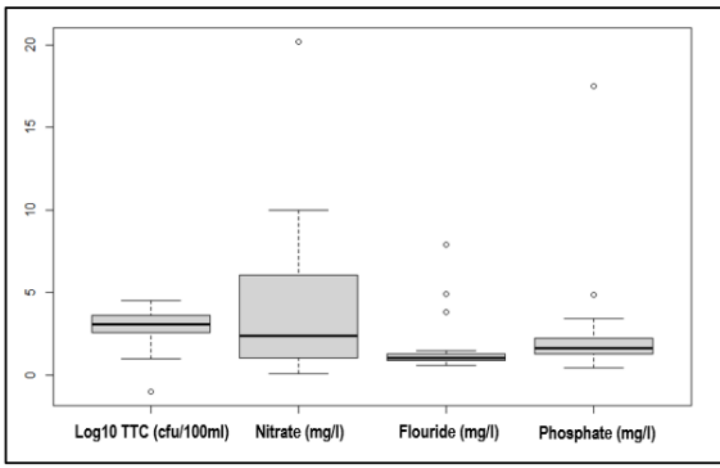
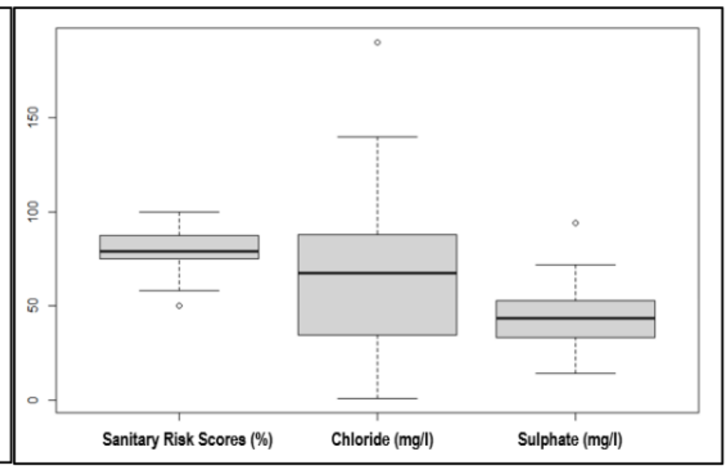


Figure 13: Box plot illustrating the distribution of the sanitary risk scores, chlorides and sulphates data throughout the sample area



3.1.3 Variation in the microbial contamination data

Prior to any statistical analysis of the microbial contamination (TTC) data, a Log10 transformation was performed in MS 365 Excel to improve the normality of the huge TTC counts and make the data modellable using a straight line (Minitab, 2020). The log10 transformed microbial contamination (TTC) data was however not normally distributed due to the high left skewing it showed, with most cfu/100ml ranging from 2 to 4 (Figure 6). The data ranged from -1 to 4.5 cfu/100ml with a median of 3.05 cfu/100ml (Table 1) and one outlier (Figure 12).

3.1.4 Variations in the chemical contamination data

Because of the considerable skewness to the right, the data on nitrates (Figure 7), fluoride (Figure 10) and phosphates (Figure 11) were not normally distributed. For nitrates, fluorides, and phosphates, the concentrations (mg/l) were commonly in the range of 0 to 10, 0 to 2, and 0 to 4, respectively. Furthermore, there were one, three, and two outliers in the nitrates, fluorides, and phosphates data, respectively (Figure 12). The chlorides (Figure 8) and sulphates (Figure 9) data distributions were somewhat skewed to the right and did not follow a typical pattern. As a result, chloride and sulphate concentrations (mg/l) were usually in the range of 0 to 100 and 20 to 60, respectively. There was one outlier in each of these two variables (chlorides and sulphates) (Figure 13).

3.2 Multiple linear regression Modelling

Bivariate analysis was carried out by undertaking a correlation analysis using a multiple linear regression model with R in RStudio. This involved using the percentage sanitary risk scores (SanRiskScore) as the independent variable and contamination parameters (Log10 transformed TTC (Log10_TTC), nitrate, chloride, sulphate, fluoride and phosphate) as the dependent variable(s).

3.2.1 Assumptions of Multiple Linear Regression Analysis

In general, linear regression modelling assumes that there is a linear relationship between the independent and dependent variable(s). For testing, the assumptions of multiple linear regression modelling (Tranmer *et al.*, 2020) were used. There should be no multi-collinearity, residuals should be normally distributed, and variances should be homoscedastic.

The model did not show multi-collinearity because all of the tolerance values (Table 4) were more than 0.2 (Menard, 1996). The residuals (errors) were normally distributed (Figure 14b). Because the average residual variance was close to zero and the residual plot was not cone shaped, the homoscedasticity of variance assumption was also met (Figure 14c).

Figure 14a: Correlation analysis of the data variables (sanitary risk scores, Log10 TTC, nitrates, chlorides, sulphates, fluorides and phosphates)

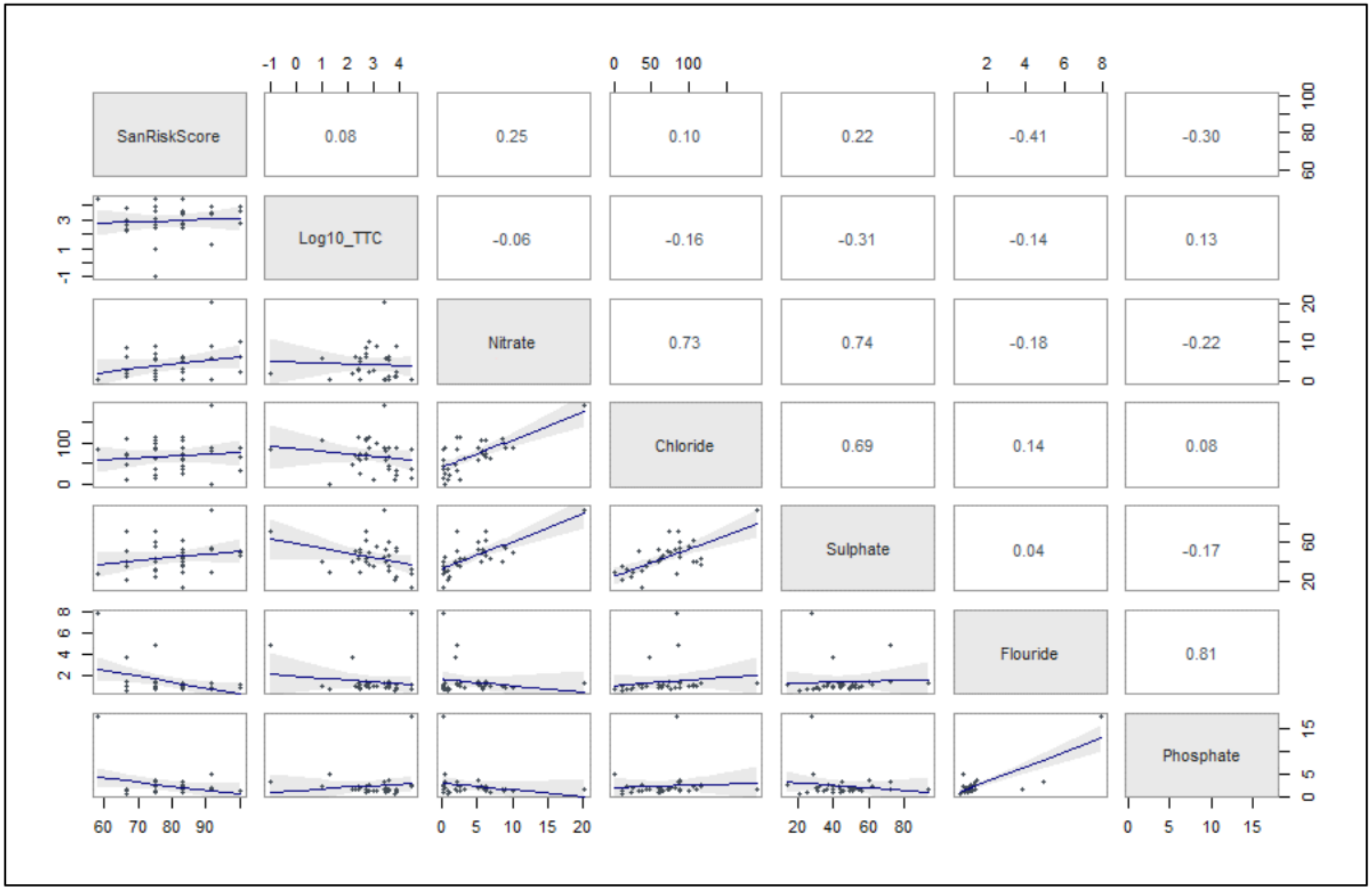


Figure 14b: Residuals' plot generated from multiple linear regression modelling of sanitary risk scores and contamination parameters

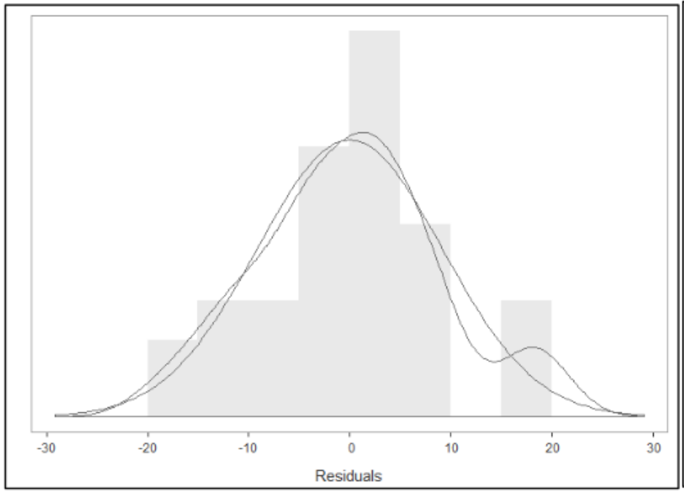


Figure 14c: Fitted values against residuals plot generated from multiple linear regression modelling of sanitary risk scores and contamination parameters

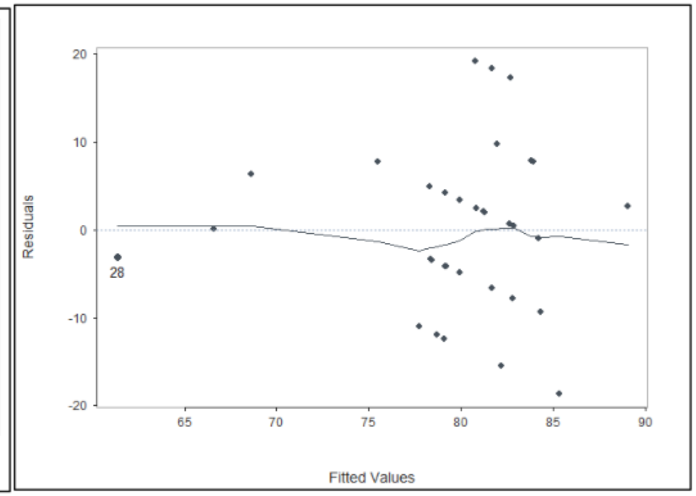


Table 3: Residuals and influence table generated from multiple linear regression modelling of sanitary risk scores and contamination parameters

