University of the Western Cape

Using geostatistical-hydrogeological approach to develop groundwater monitoring system in South Western Karoo, South Africa

Zaheed Gaffoor

2866321

This thesis is submitted in fulfilment of the requirements for the degree of Magister Scientiae in Environmental and Water Science, in the Department of Earth Sciences, University of the Western Cape.

Supervisor: Dr. Thokozani Kanyerere

Co-supervisor: Dr. Kevin Pietersen

November 2017

Declaration

I declare that "*Using geostatistical-hydrogeological approach to develop groundwater monitoring system in South Western Karoo, South Africa*" is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete reference.

Zaheed Gaffoor
Student number: 2866321

09 November 2017

Signature: ....

Acknowledgements

Many hours were spent reading, writing and analysing, sometimes towards dead-ends. However, with all this effort, this thesis would not have come to fruition without the support of some really great people. Firstly I would like to thank my supervisors, Dr. Kanyerere and Dr. Pietersen, for giving me the opportunity to do my masters and providing me with valuable guidance. Then I would sincerely like to thank the team from Department of Water and Sanitation, especially Mzu, Olwethu and Mike Smart. These individuals helped me so much in the field and provided the needed data. Lastly, to my family, thank you for putting up with me, and for supporting me. I hope these people enjoy the work in thesis, and I wish them best of luck in the future.

Abstract

Groundwater in the South Western Karoo plays a vital role in the overall water supply in the region. However, this resource is vulnerable to impacts from anthropogenic and natural activities. Mitigating the impacts on groundwater quality and quantity depends on the information provided by groundwater monitoring networks. The information provided by groundwater monitoring networks allow for timely and effective intervention to take place before widespread degradation occurs. In recent times, there has been interest in exploiting potentially vast natural resources of shale gas in the South Western Karoo. However, studies have highlighted links between shale gas development and groundwater contamination. There are concerns that these issues of groundwater contamination and overexploitation can occur in the South Western Karoo during shale gas development. One of the key features that need addressing is the lack of a statistically sound baseline that can inform on the natural conditions of the groundwater system, before development of shale gas exploitation.

The current groundwater monitoring system in the region is not adequately designed to capture the required level of information in the context of shale gas development. Hence, this research was aimed at developing a groundwater monitoring system to provide the necessary baseline data and perform detection monitoring during shale gas development. A case study area in the South Western Karoo, South Africa, was chosen to address this research problem. The objectives of the current research were to: 1) design a groundwater monitoring network, 2) establish the natural baseline using current and historical groundwater quality data, and 3) determine monitoring parameters and frequency of monitoring.

In-order to design a regional groundwater monitoring network based on multiple objectives, a geostatistical-hydrogeological approach was applied. This allowed for optimization in terms of density of monitoring network, and takes into account key hydrogeological features of interest when position monitoring points. The analysis revealed the current network, which contained 34 monitoring points, to be irregularly distributed and clustered throughout the case study area. Using kriging techniques a new network density was calculated, where monitoring points were separated by approximately 16.7 km. A systematic sampling approach was applied to a hexagonal sampling grid, which has the potential to reduce the kriging prediction standard error. Using key hydrogeological features such as contaminant transport pathways and water resources zones, 1 new monitoring per grid cell (in most cases) was placed within proximity to these features. In this manner, the groundwater monitoring

network was expanded from 34 to 91 monitoring points. The new network showed a decrease in the kriging prediction standard error compared to the existing network, which suggests a gain in information.

To establish the natural baseline, the historical and current groundwater chemistry data was analyzed using statistical methods. Only 5 groundwater chemistry points exist in the case study area with a sample size large enough to perform the statistical review. The exploratory data analysis revealed that for most analytes the distribution was spread far beyond the mean, with high variability (skewed heavily by the presence of outliers). This would imply a large population of expected values during sampling, which suggests that the aquifer is not in a steady state. All datasets were considered to be spatially and temporally independent. A total of 67.5 % of the analytes had non-normal distribution, while the natural logarithm of the datasets was only marginally better at 63.9 %. Seasonality was not present in the dataset, but more frequent annual sampling is required to accurately determine seasonality with greater confidence. Long term trends on the other hand are clearly evident within the time series of 37% of the analytes. This indicates that the aquifer is not in a steady state. Finally data should not be pooled due to differences in mean and variability between the data points. Instead an intrawell analysis should proceed, which will allow site-specific interpretations. From the results presented in this thesis the current level of data is not sufficient to statistically determine the baseline conditions of the aquifer. The high variability and secular trends should be investigated from a hydrogeological point of view.