Data Row	Log10_TTC	Nitrate	Chloride	Sulphate	Flouride	Phosphate	SanRiskScore	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
28	4.5	0.26	83	28	7.9	17.5	58.3	61.361	-3.061	-1.412	-6.35	5.548
23	1.3	0.36	1	29	0.75	4.86	91.7	81.981	9.719	1.515	1.79	0.436
37	-1	2.01	86	72	4.9	3.2	75	68.614	6.386	1.116	1.629	0.375
18	2.7	6.3	74	72	1.39	1.4	66.7	85.355	-18.655	-2.208	-1.13	0.159
40	3.9	2.04	33	51	1.2	1.35	100	82.68	17.32	1.988	0.942	0.114
30	2.8	10	87	49	0.84	1.2	100	81.654	18.346	2.078	0.877	0.097
7	2.7	8.66	110	40	1.01	1.55	66.7	78.691	-11.991	-1.348	-0.7	0.068
16	3.6	6.22	66	46	1.12	1.3	100	80.777	19.223	2.078	0.561	0.04
5	3	2.6	11	35	1.04	1.2	66.7	79.1	-12.4	-1.331	-0.53	0.039
31	3.8	0.92	10	21	0.57	0.43	66.7	77.712	-11.012	-1.191	-0.522	0.038
36	4.5	0.12	37	14	1.17	2.7	83.3	75.503	7.797	0.852	0.425	0.026
11	2.3	5.98	71	51	0.98	1.65	66.7	82.201	-15.501	-1.62	-0.413	0.023
39	3.4	20.2	190	94	1.27	1.5	91.7	89.02	2.68	0.375	0.393	0.023
19	3.6	5.18	88	60	1.49	3.4	75	84.337	-9.337	-0.988	-0.394	0.022
8	2.8	2.16	115	37	1.28	2.7	83.3	78.327	4.973	0.562	0.337	0.017
32	4.5	0.08	16	32	0.76	2.2	75	81.684	-6.684	-0.706	-0.297	0.013
29	3.9	8.92	88	54	0.88	1.7	91.7	83.912	7.788	0.802	0.269	0.01
13	1	5.62	105	40	1.08	2.3	75	78.43	-3.43	-0.394	-0.252	0.009
4	2.4	2.54	115	43	1.24	2.1	75	79.129	-4.129	-0.456	-0.251	0.009
6	3.1	8.76	99	56	0.98	1.2	75	82.813	-7.813	-0.791	-0.219	0.007

Table 4: Collinearity results from the multiple linear regression modelling of sanitary risk scores and the contamination parameters

	Tolerance	VIF
Log10_TTC	0.743	1.345
Nitrate	0.28	3.57
Chloride	0.349	2.865
Sulphate	0.31	3.226
Fluoride	0.246	4.062
Phosphate	0.245	4.08

3.2.2 Outliers

Table 3 on the first row (data row = 28) cited the outlier from the model as supported by the largest cook's distance (D = 5.542), as seen in Figure 14c. Data rows 23 and 37 in Table 3 were also probable outliers since their cook's distances (D=0.436 and D= 0.375, respectively) were considerably different from the rest of the cook's distances. For the sample to be considered a potential outlier using its cook's distance, this study honoured the work of Robinson *etal.*, (1984) who advises the scale of deviation of a sample's cooks' distance from the rest of the samples.

3.2.3 Correlation Analysis

The strength of the bivariate linear associations amongst the variables as illustrated by Figure 14a with the corresponding Pearson correlation coefficients (r) and plots are hereby substantiated as follows.

Weak and positive associations between: the percentage sanitary risk scores and Log10_TTC (r=0.08); the percentage sanitary risk scores and nitrate (r=0.25); the percentage sanitary risk scores and chloride (r=0.10); the percentage sanitary risk scores and sulphate (r=0.220); Log10_TTC and phosphate (r=0.13); chloride and fluoride (r=0.14); chloride and phosphate (r=0.08); sulphate and fluoride (r=0.04) were noticed (Figure 14a). On the contrary, weak and negative associations occurred between; Log10_TTC and nitrate (r=-0.06); Log10_TTC and chloride (r=-0.16); Log10_TTC and fluoride (r=-0.14); nitrate and fluoride (r=-0.18); nitrate and phosphate (r=-0.22); sulphate and phosphate (r=-0.17) according to Figure 14a. Furthermore, moderate and negative correlations were observed between; the percentage sanitary risk scores and fluoride (r=-0.410); the percentage sanitary risk scores and phosphates (r=-0.30); Log10_TTC and sulphate (r=-0.31) according to Figure 14a. Meanwhile, strong and positive associations were realised between; nitrate and chloride (r=0.73); nitrate and sulphate (r=0.74); chloride and sulphate (r=0.69); fluoride and phosphate (r=0.81) according to Figure 14a.

3.3 Simple linear regression analysis

Table 5: ANOVA and model fit results; confidence intervals at 95%; and model equations for the prediction of contamination parameters from sanitary risk scores, linearly regressed against microbial and chemical contamination parameters

	Regression (Reg.) of SanRiskScore (predictor variable) and:	*R-squared*	*F-value*	*P-value*	*Contamination Prediction Equations*	*Lower 95%*	*Upper 95%*
1	**Log10 TTC**	0.006	0.244	0.624	Log10 TTC = 0.007 * SanRiskScore + 2.462	-0.022	0.036
2	**Nitrate**	0.043	1.715	0.198	Nitrate = 0.071 * SanRiskScore + (-1.719)	-0.039	0.182
3	**Chloride**	0.002	0.091	0.764	Chloride = 0.169 * SanRiskScore + 53.823	-0.962	1.300
4	**Sulphate**	0.027	1.037	0.315	Sulphate = 0.221 * SanRiskScore + 27.023	-0.218	0.661
5	**Flouride**	0.121	5.222	0.028	Flouride = -0.040 * SanRiskScore + 4.519	-0.075	-0.005
6	**Phosphate**	0.091	3.087	0.089	Phosphate = -0.081 * SanRiskScore + 8.798	-0.175	0.013

In a bid to assess the relationship between the percentage sanitary risk scores and the contamination parameters (microbial and chemical), simple linear regression modelling was performed with the percentage sanitary risk scores as the predictor variable and individual contamination parameters as the response variables. The model output for each scenario were summarised in Table 5 above including the one way ANOVA results, estimation/prediction equations and confidence intervals at 95%.

3.3.1 Assumptions of Simple Linear Regression Modelling

The assumptions of simple linear regression (O'Farrell, 1970) which include; normal distribution of residuals; residuals (errors) equivalent to nearly zero at each level of the predictor; and homoscedasticity of variances were tested in each model scenario as summarised by Table 6 in the interpretation of results for the different model scenarios.

Table 6: Assumptions of Simple Linear Regression Modelling

	Statistical Assumptions of Simple Linear Regression Modelling		
	Normal distribution of residuals	**Average residuals (errors) equivalent to nearly zero at each level of prediction**	**Homoscedasticity of variances**
Simple Linear Regression Model Scenarios	Assumption was met based on the distribution of residuals plots for the different model scenarios. In this column are the reference plots for the respective model scenarios.	The assumption was met since the average residual error at each level of predictor was nearly zero based on the residuals versus fitted values plots for the different scenarios. In this column are the reference plots for the respective model scenarios.	This assumption was also met since the variances of the residuals were nearly equal across the different levels of the predictor variable based on the residuals versus fitted values plots for the different scenarios. In this column are the reference plots for the respective model scenarios.
Percentage sanitary risk scores and Log10 TTC (Microbial contamination)	Figure 15a	Figure 15b	Figure 15b
Percentage sanitary risk scores and Nitrate (Chemical contamination)	Figure 16a	Figure 16b	Figure 16b
Percentage sanitary risk scores and Chloride (Chemical contamination)	Figure 17a	Figure 17b	Figure 17b
Percentage sanitary risk scores and Sulphate (Chemical contamination)	Figure 18a	Figure 18b	Figure 18b
Percentage sanitary risk scores and Fluoride (Chemical contamination)	Figure 19a	Figure 19b	Figure 19b
Percentage sanitary risk scores and Phosphate (Chemical contamination)	Figure 20a	Figure 20b	Figure 20b

Figure 15a: Residuals' plot generated from the regression of Log10 TTC (cfu/100ml) and sanitary risk scores (%) at sample wells

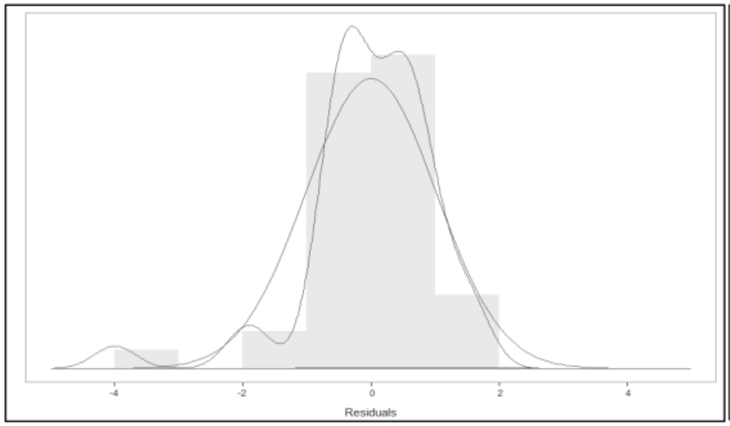


Figure 15b: Fitted values against residuals plot generated from the regression of Log10 TTC (cfu/100ml) and sanitary risk scores (%) at sample wells

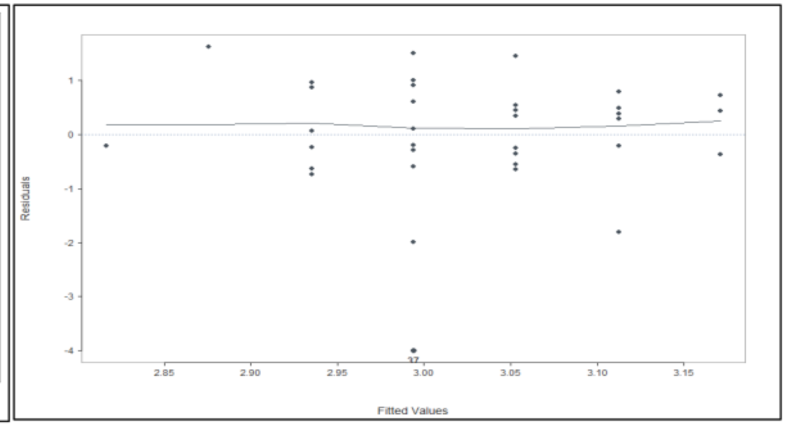
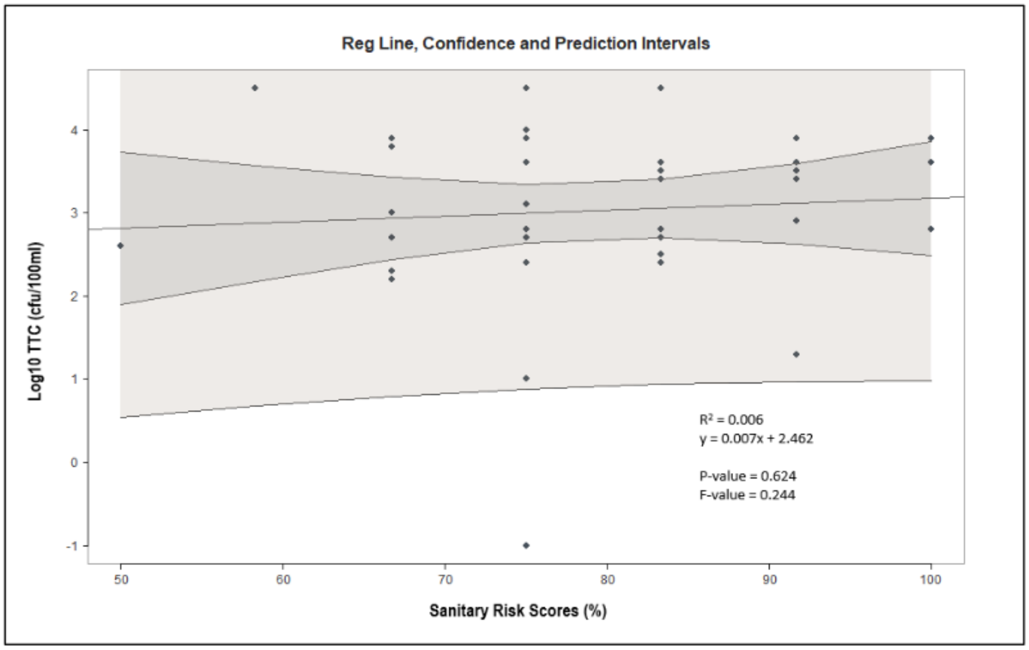


Table 7: Residuals and influence table generated from simple linear regression modelling of Log10 TTC (cfu/100ml) and sanitary risk scores (%)

Data Row	SanRiskScore	Log10_TTC	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
37	75	-1	2.994	-3.994	-5.047	-0.868	0.229
28	58.3	4.5	2.876	1.624	1.716	0.608	0.176
23	91.7	1.3	3.112	-1.812	-1.869	-0.449	0.095
13	75	1	2.994	-1.994	-2.046	-0.352	0.057
40	100	3.9	3.171	0.729	0.745	0.259	0.034
32	75	4.5	2.994	1.506	1.509	0.26	0.033
36	83.3	4.5	3.053	1.447	1.446	0.245	0.029
24	66.7	3.9	2.935	0.965	0.964	0.235	0.028
31	66.7	3.8	2.935	0.865	0.862	0.21	0.022
29	91.7	3.9	3.112	0.788	0.783	0.188	0.018
38	66.7	2.2	2.935	-0.735	-0.731	-0.178	0.016
22	75	4	2.994	1.006	0.991	0.17	0.015
11	66.7	2.3	2.935	-0.635	-0.63	-0.154	0.012
33	75	3.9	2.994	0.906	0.891	0.153	0.012
16	100	3.6	3.171	0.429	0.436	0.152	0.012
30	100	2.8	3.171	-0.371	-0.377	-0.131	0.009
20	91.7	3.6	3.112	0.488	0.482	0.116	0.007
1	50	2.6	2.817	-0.217	-0.231	-0.113	0.007
9	83.3	2.4	3.053	-0.653	-0.638	-0.108	0.006
19	75	3.6	2.994	0.606	0.592	0.102	0.005

Figure 15c: Regression of microbial contamination (Log10 TTC) and sanitary risk scores at sample wells



3.3.2 Thermotolerant Coliforms

The % sanitary risk scores and the Log10 TTC (counts per 100ml) were utilised as the predictor and response variables, respectively, in a simple linear regression model with RStudio (Figure 15c). At data row 37 (Table 7), there was an outlier following the high cook's distance (D=0.229). Other possible outliers included those in data rows 23 and 13, which had cook's distances of 0.095 and 0.057, respectively, and appeared to diverge significantly from the rest of the samples (Table 7). The WHO guideline value of 0 TTC counts in 100ml (WHO, 2001) was exceeded in all 40 samples, as shown in Table 9. In their study, Kumpel *et al.*, (2016) established that 31% of groundwater sources in the entire SSA exceeded the WHO guideline of no detectable faecal indicator bacteria per 100ml sample. Okotto-Okotto *et al.*, (2015) reports that only 5% and 9% of the baseline and follow up samples respectively had no detectable TTCs with all the rest being above the WHO guideline values. As shown in Figure 15c, there was a positive and very weak relationship between the percentage sanitary risk scores and the Log10 transformed TTC data (r^2=0.006). The relationship was however not statistically significant (p=0.624), thus there was a failure in rejecting the null hypothesis. In their study, Wright *et al.*, (2013) had the same observation as no relationship existed between thermotolerant coliform levels and the sanitary risk scores (r=0.01; p=0.91; n=191).

In the study conducted by Howard *et al.*, (2003), a strong and positive relationship occurred between sanitary risk scores and thermotolerant coliform levels (r^2=0.529; p<0.01). Having used the established and alternative models, E. Kelly *et al*., (2021) also reported in their study that no associations existed between sanitary risk scores and E. *coli* occurrence. Luby *et al*., (2008), Lloyd *et al.,* (1989) and Misati *et al*., (2017) did not attain significant associations between sanitary risk scores and TTC counts too. According to Table 5 and Figure 15c, the model developed a regression equation (y = 0.007x + 2.462) that could be used to forecast microbial contamination (y) from percentage sanitary risk scores (x), where 0.007 and 2.462 are the gradient and y intercept, respectively.

Figure 16a: Residuals' plot generated from the regression of nitrates (mg/l) and sanitary risk scores (%) at sample wells

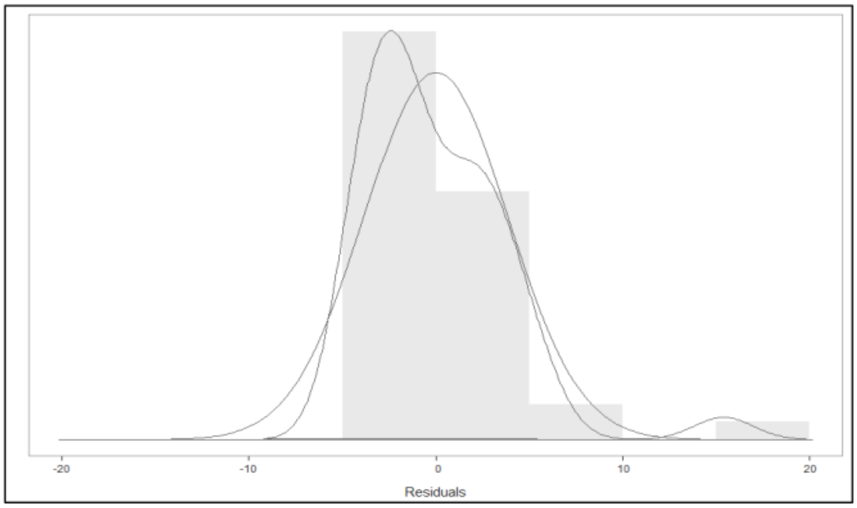


Figure 16b: Fitted values against residuals plot generated from the regression of nitrates (mg/l) and sanitary risk scores (%) at sample wells

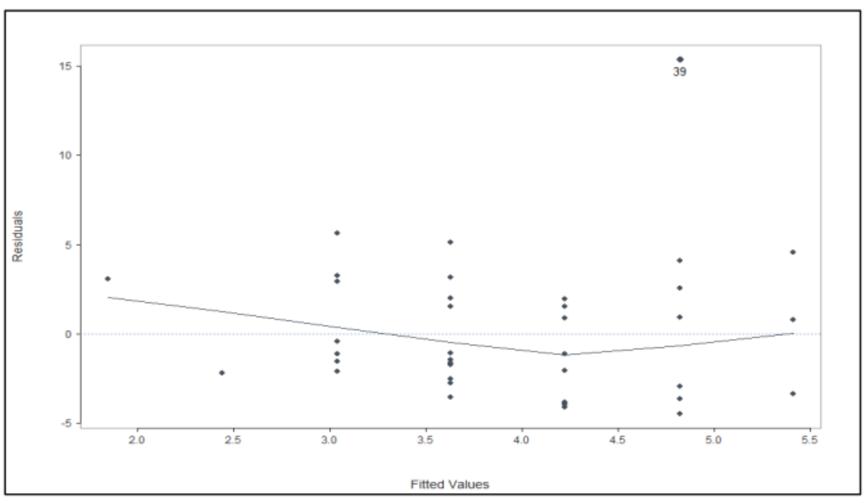
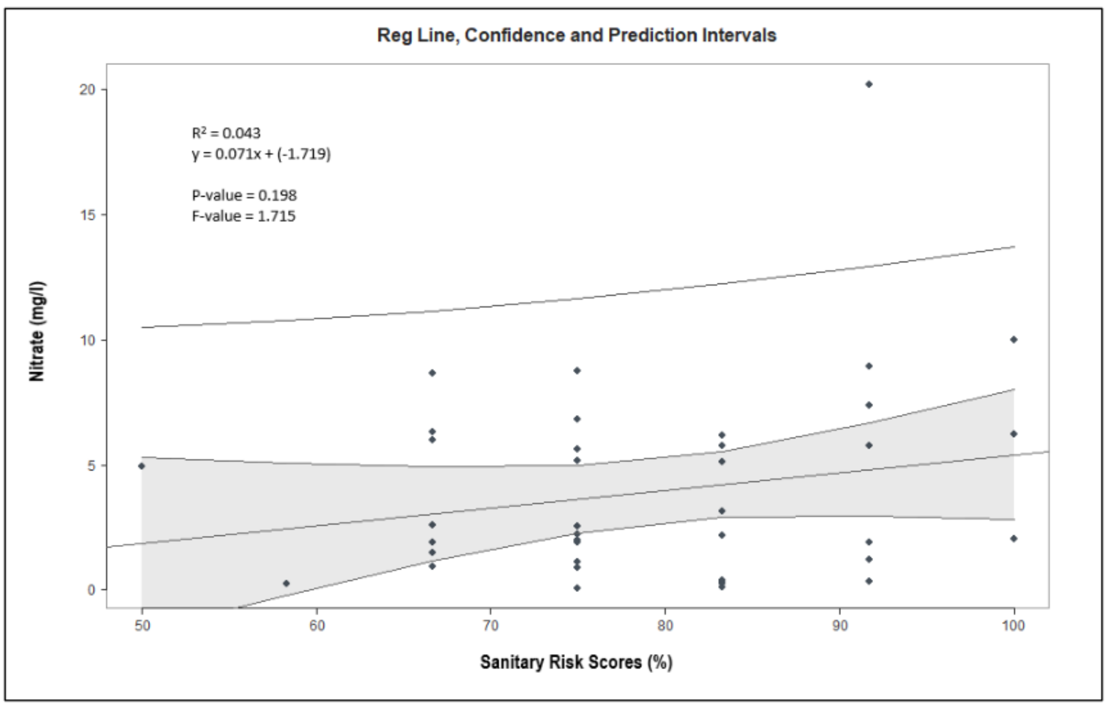


Table 8: Residuals and influence table generated from simple linear regression modelling of nitrates (mg/l) and sanitary risk scores (%)

Data Row	SanRiskScore	Nitrate	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
39	91.7	20.2	4.819	15.381	5.31	1.276	0.474
30	100	10	5.411	4.589	1.255	0.436	0.094
1	50	4.92	1.846	3.074	0.874	0.428	0.092
7	66.7	8.66	3.037	5.623	1.508	0.368	0.066
40	100	2.04	5.411	-3.371	-0.913	-0.317	0.051
23	91.7	0.36	4.819	-4.459	-1.181	-0.284	0.04
29	91.7	8.92	4.819	4.101	1.083	0.26	0.034
6	75	8.76	3.628	5.132	1.349	0.232	0.026
21	91.7	1.2	4.819	-3.619	-0.953	-0.229	0.026
18	66.7	6.3	3.037	3.263	0.858	0.21	0.022
28	58.3	0.26	2.438	-2.178	-0.587	-0.208	0.022
11	66.7	5.98	3.037	2.943	0.772	0.189	0.018
20	91.7	1.88	4.819	-2.939	-0.771	-0.185	0.017
26	83.3	0.12	4.22	-4.1	-1.068	-0.181	0.016
36	83.3	0.12	4.22	-4.1	-1.068	-0.181	0.016
27	83.3	0.24	4.22	-3.98	-1.036	-0.176	0.015
3	83.3	0.36	4.22	-3.86	-1.004	-0.17	0.014
34	83.3	0.38	4.22	-3.84	-0.998	-0.169	0.014
25	91.7	7.38	4.819	2.561	0.67	0.161	0.013
32	75	0.08	3.628	-3.548	-0.921	-0.158	0.013

Figure 16c: Regression of nitrates (mg/l) and sanitary risk scores (%) at sample wells



3.3.3 Nitrate

In this scenario, RStudio was used to perform a simple linear regression model with the percentage sanitary risk scores and nitrate levels (mg/l) as the predictor and response variables, respectively (Figure 16c). Following the maximum cook's distance of 0.474, the model identified a sample at data row 39 (Table 8) as an outlier. Furthermore, due to the considerable deviation of their cook's distances of D=0.094 and D=0.092 respectively compared to the rest of the samples, data rows 30 and 1 were identified as probable outliers. Table 9 shows that none of the 40 samples tested exceeded the WHO nitrate guideline of 50 mg/l (*WHO*, 2006). According to Okotto-Okotto *et al.*, (2015), 56% and 4% of the baseline and follow up samples were above the WHO guideline value of nitrates. In another study conducted by Wright *et al.*, (2013), ten out of 18 samples exceeded the WHO guideline value for nitrate as nitrogen. The relationship between the percentage sanitary risk scores and nitrate levels was a weak and positive one (r^2=0.043) as referenced by Figure 16c. There was however a failure in rejecting the null hypothesis since the relationship was not statistically significant (p=0.198). In a study conducted by Okotto-Okotto *et al.*, (2015), there was rather a statistically significant, strong and positive relationship between sanitary risk factors and nitrate levels (adjusted r^2=0.78; p=0.04). According to Wright *et al.*, (2013), the association between sanitary risk factors and nitrate levels which happened to be medium to large was not statistically significant (r=0.39; p=0.10; n=18). The reduction in nitrate levels with the increase in the distance from the sanitary pits was statistically significant (p=0.02) according to Ezeh *et al.*, (2021). In reference to Table 5 and Figure 16c, the model developed a linear regression equation ($y = 0.071x - 1.719$) that could be used to forecast nitrate contamination (y) from percentage sanitary risk scores (x), with the gradient and intercept being 0.071 and -1.719, respectively.