In objective 3, the literature was reviewed in order to develop a list of parameters that must be sampled during monitoring. This resulted in 6 classes of chemical parameters that include macro inorganic chemical constituents, trace elements, inorganic compounds and dissolved gasses, and radiochemistry and isotopes. Anthropogenic chemicals commonly used in hydraulic fracturing fluid, were also included. A pre-development baseline monitoring phase should be undertaken with no more than quarterly sampling, due to slow groundwater movement. A detection monitoring phase should follow during shale gas development. During detection monitoring phase, in order to manage the large list of parameters, the sampling frequency plan followed a tiered approach. Only exceedance of thresholds set per tier warranted sampling in the next tier. Nevertheless, it was determined that no more than quarterly samples should be collected, based on groundwater flow rates.

In conclusion, an optimized spatial network of monitoring points was designed, which showed an overall increase in spatial distribution and thus a gain in information across the case study area. It was shown that the current groundwater quality data is not sufficient to statistically determine the baseline with confidence. Hence new long-term monitoring should take place to increase the current level of data. The parameters and frequency of monitoring were shown to be best managed by a tiered approach, which will allow better management of monitoring cost and facilitate the early detection of contaminants. However more data is required that will allow further optimization of the monitoring network and allow the establishment of a confident and defensible baseline. In this case it shows that developing shale gas in the region with the current monitoring system and level of data may lead to groundwater contamination and overuse. It is recommended that the monitoring system presented be implemented to allow for the collection of at least 3 years of pre-development data.

Key words

Groundwater

Groundwater Monitoring

South Western Karoo

Shale gas

Hydraulic Fracturing

Indicator parameters

Baseline

Geostatistics

Water Resource Protection

Contamination

Methane gas

Abbreviations

Ca: Calcium

Cl: Chloride

DMS: Dimethyl Sulphide

F: Fluoride

K: Potassium

Mg: Magnesium

NH_4: Ammonium

NO_3: Nitrate

Na: Sodium

PO_4: Phosphate

SO_4: Sulphate

Si: Silicate

TAL: Total Alkalinity

EC: Electric Conductivity

Temp: Temperature

F: Fluoride

DOC: Dissolved organic carbon

DIC: Dissolved inorganic carbon

TDS: Total dissolved solids

DO: Dissolved oxygen

ORP: Oxygen reduction potential

Zn: Zinc

Al: Aluminium

Sb: Antimony

As: Arsenic

Cd: Cadmium

Cr: Chromium

Co: Cobalt

Cu: Copper

Fe: Iron

Pb: Lead

Mn: Manganese

Hg: Mercury

Ni: Nickel

Se: Selenium

V: Vanadium

Ba: Barium

Li: Lithium

Sr: Strontium

Br: Bromine

U: Uranium

B: Boron

Rb: Rubidium

Mo: Molybdenum

VOC: Volatile organic compounds

PAH: Polycyclic aromatic hydrocarbons

SVOC: Semi-volatile organic compounds

TPH: Total petroleum hydrocarbons

Table of Contents

Declaration II

Acknowledgements III

Abstract IV

Key words VII

Abbreviations VIII

List of figures XI

List of tables XIII

Chapter 1 : General introduction 1

- 1.1. Introduction 1
- 1.2. Background information 2
- 1.3. Problem statement 4
- 1.4. Aims and objectives 4
- 1.5. Scope of the study 5
- 1.6. Rationale of the research 6
- 1.7. Research framework 7
- 1.8. Outline of the thesis 8

Chapter 2 : Literature review 9

- 2.1. Introduction 9
- 2.2. Shale gas and hydraulic fracturing 9
- 2.3. Hydraulic fracturing and contamination 10
- 2.4. Current literature pertaining to research objectives 13
  - 2.4.1. Groundwater monitoring network design 13
  - 2.4.2. Towards a groundwater baseline in the South Western Karoo 17
  - 2.4.3. Parameters of concern in shale gas environment 20
- 2.5. Key insights from the reviewed literature 24

Chapter 3 : Research design and methods 26

- 3.1. Introduction 26
- 3.2. Research design 26
  - 3.2.1. General description of case-study area 27
  - 3.2.2. General geology and hydrogeology of the case study area 29
  - 3.2.3. Design of sampling procedure 35
  - 3.2.4. Parameters of concern during data collection and analysis 35
  - 3.2.5. Data types and sources 36