Table 9: Comparison of sample values with the guideline values

	Guideline Value	Samples that exceeded the guideline value	
		Number	Percentage
TTC	0 in 100ml	40	100
Nitrate	50	0	0
Chloride	250	0	0
Sulphate	250	0	0
Fluoride	1.5	3	7.5
Phosphate	30	0	0

Figure 17a: Residuals' plot generated from the regression of chlorides (mg/l) and sanitary risk scores (%) at sample wells

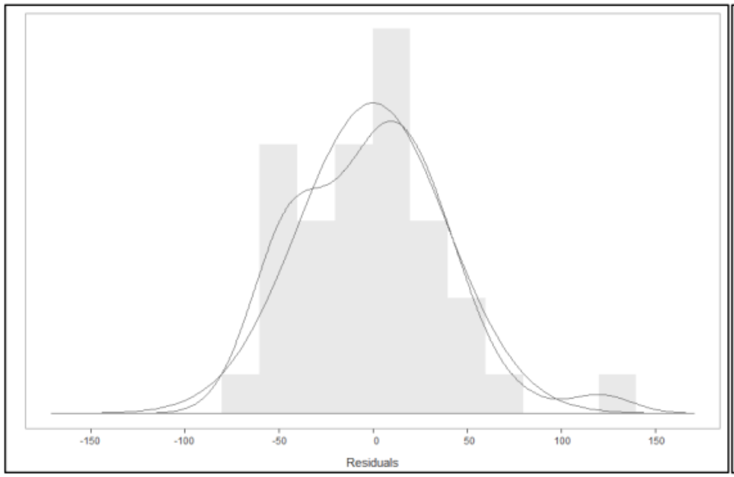


Figure 17b: Fitted values against residuals plot generated from the regression of chlorides (mg/l) and sanitary risk scores (%) at sample wells

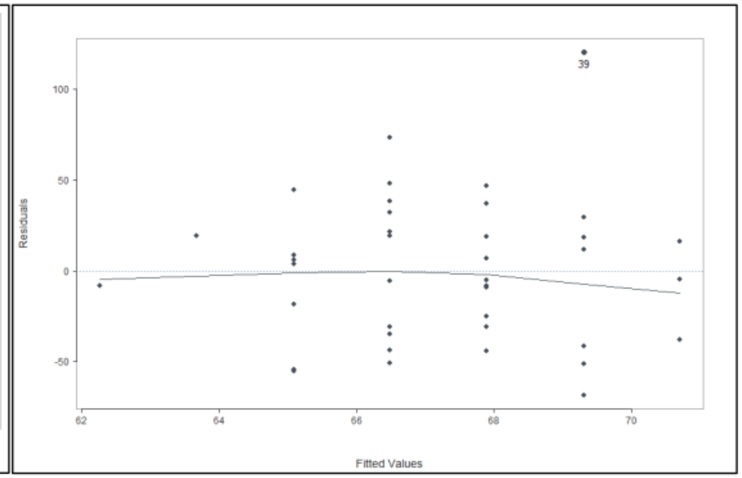
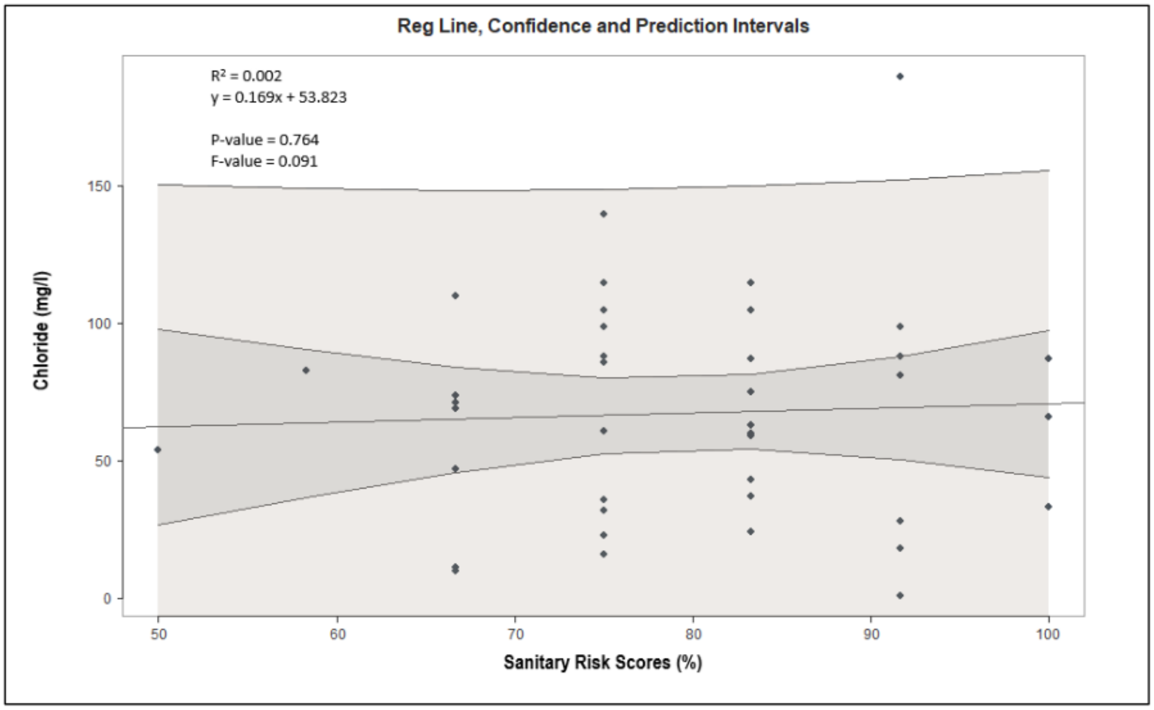


Table 10: Residuals and influence table generated from simple linear regression modelling of chlorides (mg/l) and sanitary risk scores (%)

Data Row	SanRiskScore	Chloride	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
39	91.7	190	69.305	120.695	3.539	0.85	0.277
23	91.7	1	69.305	-68.305	-1.806	-0.434	0.089
40	100	33	70.707	-37.707	-0.997	-0.347	0.06
31	66.7	10	65.085	-55.085	-1.436	-0.351	0.06
5	66.7	11	65.085	-54.085	-1.408	-0.344	0.058
22	75	140	66.486	73.514	1.928	0.332	0.051
21	91.7	18	69.305	-51.305	-1.331	-0.32	0.05
7	66.7	110	65.085	44.915	1.16	0.283	0.04
20	91.7	28	69.305	-41.305	-1.063	-0.255	0.032
32	75	16	66.486	-50.486	-1.29	-0.222	0.024
4	75	115	66.486	48.514	1.238	0.213	0.022
8	83.3	115	67.887	47.113	1.2	0.204	0.021
33	75	23	66.486	-43.486	-1.105	-0.19	0.018
34	83.3	24	67.887	-43.887	-1.115	-0.189	0.018
25	91.7	99	69.305	29.695	0.758	0.182	0.017
28	58.3	83	63.666	19.334	0.507	0.18	0.016
13	75	105	66.486	38.514	0.975	0.168	0.014
14	83.3	105	67.887	37.113	0.939	0.159	0.013
12	75	32	66.486	-34.486	-0.871	-0.15	0.011
30	100	87	70.707	16.293	0.426	0.148	0.011

Figure 17c: Regression of chlorides (mg/l) and sanitary risk scores (%) at sample wells



3.3.4 Chloride

Using the percentage sanitary risk scores and chloride levels as the predictor and response variables respectively, a simple linear regression model was run with RStudio (Figure 17c). The model discovered one outlier, which was identified by data row 39 (Table 10) and was the data sample with the greatest cook's distance. The other possible outlier was at data row 23, where the cook's distance (D) was 0.089, which was similarly considerably different from the D values of the other samples. Table 9 reveals that none of the 40 samples exceeded the WHO guideline value of 250mg/l for chloride (*WHO,* 2006). Wright *et al.*, (2013) also reported that none of the samples exceeded the WHO guideline value of 250mg/l for chloride. According to Figure 17c, the weak and positive relationship between sanitary risk scores and chloride levels (r^2=0.002) was also not statistically significant (p=0.764), thus there was a failure in rejecting the null hypothesis. Wright *et al.*, (2013) got a similar observation, in that the small to medium relationship that existed between sanitary risk scores and chloride levels was not statistically significant (r=0.27; p=0.67; n=18). A simple linear regression equation (y=0.169x + 53.823) that could be used to predict chloride contamination (y) in mg/l from the percentage sanitary risk scores (x) was generated by the model (Table 5 and Figure 17c) where 0.169 and 53.823 were the gradient of the line and the y intercept respectively.

Figure 18a: Residuals' plot generated from the regression of sulphates (mg/l) and sanitary risk scores (%) at sample wells

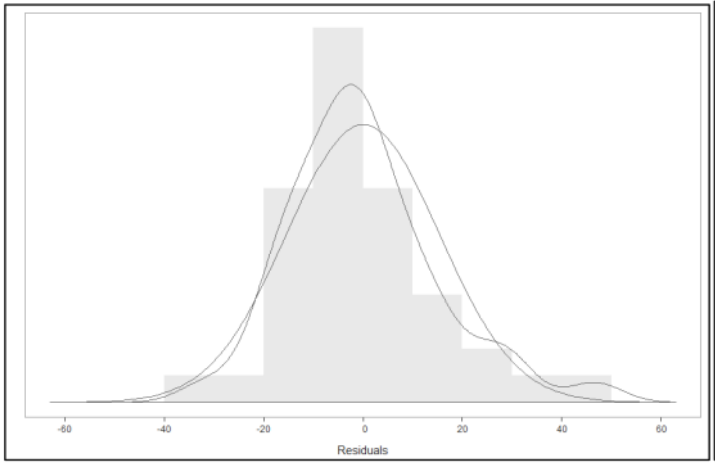


Figure 18b: Fitted values against residuals plot generated from the regression of sulphates (mg/l) and sanitary risk scores (%) at sample wells

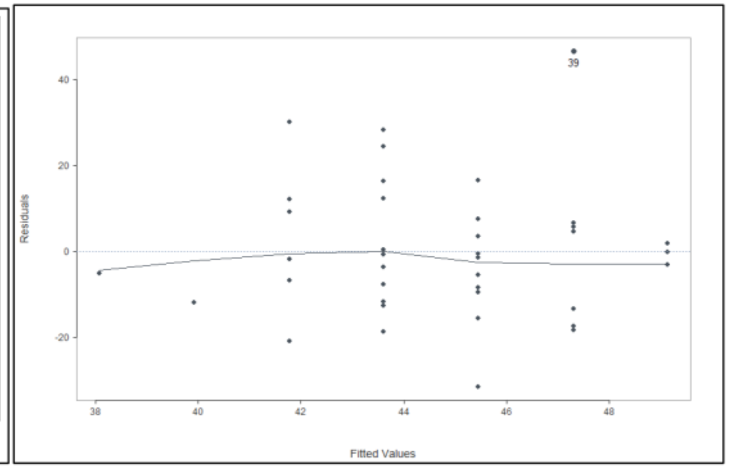
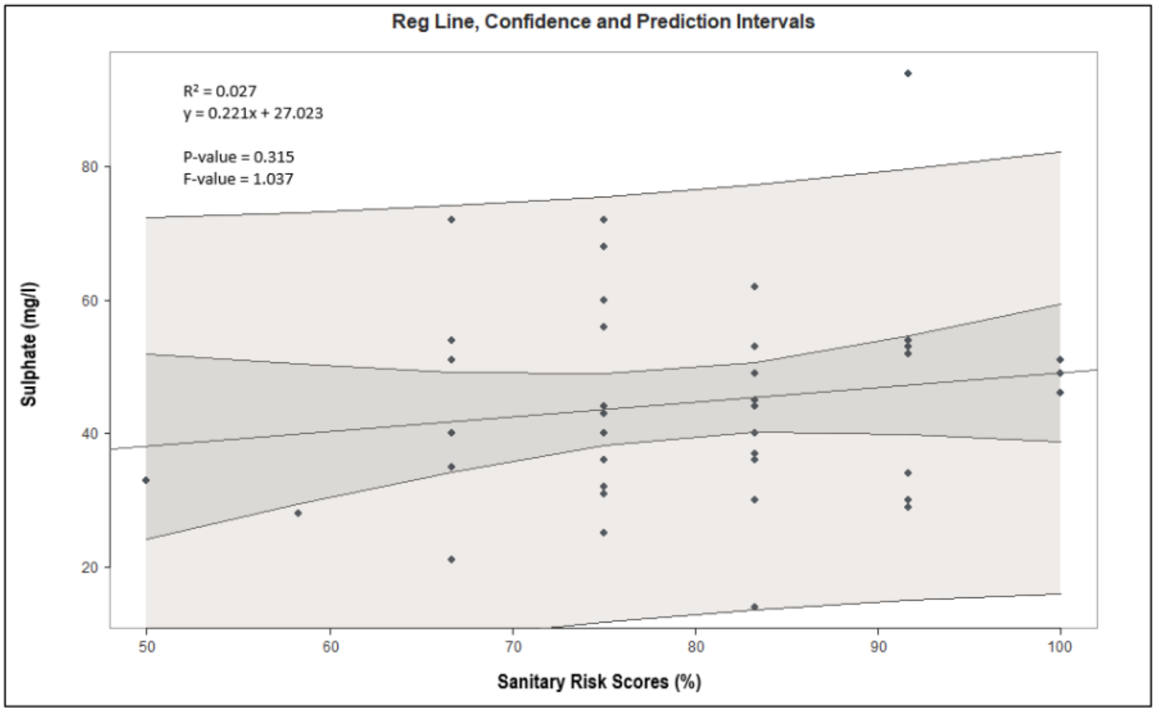


Table 11: Residuals and influence table generated from simple linear regression modelling of sulphates (mg/l) and sanitary risk scores (%)

Data Row	SanRiskScore	Sulphate	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
39	91.7	94	47.3	46.7	3.518	0.845	0.275
18	66.7	72	41.772	30.228	2.086	0.509	0.119
36	83.3	14	45.442	-31.442	-2.144	-0.364	0.06
31	66.7	21	41.772	-20.772	-1.391	-0.34	0.056
37	75	72	43.607	28.393	1.915	0.329	0.051
23	91.7	29	47.3	-18.3	-1.217	-0.292	0.042
28	58.3	28	39.914	-11.914	-0.809	-0.286	0.041
21	91.7	30	47.3	-17.3	-1.148	-0.276	0.038
22	75	68	43.607	24.393	1.624	0.279	0.037
20	91.7	34	47.3	-13.3	-0.876	-0.211	0.022
33	75	25	43.607	-18.607	-1.221	-0.21	0.022
24	66.7	54	41.772	12.228	0.805	0.197	0.02
19	75	60	43.607	16.393	1.071	0.184	0.017
14	83.3	62	45.442	16.558	1.082	0.184	0.017
1	50	33	38.079	-5.079	-0.359	-0.176	0.016
34	83.3	30	45.442	-15.442	-1.007	-0.171	0.015
11	66.7	51	41.772	9.228	0.605	0.148	0.011
35	75	31	43.607	-12.607	-0.818	-0.141	0.01
6	75	56	43.607	12.393	0.804	0.138	0.01
32	75	32	43.607	-11.607	-0.752	-0.129	0.008

Figure 18c: Regression of sulphates (mg/l) and sanitary risk scores (%) at sample wells



3.3.5 Sulphate

Simple linear regression modelling with RStudio was undertaken using the percentage sanitary risk scores and sulphate levels as the predictor and response variables respectively (Figure 18c). Following the sample data's high cook's distance (D=0.275), there was an outlier at data row 39, according to Table 11. Other potential outliers, using the same reference table, were those at data rows 18, 36, 31, and 37, which had D values of 0.119, 0.06, 0.056, and 0.051, respectively, which differed significantly from the rest of the samples. In reference to Table 9, none of the 40 samples exceeded the WHO guideline value of 250mg/l for sulphate (*WHO*, 2006). A weak and positive relationship was produced by the model between the percentage sanitary risk scores and sulphate levels (r^2=0.027) as referenced by Figure 18c. The relationship happened not be statistically significant (p=0.315) implying a failure in rejecting the null hypothesis. According to Ezeh *et al.*, (2021), there was a statistically significant relationship (p=0.01) where the sulphate levels reduced with the increase in the distance from the sanitary pits. According to Table 5 and Figure 18c, the model created a simple linear regression equation (y = 0.221x + 27.023) that could be used to forecast sulphate contamination levels (y) from the percentage sanitary risk scores (x), with the gradient and y intercept of the regression equation being 0.221 and 27.023, respectively.

Figure 19a: Residuals' plot generated from the regression of fluorides (mg/l) and sanitary risk scores (%) at sample wells

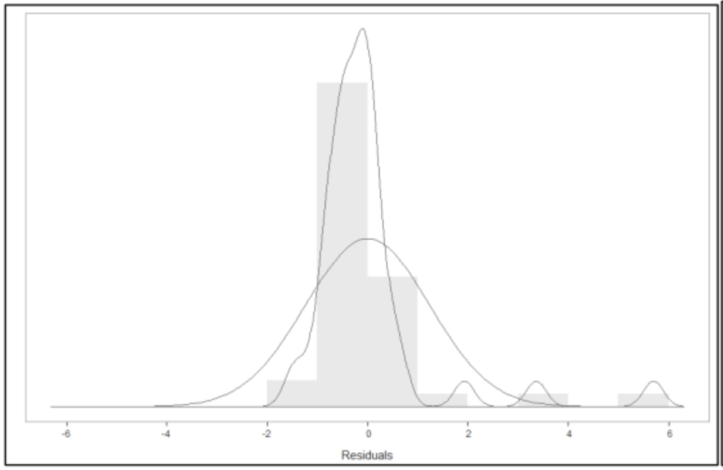


Figure 19b: Fitted values against residuals plot generated from the regression of fluorides (mg/l) and sanitary risk scores (%) at sample wells

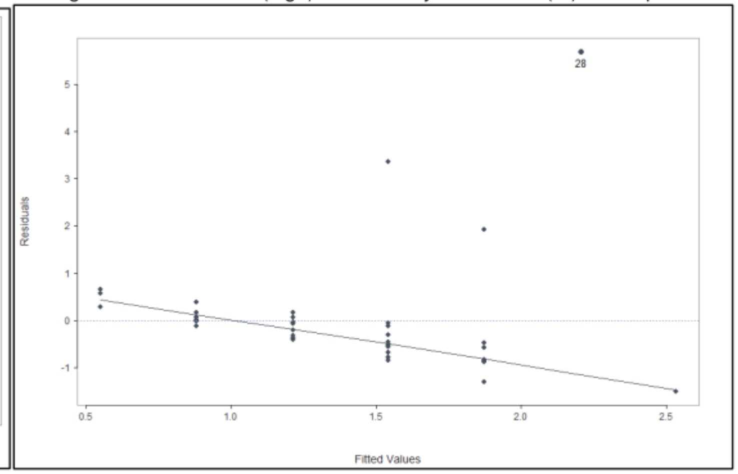
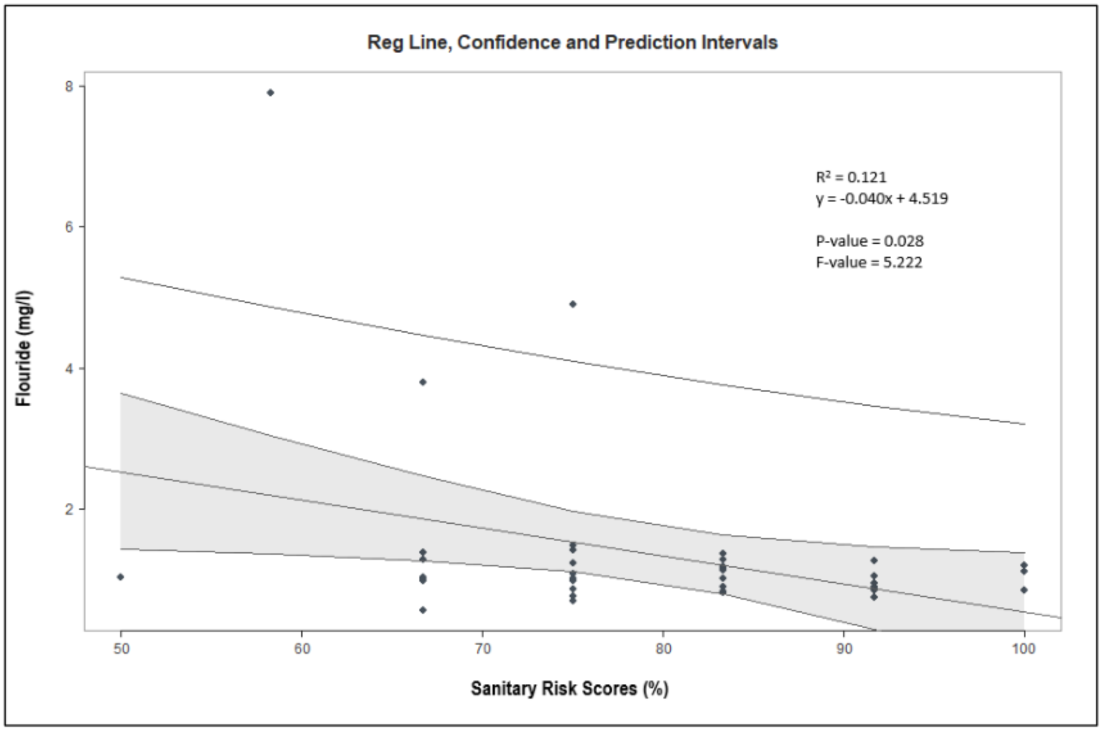


Table 12: Residuals and influence table generated from simple linear regression modelling of fluorides (mg/l) and sanitary risk scores (%)

Data Row	SanRiskScore	Flouride	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
28	58.3	7.9	2.205	5.695	7.766	2.751	1.478
1	50	1.03	2.535	-1.505	-1.36	-0.665	0.216
37	75	4.9	1.542	3.358	3.014	0.518	0.111
38	66.7	3.8	1.872	1.928	1.629	0.398	0.076
31	66.7	0.57	1.872	-1.302	-1.078	-0.263	0.035
40	100	1.2	0.55	0.65	0.548	0.191	0.018
11	66.7	0.98	1.872	-0.892	-0.733	-0.179	0.016
7	66.7	1.01	1.872	-0.862	-0.708	-0.173	0.015
16	100	1.12	0.55	0.57	0.48	0.167	0.014
5	66.7	1.04	1.872	-0.832	-0.683	-0.167	0.014
33	75	0.69	1.542	-0.852	-0.69	-0.119	0.007
24	66.7	1.29	1.872	-0.582	-0.476	-0.116	0.007
32	75	0.76	1.542	-0.782	-0.632	-0.109	0.006
18	66.7	1.39	1.872	-0.482	-0.394	-0.096	0.005
35	75	0.86	1.542	-0.682	-0.551	-0.095	0.005
30	100	0.84	0.55	0.29	0.244	0.085	0.004
6	75	0.98	1.542	-0.562	-0.453	-0.078	0.003
12	75	0.98	1.542	-0.562	-0.453	-0.078	0.003
39	91.7	1.27	0.879	0.391	0.319	0.077	0.003
17	75	1.02	1.542	-0.522	-0.421	-0.072	0.003

Figure 19c: Regression of fluorides (mg/l) and sanitary risk scores (%) at sample wells



3.3.6 Fluoride

The percentage sanitary risk scores and fluoride levels were used as predictor and response variables, respectively, in simple linear regression modelling with RStudio (Figure 19c). Data rows 28 and 1 (Table 12) were identified as outliers because they had the highest cook's distances of 1.478 and 0.216, respectively. Following the large departure of its D value compared to the rest of the samples, data row 37 with a cook's distance of 0.111 may be considered another likely outlier (Table 12). Only three samples out of 40 surpassed the WHO fluoride guideline value of 1.5 mg/l, according to Table 9 (*WHO*, 2006). In the study conducted by Wright *et al.*, (2013), all the fluoride samples exceeded the WHO guideline of 1.5mg/l. According to Figure 19c, the moderate inverse relationship between sanitary risk scores and the fluoride levels (r^2=0.121) was statistically significant (p=0.028; f=5.222). The null hypothesis was therefore rejected in this model scenario. Fluoride contamination levels (y) could be predicted from percentage sanitary risk scores (x) using the linear regression equation (y= -0.040x + 4.419) shown in Figure 19c and Table 5, where -0.040 and 4.419 were the gradient and y intercept, respectively.