3.3. Data collection methods 36
3.3.1. Groundwater sampling methods 36
3.3.2. Tools used during groundwater sampling 39
3.3.3. Procedure followed during groundwater sampling 39
3.4. Data analysis methods 40
3.4.1. Design of groundwater monitoring network 40
3.4.2. Analysis for determining baseline characteristics of aquifer 46
3.4.3. Determining monitoring parameters and frequency 51
3.5. Quality Control 53
3.5.1. Reliability 53
3.5.2. Validity 53
3.6. Statement on research integrity 54
3.7. Limitations of the study 54
Chapter 4 : Design of groundwater monitoring network 55
4.1. Introduction 55
4.2. Exploratory geostatistical analysis 55
4.3. Determining spatial autocorrelation 59
4.4. Design of the sampling grid 62
4.5. Hydrogeological placement of monitoring points 62
4.6. Re-evaluation of designed network 67
4.7. Discussion on results of designing groundwater monitoring network 67
4.8. Summary of results and discussion 72
Chapter 5 : Determining baseline in South Western Karoo 73
5.1. Introduction 73
5.2. Exploratory Data Analysis 73
5.3. Analysis of spatial and temporal data independence 75
5.4. Test to determine normality of dataset 75
5.5. Determination of seasonality within the dataset 77
5.6. Determination of secular trends within datasets 79
5.7. Determining the need to pool data 80
5.8. Discussion on the determination of baseline in South Western Karoo 81
5.9. Summary of results and discussion 84
Chapter 6 : Parameters and Frequency of Monitoring 86
6.1. Introduction 86

6.2. Parameters of concern 86
6.3. Frequency of monitoring 88
6.4. Discussion on parameters and frequency of monitoring 90
6.5. Summary of results and discussion 92
Chapter 7 : Conclusions & Recommendations 94
7.1. Introduction 94
7.2. Conclusions 94
7.3. Recommendations 96
7.4. Concluding remarks 96
References 97
Annex 1: Supplementary statistical information 113

List of figures

Figure 1.1. Schematic showing the geology of various hydrocarbon reservoirs. Reservoir such as gas-rich shales, tight sands and coalbed methane are forms of unconventional hydrocarbon deposits (U.S. Energy Information Administration, 2011). 2

Figure 1.2. Map of the general study area located in the upper Breede-Gouritz catchment area (red line). The Hydstra boreholes are shown to indicate the available data points (i.e. scope of the available data). 6

Figure 2.1. Infographic illustrating the typical application of shale gas extraction using hydraulic fracturing (ProPublica, 2017). 10

Figure 3.1. Map of the case study area, showing the various catchments as well as the current monitoring network. These points represent the data points used in this study. 28

Figure 3.2. Simplified geology of the case study area. 30

Figure 3.3. Groundwater flow direction and compartmentalization in the Beaufort West area (Rose, 2008). 33

Figure 3.4. Conceptual model of the groundwater system in the Beaufort West area, from north to south (Rose, 2008) 34

Figure 3.5. Example of experimental variogram, fitted with a corresponding mathematical model. C_0 is the nugget effect, C_1 is the sill range, and a is the spatial autocorrelation range (Caeiro *et al.*, 2003). 43

Figure 3.6. Flow-chart process followed to establish the relevant baselines for the study area (adapted from the Department of Environmental Quality, 2014). 49

Figure 4.1. Mean depth to water for current monitoring points in the study area. 55

Figure 4.2. Histogram for mean depth to water observations (A) and log-transformed depth to water (B), for the current monitoring system. The normal curve is also show in order to evaluate normality of the dataset. 57

Figure 4.3. Normal Q-Q plots of mean depth to water (A) and log-transformed depth to water (B) values, for the current monitoring network. 57

Figure 4.4. Trend analysis graph for mean depth to water observations in the study area. Figure A displays the calculated trend for the 0^0 direction while figure B is rotated 90^0. 58

Figure 4.5. Trend analysis graphs for log-transformed depth to water values. Figure A is for the 0^0 direction, while figure B is for the 90^0 direction. 59

Figure 4.6. Map of the prediction standard errors according to the kriging model 1. 61

Figure 4.7. Map of the designed hexagon sampling grid, showing the current monitoring points as well as the prediction standard errors. 62
Figure 4.8. Map depicting the rate of recharge for the study area. 63
Figure 4.9. Map depicting the outcropping of dolerite intrusion within the study area, including dikes. 64
Figure 4.10. Map of the aquifer yield and aquifer type in the study area. 65
Figure 4.11. Map depicting areas of low electric conductivity (i.e. favourable water quality). 65
Figure 4.12. Map of all new monitoring points. Existing monitoring points are including in the newly designed network..66
Figure 4.13. Prediction standard error map for the newly designed monitoring network. 67
Figure 6.1. Diagram showing the tiered approach to monitoring frequency and parameters. Successive tiers are only activated when there is an exceedance beyond the threshold limit. 90

List of tables

Table 2.1. Proposed baseline monitoring parameters for shallow aquifers, according to O'Brien *et al.* (2013).......21

Table 3.1. Setup parameters for models in the kriging tool.......45

Table 3.2. Criteria used for hydrogeological zone mapping.......46

Table 3.3. Methodological approaches used to establish baseline characteristics of an aquifer (Shand *et al.*, 2007).......47

Table 4.1. Summary statistics for depth to water for current monitoring points.......56