Figure 20a: Residuals' plot generated from the regression of phosphates (mg/l) and sanitary risk scores (%) at sample wells

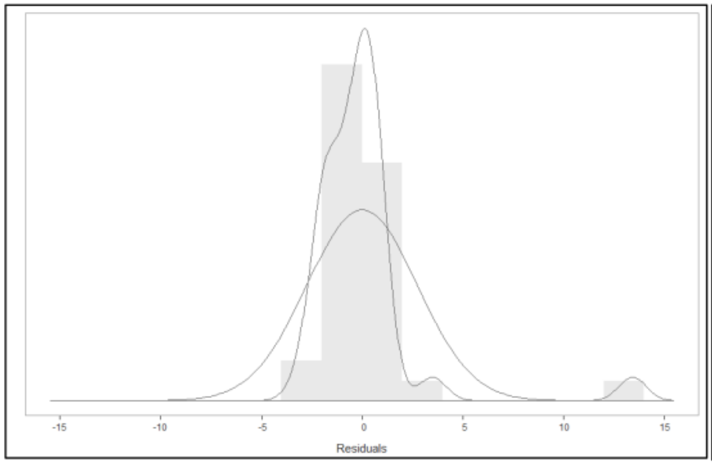


Figure 20b: Fitted values against residuals plot generated from the regression of phosphates (mg/l) and sanitary risk scores (%) at sample wells

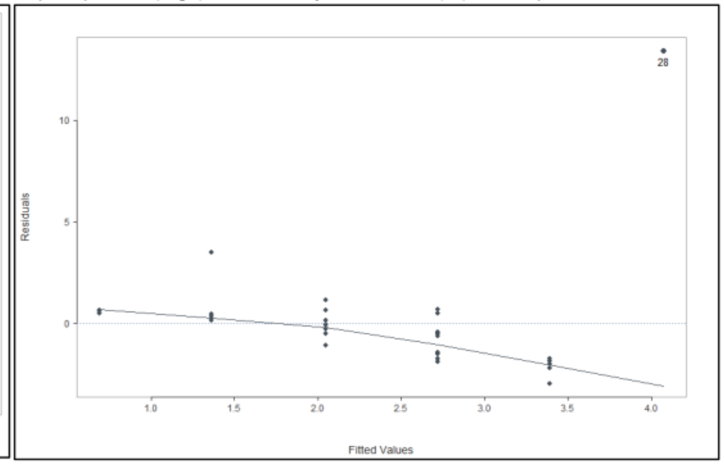
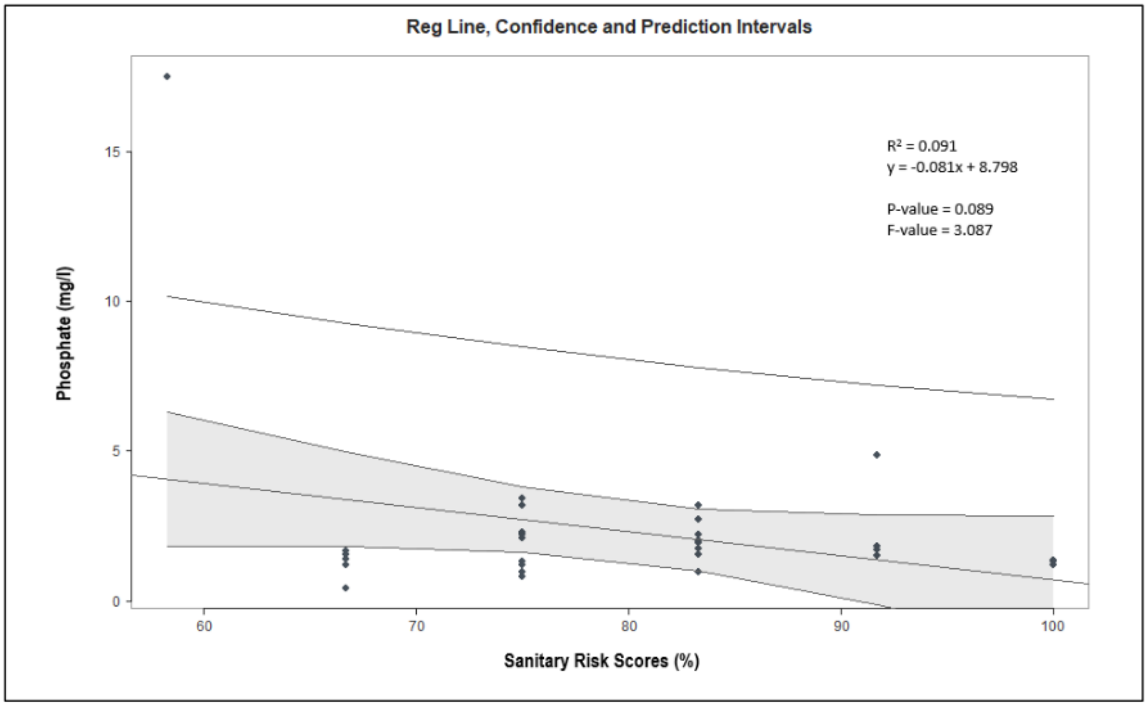


Table 13: Residuals and influence table generated from simple linear regression modelling of phosphates (mg/l) and sanitary risk scores (%)

Data Row	SanRiskScore	Phosphate	Fitted	Residual	Studentized Residual	Dffits	Cook's Distance
28	58.3	17.5	4.072	13.428	16.313	7.073	2.618
23	91.7	4.86	1.365	3.495	1.322	0.361	0.064
31	66.7	0.43	3.391	-2.961	-1.117	-0.324	0.052
5	66.7	1.2	3.391	-2.191	-0.819	-0.238	0.029
18	66.7	1.4	3.391	-1.991	-0.742	-0.216	0.024
38	66.7	1.4	3.391	-1.991	-0.742	-0.216	0.024
7	66.7	1.55	3.391	-1.841	-0.686	-0.199	0.02
11	66.7	1.65	3.391	-1.741	-0.648	-0.188	0.018
33	75	0.82	2.719	-1.899	-0.692	-0.135	0.009
17	75	0.96	2.719	-1.759	-0.64	-0.125	0.008
6	75	1.2	2.719	-1.519	-0.552	-0.108	0.006
40	100	1.35	0.692	0.658	0.252	0.103	0.005
35	75	1.3	2.719	-1.419	-0.515	-0.1	0.005
16	100	1.3	0.692	0.608	0.233	0.095	0.005
30	100	1.2	0.692	0.508	0.195	0.08	0.003
3	83.3	3.2	2.046	1.154	0.418	0.078	0.003
27	83.3	0.96	2.046	-1.086	-0.393	-0.073	0.003
19	75	3.4	2.719	0.681	0.247	0.048	0.001
8	83.3	2.7	2.046	0.654	0.236	0.044	0.001
36	83.3	2.7	2.046	0.654	0.236	0.044	0.001

Figure 20c: Regression of phosphates (mg/l) and sanitary risk scores (%) at sample wells



3.3.7 Phosphate

For this scenario, a simple linear regression model with RStudio was run with the percentage sanitary risk scores and phosphate levels as the predictor and response variables respectively (Figure 20c). Table 13 shows that the model identified outliers at data rows 28, 23, and 31 because they had the highest cook's distances (D) of 2.618, 0.064, and 0.052, respectively, when compared to the remainder of the data samples. Table 9 shows that none of the 40 samples tested exceeded the Kenyan Water Services Regulatory Board's phosphate guideline value of 30 mg/l (WASREB, 2008). According to Figure 20c, the inverse moderate relationship between sanitary risk scores and the phosphate levels (r^2=0.091) was not statistically significant (p=0.089). A linear regression equation (y=-0.081x + 8.798) that was generated by the model (Figure 20c and Table 5) could be used to predict the phosphate contamination levels (y) from the percentage sanitary risk scores (x) where -0.081 and 8.798 were the gradient and y intercept respectively.

4 Discussion

This section presents the discussions of the findings in relation to the aim and objectives of this research project. This study focused on assessing the relationship between sanitary risk inspection and shallow groundwater contamination in the peri-urban areas of Kisumu, Kenya.

The significance of this work stems from the rising urbanisation of Sub-Saharan Africa, as well as the exposure of this large population to potentially contaminated water from shallow wells (Okotto-Okotto *et al*., 2015). Secondary data analysis revealed that the average percentage of sanitary risk scores (Table 1) was high (79.38%), with the risk factors of scattered garbage within 30 metres of the well and unsanitary well covers posing the highest risk (95%) of polluting shallow ground water wells (Figure 5). The % sanitary risk scores had no statistically significant relationship with five (TTC, Nitrate, Chloride, Sulphate, and Phosphate) of the six contamination parameters ($p>0.05$) according to multivariate statistical analysis using simple linear regression modelling with R in RStudio. The only model scenario which exhibited a statistically significant relationship was the one between the percentage sanitary risk scores and fluoride ($p<0.05$). The regression models produced some outliers, which appeared to increase data variability and hence reduce statistical power, but this study did not attempt to eliminate any of the outliers in any model scenario. This is because they have the potential to reflect cases that may have been under-sampled during data collection. The study also discovered that all TTC counts were higher than the WHO's recommended levels (Table 9). All chemical parameters, on the other hand, were below the WHO guidelines values, with the exception of fluoride, where 7.5 percent of the samples exceeded the fluoride guideline threshold (Table 9). Nonetheless, the model scenarios that failed to reject the null hypothesis (p0.05) showed positive weak connections ($r^2<0.09$) with the exception of the relationship between % sanitary risk scores and phosphate, which had an inverse moderate relationship ($r^2=0.091$). The only model scenario that rejected the null hypothesis ($p=0.028$) was a moderate inverse relationship ($r^2=0.121$) between percentage sanitary risk scores and fluoride levels.

Following the high TTC counts beyond the guideline value, the findings suggest that medium and high risk of latrines proximity (Figure 5) to the wells directly influenced microbial groundwater contamination. This is owing to the fact that the majority of people in the study area utilise pit latrines (Manyatta: 91 percent and Migosi: 38 percent), with open defecation being performed by less than 1% of the population (Okotto-Okotto *et al*., 2015). This faecal matter subsequently infiltrates the aquifer system, causing measurable TTCs to be present in the groundwater. During the wet season, the mobility of these contamination hazards is likely to rise. The high fluoride contents in three of the forty samples could be attributed to geogenic causes, such as rock–water interactions with fluorine-bearing minerals found in a few porphyritic granites (RDWSSP, 1987) in some places of the study region.

The acceptable concentration of rest of the chemical parameters (nitrate, chloride, sulphate, phosphate and 92.5% of fluoride samples) based on guideline values could have been affected by confounding factors. These confounding factors as substantiated below have been suggested to affect all the relationships that this project was striving to establish.

According to the findings of the study, there could have been some unmeasured factors that caused the substantial variations in the results. The study suggests as advised by Kelly *et al.*, (2021), that the aforementioned weak (small r^2) and non statistically significant relationships could have been caused by 1) single measurements that do not accurately represent water quality of a given water system over time; 2) the current WHO sanitary inspection forms that do not adequately identify all pertinent sources, carriers, and barrier breakdowns affecting water systems; 3) the use of models which do not weight risk factors misrepresents associations in which a small number of important factors dominate the risk (Kelly *et al.*, 2021). Microbial contamination has for instance been reported by Stukel *et al.*, (1990) to radically vary over brief time periods but also affected by seasonal weather patterns and recent rainfall. In light of the aforementioned, the findings of this study could have been affected by measurement errors as there was a lack of replicability in sanitary risk inspections (Okotto-Okotto *et al.*, 2015) which reduce statistical power to detect changes in risk scores. Furthermore, the fact that samples for water quality tests had to be transported from sample collection points could also possibly affect the quality of the test results as shallow well water happens to be prone to transient contamination events (Godfrey *et al.*, 2005). The likes of water system age (Ercumen *et al.*, 2017), topography (Engström *et al.*, 2015) and population density (Howard *et al.*, 2003) are also not incorporated in the WHO sanitary risk inspection forms yet identified as determinants of water system contamination. Seasonality was not taken into account in this study, according to Okotto-Okotto *et al.*, (2014)), who state that the samples were collected between late February and early March of 2014, which coincided with the start of the rainy season in the study area (WWCI, 2021). This was also a constraint because research suggests that extreme seasonal occurrences such as rainfall have an impact on contamination magnitude (Godfrey *et al.*, 2005). As if that weren't enough, the study's spatial analysis was unfeasible due to a lack of georeferenced data for the sample wells. Regarding the WHO sanitary inspection forms not weighting the risk factors much as some factors may be more influential than others, the presence of cracks in the cement floor could for instance be more or less strongly connected with contamination than an unsanitary well cover. Setting, season, and/or construction standards may all have a role in the relative relevance of distinct risk factors fluctuating. One significant barrier collapse may, in some ways, trump all other risk factors in that group. As a result, all sanitary risk variables need probably be noted and addressed in order to regard the present sanitary inspection approach as a more appropriate risk assessment tool (Kelly *et al.*, 2021).

In the SSA, where water quality monitoring with the frequency recommended by (Peletz *et al*., 2016) is not conducted, it is very important to prioritize sanitary risk inspections over water quality analysis in a bid to identify and prioritize repairs to water systems. This could include sealing fractures and leaks, repairing or replacing damaged or missing components, and operators should be trained to do basic remedial activities or seek outside assistance for more extensive restoration (Kelly *et al*., 2021). In terms of policy, sanitary inspection forms should include more evidence-based risk transporters and sources, as well as access to open data that might be included as and when sanitary inspection data is collected (Kelly., *et al* 2021). In a bid to improve the current sanitary inspection tools and forms, larger and richer datasets should be developed linking sanitary risk inspection data with water quality, health outcomes and meteorological data (Kelly *et al*., 2021). Such endeavours may include the investigation of the relevance of additional risk factors like well depth, soil characteristics among others in sanitary risk inspections. In a way, the continuous collection and compilation data on water quality, sanitary risks and the relevant outcomes coupled with continuous explorations of the relationships between water system vulnerability and contamination will sustainably inform the concerned stakeholders on how to address the issues of safety for drinking water supply.

5 Conclusion

This section presents the conclusions of the study based on the findings and discussions in relation to the aim and objectives of this research project.

The study's limitations preclude it from drawing firm conclusions about the relationship between sanitary risk variables and contamination measures. The study does, however, show that sanitary risk factors play a substantial role in microbial contamination of shallow wells, as all samples were above the WHO guideline range for detectable TTC in drinking water. The study, on the other hand, reveals that sanitary risk factors have a limited contribution towards chemical contamination because the concentrations of all chemical parameters were below the standard limits, with the exception of fluoride, where 7.5 percent of samples were over the guideline values. The results also show that there was no relationship between sanitary risk scores and TTC, nitrate, chloride, sulphate, or phosphate, since the p values in all five model scenarios were more than 0.05. However, a statistically significant negative inverse association between sanitary risk scores and fluoride ($p<0.05$) was discovered in the study. Despite the fact that statistical power was low in five model scenarios, the study found that multivariate statistical analysis is likely to produce model scenarios in which TTC counts, nitrate, chloride, and sulphate levels increase in statistically significant relationships with the percentage of sanitary risk scores.

To address the study's limitations, the scholar recommends that: 1) during the data collection process, frequent and multiple measurements, as well as sanitary inspections, be conducted; 2) the current WHO sanitary inspection forms be revised to adequately identify all relevant sources, carriers, and barrier breakdowns contributing to groundwater contamination; and 3) risk factors be weighted using appropriate models for better representation of the contribution of individual factors towards groundwater contamination.

It is therefore imperative to incorporate more appropriate datasets as expounded in the discussion section while establishing the relationship between sanitary risk inspection and the contamination of shallow groundwater wells in the peri-urban areas of the sub-Saharan Africa and elsewhere.

6 Acknowledgements

I wish to express my sincere gratitude towards Dr Clare Robinson for her supervision and guidance as well as for providing me with a sound basis to exercise an independent thought process and execution of tasks. Many thanks also to Dr Andrew Lowe who held the fort during the periods when Clare was not available. The phenomenal job that Andrew did during the tutorials is much appreciated. I am forever indebted to Clare and Andrew for their support without which I would possibly not have made it this far. May the Most High's blessings be upon you!

7 References

Andriesse, W.. & Pouw, B.J.A. van der (1985) *Reconnaissance soil map of the lake basin development authority area Western Kenya, scale 1 : 250.000, legend and rating tables*. [Online]. Available from: https://library.wur.nl/WebQuery/edepot/505358.

Ayse Ercumen, Abu Mohd Naser, Benjamin F. Arnold, Leanne Unicomb, John M. Colford Jr., and S.P.L. (2017) Can Sanitary Inspection Surveys Predict Risk of Microbiological Contamination of GroundwaterSources.pdf. *The American Society of Tropical Medicine and Hygiene*. [Online] Available from: doi:10.4269/ajtmh.16-0489.

Bain, R., Cronk, R., Hossain, R., Bonjour, S., et al. (2014) Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Tropical Medicine and International Health*. [Online] 19 (8), 917–927. Available from: doi:10.1111/tmi.12334.

Barrett, M., Howard, G., Pedley, S., Taylor, R., et al. (2000) A comparison of the extent and impacts of sewage contamination on urban groundwater in developed and developing countries. *Schriftenreihe des Vereins fur Wasser-, Boden- und Lufthygiene*. [Online] 105, 179–185. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10842812.

Cronin, A.A., Breslin, N., Gibson, J. & Pedley, S. (2006) Monitoring source and domestic water quality in parallel with sanitary risk identification in Northern Mozambique to prioritise protection interventions. *Journal of Water and Health*. [Online] 4 (3), 333–345. Available from: doi:10.2166/wh.2006.029.

Cronin, A.A., Pedley, S., Hoadley, A.W., Kouonto Komou, F., et al. (2007) Urbanisation effects on groundwater chemical quality: findings focusing on the nitrate problem from 2 African cities reliant on on-site sanitation. *Journal of water and health*. [Online] 5 (3), 441–454. Available from: doi:10.2166/wh.2007.040.

Emma Kelly, Ryan Cronk, M.F. and J.B. (2021) Sanitary inspection, microbial water quality analysis,.pdf. *npj Clean Water*. [Online] Available from: doi:10.1038/s41545-020-00093-z.

Engström, E., Balfors, B., Mörtberg, U., Thunvik, R., et al. (2015) Prevalence of microbiological contaminants in groundwater sources and risk factor assessment in Juba, South Sudan. *Science of The Total Environment*. [Online] 515–516, 181–187. Available from: doi:10.1016/j.scitotenv.2015.02.023.

Ezeh, E., Akagha, C.I., Okeke, O., Aniobi, C.C., et al. (2021) *Effect of Distance of Sanitary Pits on the Physicochemical Properties of Hand Dug Well Water Samples Consumed by People Living in Akwuke Community , Enugu South Local Government Area of Enugu State*. [Online] 340–349. Available from: doi:10.4236/jwarp.2021.135021.

Godfrey, S., Timo, F. & Smith, M. (2005) Relationship between rainfall and microbiological contamination of shallow groundwater in Northern Mozambique. *Water SA*. [Online] 31 (4), 609–614. Available from: doi:10.4314/wsa.v31i4.5152.

Government of Kenya (2010) *The 2009 Kenya Population and Housing Census - Population and Household Distribution by Socio-economic Characteristics*. [Online] II, 451. Available from: https://s3-eu-west-1.amazonaws.com/s3.sourceafrica.net/documents/21195/Census-2009.pdf.

Grönwall, J.T., Mulenga, M. & McGranahan, G. (2010) *Groundwater, self-supply and poor urban dwellers A review with case studies of Bangalore and Lusaka*. [Online]. Available from: https://pubs.iied.org/sites/default/files/pdfs/migrate/10584IIED.pdf.

Havelaar, A., Blumenthal, U.J., Strauss, M., Kay, D., et al. (2001) Guidelines: the current position. *Water Quality: Guidelines, Standards and Health*. [Online] 18–42. Available from: https://www.who.int/water_sanitation_health/dwq/iwachap2.pdf?ua=1.

Howard, G., Pedley, S., Barrett, M., Nalubega, M., et al. (2003) Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda. *Water Research*. [Online] 37 (14), 3421–3429. Available from: doi:10.1016/S0043-1354(03)00235-5.

W.W.& C. (2021) *World Weather & Climate Information*. [Online]. 2021. Available from: https://weather-and-climate.com/average-monthly-precipitation-Rainfall,kisumu-ke,Kenya [Accessed: 13 August 2021].

Kumpel, E., Peletz, R., Bonham, M. & Khush, R. (2016) Assessing Drinking Water Quality and Water Safety Management in Sub-Saharan Africa Using Regulated Monitoring Data. *Environmental Science and Technology*. [Online] 50 (20), 10869–10876. Available from: doi:10.1021/acs.est.6b02707.

Lapworth, D.J., Macdonald, A.M., Tijani, M.N., Darling, W.G., et al. (2013) *Residence times of shallow groundwater in West Africa : implications for hydrogeology and resilience to future changes in climate*. [Online] 673–686. Available from: doi:10.1007/s10040-012-0925-4.

Lewis, W.J., Stephen S.D. Foster & S., B.D. (1984) Risk of Groundwater Pollution Through on-Site Sanitation in Developing Countries. *Journal of the Institution of Engineers (India): Environmental Engineering Division*. [Online]. 65 (01). Available from: https://www.ircwash.org/sites/default/files/244-82RI-15490.pdf.

Lloyd, B. & Suyati, S. (1989) A pilot rural water surveillance project in Indonesia. *Waterlines*. [Online] 7 (3), 10–13. Available from: doi:10.3362/0262-8104.1989.004.

Lloyd, B.J. & Bartram, J.K. (1991) Surveillance Solutions to Microbiological Problems in Water Quality Control in Developing Countries. *Water Science and Technology*. [Online] 24 (2), 61–75. Available from: doi:10.2166/wst.1991.0031.

Luby, S.P., Gupta, S.K., Sheikh, M.A., Johnston, R.B., et al. (2008) Tubewell water quality and predictors of contamination in three flood-prone areas in Bangladesh. *Journal of Applied Microbiology*. [Online] 105 (4), 1002–1008. Available from: doi:10.1111/j.1365-2672.2008.03826.x.

MacDonald, A.M., Bonsor, H.C. & Taylor, B.E.O.D.´ and R.G. (2012) Quantitative maps of groundwater resources in Africa. *Environmental Research Letters*. [Online] Available from: doi:10.1088/1748-9326/7/2/024009.

MacDonald AM, Bonsor HC, Calow RC, Taylor RG, Lapworth DJ, Maurice L, Tucker J, O.D.B. (2011) *Groundwater resilience to climate change in Africa. British Geological Survey Open Report*. [Online]. Available from: doi:Open Report OR/11/031.

Menard, S. (1996) *Applied Logistic Regression Analysis*. [Online]. Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality. Available from: doi:10.4135/9781412983433.