Table 4.2. Results of kriging analysis.......59

Table 4.3. Results of the cross-validation analysis.......61

Table 5.1. Data summary for 5 ZQM groundwater quality monitoring points in the study area. 73

Table 5.2. Statistical summary for various chemical analytes measured during sampling of ZQM boreholes.......74

Table 5.3. Results of the Shapiro-Wilks test for normality performed on original dataset.......76

Table 5.4. Results of the Shapiro-Wilks test for normality performed on transformed values. 77

Table 5.5. Results of Kruskal-Wallis test using seasonality variables.......78

Table 5.6. Results of the Mann-Kendall test for secular trend.......79

Table 5.7. Results for ANOVAs test for homogeneity of variances between ZQMWVL1 and ZQMWVL2.......80

Table 5.8. Results of the Kruskal-Wallis test for similar means between ZQMWVL1 and ZQMWVL2.......81

Table 6.1. Target chemical parameters to include during monitoring (Modified from O'Brien et al., 2013).......87

Chapter 1 : General introduction

1.1. Introduction

In the semi-arid regions of the South Western Karoo groundwater forms an integral part of the overall freshwater supply. The local population in the South Western Karoo depends on groundwater for domestic, agricultural and industrial purposes. However, groundwater in this region is vulnerable to natural and anthropogenic influence. In light of changing climatic conditions, recurring droughts, population growth, and increasing urbanization, greater effort is needed to protect this vital resource.

Best management strategies are usually employed to address issues of groundwater protection and sustainability. These strategies are often informed by continuous monitoring of the resource, which provides the information during decision making processes in management situations. Regular monitoring of groundwater resources is performed in-order to provide information on natural background (baseline) conditions of groundwater quantity (water levels) and quality, as well as to detect any changes is these properties. Furthermore it also informs authorities as to the cause of the changes, whether it be as a result of contamination or within natural fluctuations (Vrba and Adams, 2008). Therefore groundwater monitoring is a "must-do" practice to ensure groundwater protection.

Groundwater monitoring programs exists in either one of two forms depending on the objectives: regional/national scale schemes that are intended to classify background values and trends (baselines), or local/site-specific scale schemes that are intended to address issues of compliance or detection monitoring (Loaiciga *et al.*, 1992; Vrba and Adams, 2008). It is advised that an early warning mechanism be incorporated into both these schemes, and indeed site-specific monitoring does normally fulfil this need, but regional schemes generally do not (Vrba and Adams, 2008). Early detection of contamination (or changes) is an excellent tool to combat the effects of contaminants on groundwater quality (Gullick *et al.*, 2003; Vrba and Adams, 2008). Limiting degradation in the event of groundwater contamination and overexploitation can ensure that remediation is feasible and manageable.

In the South Western Karoo, oil and gas companies are interested in exploiting shale gas resources in the region. With this comes the added risk of contamination to regional groundwater resources (Vengosh *et al.*, 2014). Therefore this study proposes the development of a regional groundwater monitoring system, presenting in it the information necessary to design an effective system for protection against the risks of shale gas development.

1.2. Background information

In the last decade shale gas has emerged as a new and cleaner energy source, compared to coal and other fossil fuels (Wait and Rossouw, 2014). Unconventional hydrocarbon reservoirs, such as shale gas, compared to conventional reservoirs require unconventional techniques to develop due to their low permeabilities (Fig. 1.1.) (Boyer *et al.*, 2011). Shale gas is contained within deeply buried organic rich shale formation, and requires alternative extraction techniques to produce, such as horizontal drilling and hydraulic fracturing (U.S. Environmental Protection Agency, 2017).

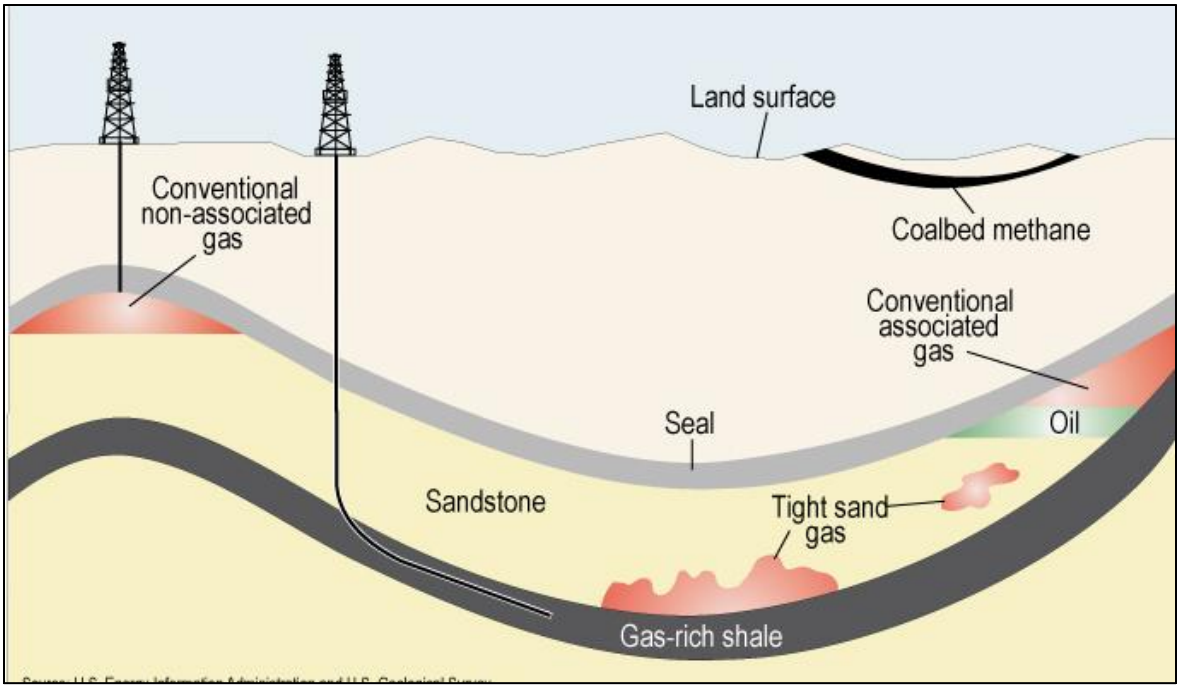


Figure 1.1. Schematic showing the geology of various hydrocarbon reservoirs. Reservoir such as gas-rich shales, tight sands and coalbed methane are forms of unconventional hydrocarbon deposits (U.S. Energy Information Administration, 2011).

The United States of America (U.S.A.), using horizontal drilling and hydraulic fracturing, were the first to exploit shale gas reservoirs, with great success. Production of shale gas from plays such as the Marcellus shale, Barnett shale and Eagle Ford shale have made the U.S.A. one of the biggest natural gas producers in the world (U.S. Energy Information Administration, 2016a; International Energy Agency, 2017). The success experienced by the U.S.A. has generated shale gas interests in other countries, including South Africa. Vast reserves of shale gas have been predicted in the deep formation of the Karoo basin (Geel *et* *al.*, 2013; Pietersen *et al.*, 2016). A number of international oil and gas companies have since applied for exploration rights in the Southern Karoo region, to exploit these reserves (Petroleum Agency of South Africa, 2013).

However, hydraulic fracturing has been criticised for being an environmentally detrimental practice, specifically on the shallow groundwater environment (Olmstead *et al.*, 2013; Vidic *et al.*, 2013; Jackson *et al.*, 2014; Vengosh *et al.*, 2014). Osborn *et al.*, (2011) and Jackson *et al.*, (2013) highlighted the issue of methane gas migration in the shallow groundwater systems in the Appalachian basin, which has brought to light the serious issue of faulty well barrier construction (Davies, 2011). Steyl and Van Tonder (2013) argue that the toxic nature of hydraulic fracturing fluid could contaminate shallow groundwater resources. Many of the chemicals used in hydraulic fracturing fluid were known to be carcinogens (e.g. Benzene, Naphthalene); others are neurotoxins and reproductive toxins (e.g. Isopropanol, Ethylene Glycol), and that these chemicals can impact the health of end users (Colborn *et al.*, 2011; Committee On Energy and Commerce, 2011; Maule *et al.*, 2013; National Toxic Network, 2013; Bamberger and Oswald, 2014). Also flow-back fluids and produced waters which are highly saline and contain above normal levels of naturally occurring radioactive material (N.O.R.M), may pose a contamination risk to shallow groundwater systems (Perry, 2011; Haluszczak *et al.*, 2013).

There are concerns amongst scientists and stakeholders that stray gas, hydraulic fracturing fluid, deep saline brines and N.O.R.M. can infiltrate into the shallow groundwater aquifers in the Karoo (Steyl and Van Tonder, 2013; Van Tonder *et al.*, 2013). This may impact the shallow groundwater quality, and its use as a freshwaters source may no longer be functional (Jackson *et al.*, 2014). The situation may be exacerbated by the presence of dolerite dyke that intrude through the Karoo Supergroup. These dykes complicate the structure of the basin and provide pathways for fluid movement from deeper formations to the shallow aquifers (Woodford and Chevallier, 2002; Steyl and Van Tonder, 2013; Van Tonder *et al.*, 2013).

The complexities associated with shale gas productions in South Africa (as highlighted in section 1.2.), necessitates the importance to establish effective groundwater governance protocols (Pietersen *et al.*, 2016). This introduces the importance of monitoring which informs the decision making process (O'Brien *et al.*, 2013; Pietersen *et al.*, 2016). The need for pre-drilling baseline and the parameters that compose it will aid the interpretation of data during possible shale gas contamination (Davies, 2011; Jackson *et al.*, 2013; O'Brien *et al.*, 2013).