Minitab (2020) *Natural log (log base e) function - Minitab*. [Online]. 2020. Available from: https://support.minitab.com/en-us/minitab/19/help-and-how-to/calculations-data-generation-and-matrices/calculator/calculator-functions/logarithm-calculator-functions/natural-log-log-base-e-function/ [Accessed: 11 August 2021].

Misati, A.G., Ogendi, G., Peletz, R., Khush, R., et al. (2017) Can sanitary surveys replace water quality testing? Evidence from Kisii, Kenya. *International Journal of Environmental Research and Public Health*. [Online] 14 (2). Available from: doi:10.3390/ijerph14020152.

Mondal, D., Ganguli, B., Sen Roy, S., Halder, B., et al. (2014) Diarrhoeal health risks attributable to water-borne-pathogens in arsenic-mitigated drinking water in west Bengal are largely independent of the microbiological quality of the supplied water. *Water (Switzerland).* [Online] 6 (5), 1100–1117. Available from: doi:10.3390/w6051100.

O'Farrell, M.A.P. and P.N. (1970) The Assumptions of the Linear Regression Model. *Transactions of the Institute of British Geographers*. [Online] Available from: doi:10.2307/621706.

Okotto-Okotto, J.; Pedley, S.; Okotto, L.G.; Price, H.; Wright, J. (2014) *Sanitary risk inspections of shallow wells, boreholes and springs in Kisumu, Kenya*. [Online]. 2014. NERC Environmental Information Data Centre. (Dataset). Available from: https://doi.org/10.5285/bc1a979b-7cc9-4c9d-9fe4-a8510cd62f8e.

Okotto-Okotto, J., Okotto, L., Price, H., Pedley, S., et al. (2015) A longitudinal study of long-term change in contamination hazards and shallow well quality in two neighbourhoods of Kisumu, Kenya. *International Journal of Environmental Research and Public Health*. [Online] 12 (4), 4275–4291. Available from: doi:10.3390/ijerph120404275.

Okotto, L., Okotto-Okotto, J., Price, H., Pedley, S., et al. (2015) Socio-economic aspects of domestic groundwater consumption, vending and use in Kisumu, Kenya. *Applied Geography*. [Online] 58, 189–197. Available from: doi:10.1016/j.apgeog.2015.02.009.

Omara, T., Nassazi, W., Adokorach, M. & Kagoya, S. (2019) Physicochemical and Microbiological Quality of Springs in Kyambogo University Propinquity. *Open Access Library Journal*. [Online] 6, 1–13. Available from: doi:10.4236/oalib.1105100.

Patrick, J.M., Murcott, S. & Punsalan, J. (2011) Coupling microbiological testing and sanitary surveys in drinking water quality programs: results from Capiz Province, Philippines. *Journal of Water, Sanitation and Hygiene for Development*. [Online] 1 (2), 124–135. Available from: doi:10.2166/washdev.2011.033.

Pedley, S.; Okotto-Okotto, J.; Okotto, L.; Price, H.; Wright, J. (2014) (2014) *Groundwater microbiology and chemistry of shallow wells, boreholes and springs in Kisumu, Kenya in 2014*. [Online]. 2014. NERC Environmental Information Data Centre. (Dataset). Available from: https://doi.org/10.5285/4062e6d9-2e90-4775-87f1-179dea283ef1.

Pedley, S. & Howard, G. (1997) The public health implications of microbiological contamination of groundwater. *Quarterly Journal of Engineering Geology*. [Online] 30 (2), 179–188. Available from: doi:10.1144/GSL.QJEGH.1997.030.P2.10.

Peletz, R., Kumpel, E., Bonham, M., Rahman, Z., et al. (2016) To what extent is drinking water tested in sub-saharan Africa? A comparative analysis of regulated water quality monitoring. *International Journal of Environmental Research and Public Health*. [Online] 13 (3). Available from: doi:10.3390/ijerph13030275.

Project, KOD. (2016) *Kenya Open Data*. [Online]. 2016. Available from: https://www.opendata.go.ke/datasets/abba736b3287484aac696e9766b6db36/explore [Accessed: 25 June 2021].

RDWSSP (1987) *Rural Domestic Water Resources Assessment Kisumu District —Water Resources Survey and Training Programme: Report on Nyakach. Division; DHV Consulting Engineers*. [Online]. Available from: https://www.ircwash.org/sites/default/files/824-5816.pdf.

Robinson, A., Cook, R.D. & Weisberg, S. (1984) Residuals and Influence in Regression. *Journal of the Royal Statistical Society. Series A (General)*. [Online] 147 (1), 108. Available from: doi:10.2307/2981746.

Shamim, A., Basant, M. & Sabiiti Elly, B.L.N.& B.B. (2019) Status of Waste Management in the East African Cities : Understanding the Drivers of Waste Generation , Collection and Disposal and Their Impacts on Kampala City ' s Sustainability. *MDPI*. [Online] 1–16. Available from: doi:10.3390/su11195523.

Sibanda, L.K., Obange, N. & Awuor, F.O. (2017) Challenges of Solid Waste Management in Kisumu, Kenya. *Urban Forum*. [Online] 28 (4), 387–402. Available from: doi:10.1007/s12132-017-9316-1.

Snoad, C., Nagel, C., Bhattacharya, A. & Thomas, E. (2017) The Effectiveness of Sanitary Inspections as a Risk Assessment Tool for Thermotolerant Coliform Bacteria Contamination of Rural Drinking Water: A Review of Data from West Bengal, India. *The American journal of tropical medicine and hygiene*. [Online] 96 (4), 976–983. Available from: doi:10.4269/ajtmh.16-0322.

Sorensen, J.P.R., Lapworth, D.J., Nkhuwa, D.C.W., Stuart, M.E., et al. (2015) Emerging contaminants in urban groundwater sources in Africa. *Water Research*. [Online] 72, 51–63. Available from: doi:10.1016/j.watres.2014.08.002.

Stukel, T.A., Greenberg, E.R., Dain, B.J., Reed, F.C., et al. (1990) A Longitudinal Study of Rainfall and Coliform Contamination in Small Community Drinking Water Supplies. *Environmental Science and Technology*. [Online] 24 (4), 571–575. Available from: doi:10.1021/es00074a610.

Tranmer, M., Murphy, J., Elliot, M. & Pampaka, M. (2020) Multiple Linear Regression (2nd Edition). *Cathie Marsh Institute Workin Paper*. [Online] (01), 59. Available from: https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/2020-1-multiple-linear-regression.pdf.

UN-Habitat (2014) *The State of African Cities, 2014, Re-imagining sustainable urban transitions*. [Online]. Available from: https://unhabitat.org/sites/default/files/download-manager-files/State of African Cities 2014.pdf.

UN (2015) *United Nations, Department of Economic and Social Affairs, Population Division, World Urbanization Prospects: The 2014 Revision*. [Online]. Available from: https://population.un.org/wup/Publications/Files/WUP2014-Report.pdf.

United Nations (2016) The human rights to safe drinking water and sanitation. *Treaty Series*. [Online] 660 (9464), 27531. Available from: https://digitallibrary.un.org/record/821067/files/A_RES_70_169-EN.pdf.

United Nations Human Settlements Programme (UN-HABITAT) (2005) *Situation Analysis of Informal Settlements in Kisumu Kenya Slum*. [Online]. Available from: https://unhabitat.org/sites/default/files/download-manager-files/Situation Analysis of informal settlements in Kisumu.pdf.

WASREB (2008) *Guidelines on Water Quality and Effluent Monitoring*. [Online]. WASREB. Available from: https://wasreb.go.ke/downloads/Water_Quality_&_Effluent_Monitoring_Guidelines.pdf.

WHO/UNICEF (2017) Progress on Drinking Water, Sanitation and Hygiene - Joint Monitoring Programme 2017 Update and SDG Baselines. *World Health Organization*. [Online] 66. Available from:http://apps.who.int/bookorders.%0Ahttp://apps.who.int/iris/bitstream/10665/258617/1/9789241512893-eng.pdf%0Ahttp://www.wipo.int/amc/en/%0Ahttp://www.wipo.int/amc/en/.

WHO (2004) *Guidelines for Drinking-water Quality, 3rd Edition*. Third. [Online]. Available from: https://www.who.int/water_sanitation_health/dwq/GDWQ2004web.pdf.

WHO (1997) *Guidelines for Drinking-water Quality*. second. [Online]. Available from: doi:10.1016/S1462-0758(00)00006-6.

WHO (2006) *WHO guidelines for drinking-water quality*. 3rd edition. [Online]. Available from: https://www.who.int/water_sanitation_health/dwq/gdwq0506.pdf.

Wright, J.A., Cronin, A., Okotto-Okotto, J., Yang, H., et al. (2013) A spatial analysis of pit latrine density and groundwater source contamination. *Environmental Monitoring and Assessment*. [Online] 185 (5), 4261–4272. Available from: doi:10.1007/s10661-012-2866-8.

Wright, R.C. (1986) The seasonality of bacterial quality of water in a tropical developing country (Sierra Leone). *Journal of Hygiene*. [Online] 96 (1), 75–82. Available from: doi:10.1017/S0022172400062550.

Clare Robinson - 10785991 - EART60372 MSc Project

GRADEMARK REPORT

FINAL GRADE

88/100

GENERAL COMMENTS

Instructor

An excellent project. Well-written and imaginatively analysed with a thorough review of secondary data and published literature. Critical analysis of data and literature with limitations of the study recognised.

First marker mark: 85%

I confirm the plagiarism report is acceptable.

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

GRADING FORM: MPEC PROJECT GRADING FORM 2021

BRIAN BUKENYA	**501**

RESEARCH QUESTION

Was the research question clearly stated ? Was the context (including societal importance) of the question clear ? Were the research project aims & objectives clear and appropriate? [cf. ILO EART60372-02]

Excellent aims, objectives and research question. **85**

CONTEXT

Were key theories, concepts and/or knowledge relevant to the specified research question presented in the introduction or literature review in sufficient breadth, detail and clarity ? As appropriate, were (hypothetical or real) data outputs that would be consistent with different relevant theoretical models or hypotheses contrasted & compared ? [cf. ILO EART60372-01]

Excellent context in the Introduction

85

EXPERIMENTAL DESIGN

Were sample, data and data quality requirements, clear ? Was the project design appropriate for the research project aims ? Was the project design justified/explained (e.g. reasons for selecting particular samples, methods or techniques) ? [cf. ILO EART60372-03]

Experimental design was made by the source of the secondary data and this was made clear. Statistical analyses were designed and carried out by the student.

83

DATA ACQUISITION

Does the thesis indicate use of literature review, computer modelling, field and/or laboratory protocols relevant to obtaining data to address the aims ? Were data & other outputs presented in a clear manner ? Were appropriate descriptive statistics used ? Were data quality parameters reported ? [cf. ILO EART60372-06; ILO EART60372-04] If you were also the project supervisor, you are encouraged to give an indication of the level of diligence or skill of the student in acquiring *data (secondary or primary) and in its interpretation. These comments may contribute to the evidence considered in the event of differences between examiner marks requiring a discussion and resolution*

Data acquired relevant to addressing aim, objectives and research question.

80

DATA INTERPRETATION

Were appropriate and justified statistical tests used to interpret the data ? Were the implications & significance of the data interpreted, demonstrating appropriate use of key concepts and arguments ? [cf.ILO EART60372-05, ILO EART60372-07] If you were also the project supervisor, you are encouraged to give an indication of the level of diligence or skill of the student in acquiring data (secondary or primary) and in its interpretation. These comments may contribute to the evidence considered in the event of differences between examiner marks requiring a discussion and resolution

Excellent data interpretation designed by the student. Excellent use of statistics.

85

PROFESSIONAL PRESENTATION

Did the general presentation (including timely submission, format & structure compared to requirements & standards of the University of Manchester, clear tables & figures with useful captions, few proof-reading errors) reflect skills & attributes of those seeking employment in the environment sector ? [cf. ILO EART60372-08] The research project is an exercise in both carrying out and reporting an investigation, consequently clarity and concision in the thesis are an integral part of the project.

Excellent professional presentation.

83