1.3. Problem statement

Although monitoring is performed in the region, the design and set-up of this monitoring system is not adequate to provide the necessary level of data that can inform effective decision making. The objective of the current groundwater monitoring network in the South Western Karoo is more concerned with monitoring groundwater resource availability. It is not designed to collect regional background (baseline) groundwater data. In addition it is not designed to monitor trends in groundwater environment for detection of contamination. In the context of this study, without effective decision making, the contamination that might arise from shale gas development could negatively affect groundwater resources.

Groundwater resources of the South Western Karoo is a scarce resource, that is already vulnerable to contamination and misuse (Le Maitre *et al.*, 2009). With the possible development of shale gas resources in the future, groundwater resources in the South Western Karoo may be placed at even greater risk. Hence, if no effort is made to develop an effective groundwater monitoring system before shale gas development commences, then widespread contamination may occur that will deteriorate the groundwater resources. Therefore, the main problem that this study is intended to solve is, what groundwater monitoring system configuration can be designed to help mitigate risks associated with shale gas development.

1.4. Aims and objectives

The aim of this research is to develop a groundwater monitoring system for the collection of baseline date and detection monitoring of contamination from shale gas development.

In order to achieve this aim, 3 objectives were set, as follows:

I. Design a regional groundwater monitoring network that will effectively protect groundwater resource.
II. Assess current and historic groundwater quality data in-order to develop baseline conditions of relevant groundwater quality parameters.
III. Identify monitoring parameters and frequency of monitoring that are require to develop baseline conditions and predict contamination from shale gas development.

1.5. Scope of the study

Shale gas development does not only pose a risk to groundwater resources, but also to surface water resources and air quality as well (De Wit, 2011; Jackson *et al.*, 2014). However to ensure that the methodology is not complicated by an increase in variables, this research will focus solely on researching the development of a groundwater monitoring system. In doing so the validity of the results achieved can be ensured. Extensions into other environmental domains are beyond the scope and expertise of this research.

The concept of the objectives of a monitoring system was introduced. The objectives influence the design of the monitoring system. In this research the scope/objectives (purpose) of the designed groundwater monitoring system is 1) collect baseline groundwater quality data, and 2) provide detection monitoring. The scope of this groundwater monitoring system has been limited to these 2 objectives to avoid increasing the complications, and to ensure methodologies can be applied in a valid manner. Hence the objectives set in section 1.5. was set-up specifically to focus on achieving the scope of the groundwater monitoring system. Objective 1 will focus on the position of the monitoring points, objective 2 will focus on the baseline (reference) conditions, and objective 3 will focus on what parameters to measure and when to measure. In essence these or three components of a baseline and detection monitoring system.

The general study area of this research is located within the South Western Karoo (Fig. 1.2). It is also located within the upper Breede-Gouritz catchment area, and contains a large local population that is dependent on groundwater resources. The general study area represents the regional aspect of this research. Its extent is intended to cover portions of the most prospective areas for shale gas development. However, this research will focus on a smaller more restricted case study area (see chapter 3) was chosen in-order to effectively apply the methodology used in this thesis. A sub-set of the Hydstra data points shown in figure 1.2. are part of the current monitoring network.

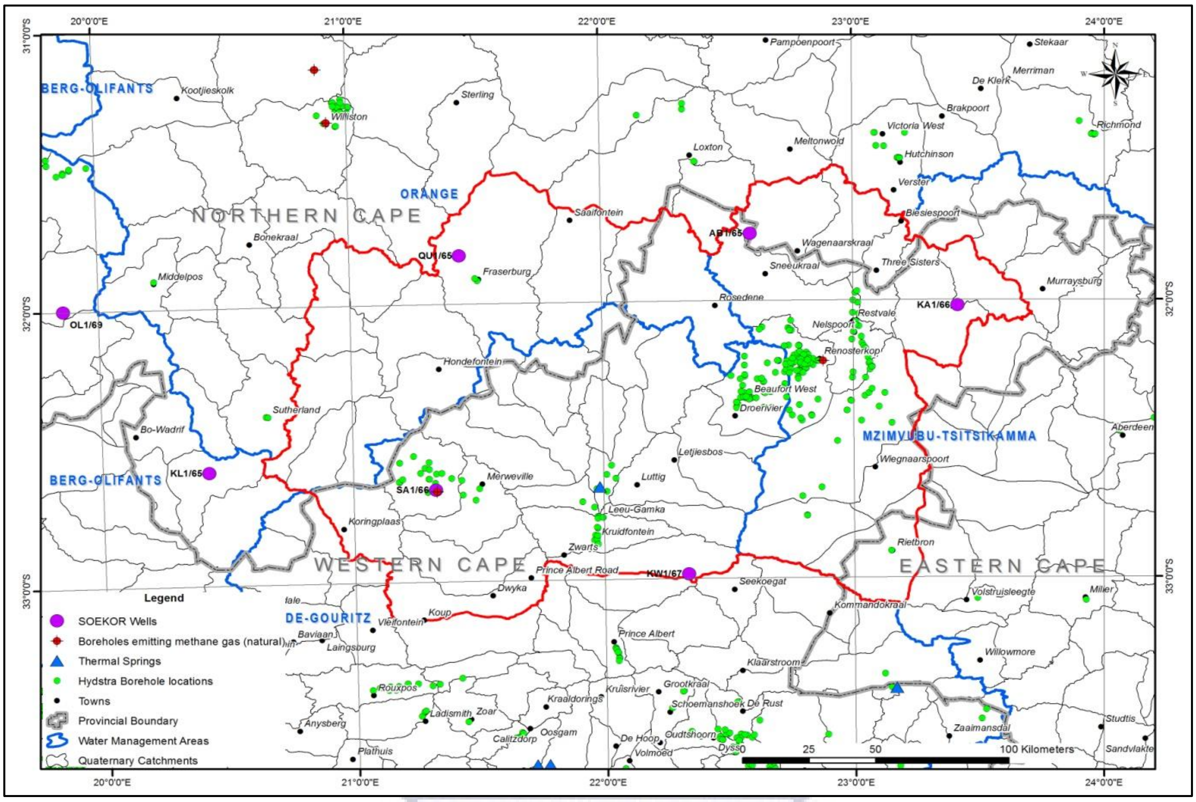


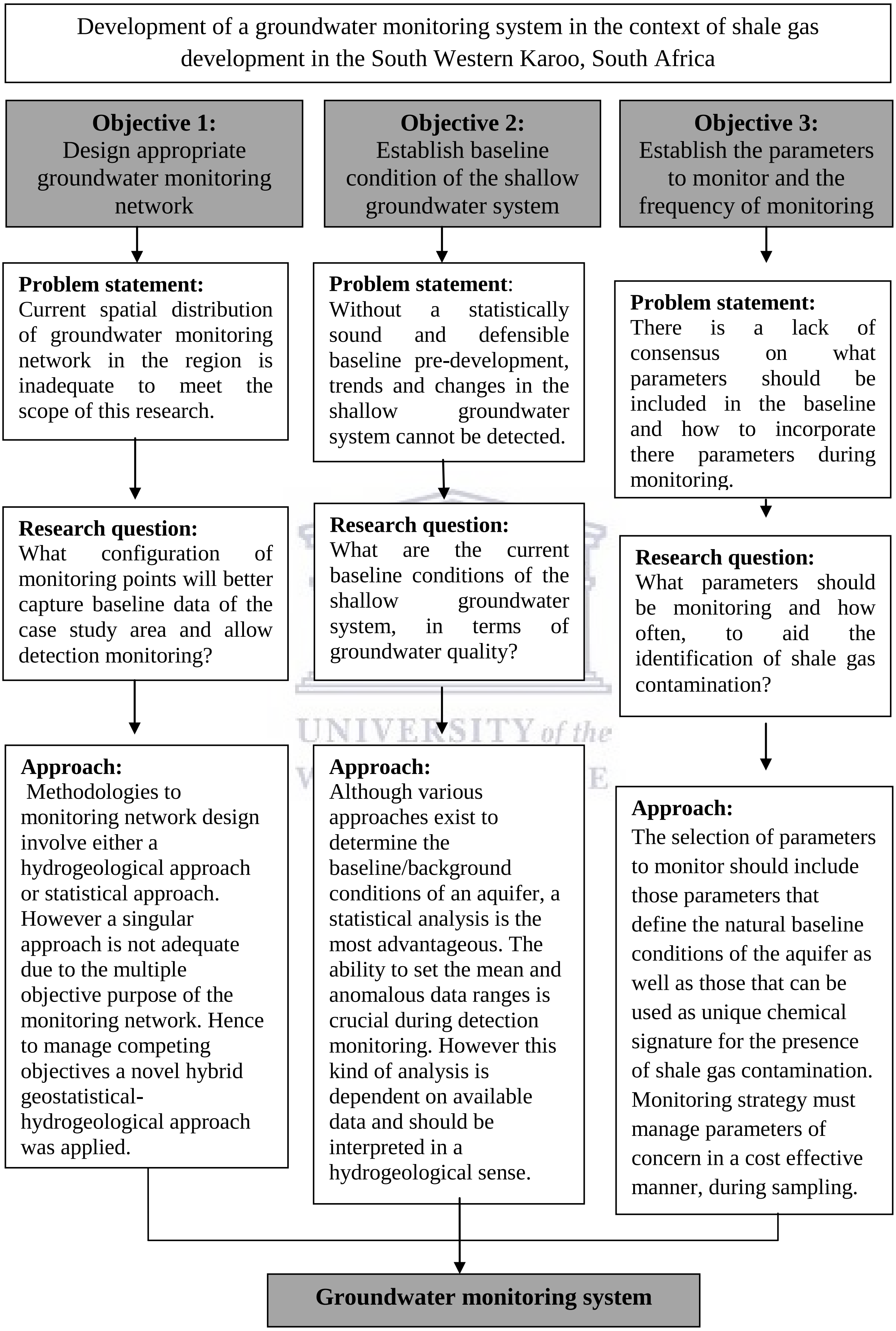
Figure 1.2. Map of the general study area located in the upper Breede-Gouritz catchment area (red line). The Hydstra boreholes are shown to indicate the available data points (i.e. scope of the available data).

1.6. Rationale of the research

This research is part of a more expansive Water Research Commission (WRC) project related to unconventional gas development in South Africa, and its links to the groundwater environment. As part of expert discussions during the WRC project, the need for baseline conditions of groundwater system to be established, was highlighted. The reason being that understanding existing natural conditions within the aquifers before shale gas development will allow us to detected and interpret changes in the aquifer after shale gas development (Davies, 2011; Jackson *et al.*, 2013). Intrinsic to achieving a baseline understanding of the aquifer, is a comprehensive groundwater monitoring system without which the necessary data cannot be acquired.

From an academic standpoint, the literature is limited with regard to groundwater monitoring network design in the context of shale gas development. Not much research has been done on this topic globally and within South Africa. Detection monitoring for instance is generally only applied to site-specific situation, and not on a regional scale (Vrba and Adams, 2008). Hence the novelty of the research, and the methodology therein, can be seen as a new approach to groundwater monitoring system design, at least within South Africa.

1.7. Research framework



Development of a groundwater monitoring system in the context of shale gas development in the South Western Karoo, South Africa

Objective 1:
Design appropriate groundwater monitoring network

Problem statement: Current spatial distribution of groundwater monitoring network in the region is inadequate to meet the scope of this research.

Research question: What configuration of monitoring points will better capture baseline data of the case study area and allow detection monitoring?

Approach: Methodologies to monitoring network design involve either a hydrogeological approach or statistical approach. However a singular approach is not adequate due to the multiple objective purpose of the monitoring network. Hence to manage competing objectives a novel hybrid geostatistical-hydrogeological approach was applied.

Objective 2:
Establish baseline condition of the shallow groundwater system

Problem statement: Without a statistically sound and defensible baseline pre-development, trends and changes in the shallow groundwater system cannot be detected.

Research question: What are the current baseline conditions of the shallow groundwater system, in terms of groundwater quality?

Approach: Although various approaches exist to determine the baseline/background conditions of an aquifer, a statistical analysis is the most advantageous. The ability to set the mean and anomalous data ranges is crucial during detection monitoring. However this kind of analysis is dependent on available data and should be interpreted in a hydrogeological sense.

Objective 3:
Establish the parameters to monitor and the frequency of monitoring

Problem statement: There is a lack of consensus on what parameters should be included in the baseline and how to incorporate there parameters during monitoring.

Research question: What parameters should be monitoring and how often, to aid the identification of shale gas contamination?

Approach: The selection of parameters to monitor should include those parameters that define the natural baseline conditions of the aquifer as well as those that can be used as unique chemical signature for the presence of shale gas contamination. Monitoring strategy must manage parameters of concern in a cost effective manner, during sampling.

Groundwater monitoring system

1.8. Outline of the thesis

Chapter 1 is an introduction of this research, introducing the topic and the importance of it. Chapter 2 is a comprehensive and critical review of the literature, starting with a brief history of the shale gas and hydraulic fracturing. Followed by a discussion of literature regarding the objectives mentioned in section 1.4. Chapter 3 details the research methodology and general approach that is utilised to accomplish the aim of this research. Chapter 3 also includes a description of the study area, in terms of its demographics, geology and hydrology, among other aspects. Chapter 4 presents the results of objective 1 followed by discussion of the results. Chapter 5 presents the results of objective 2 and followed also by a discussion of the results. Similarly for chapter 6 which describes the results of objectives 3 and a discussion of it. Finally chapter 7 is a succinct summary and conclusion of the entire research, and will recommend ideas for further research.

Chapter 2 : Literature review

2.1. Introduction

Groundwater monitoring has become an integral part of managing groundwater resources. Groundwater monitoring systems have evolved significantly as new knowledge has become available. Technological advances in the field have propelled this evolution, from conventional ground based observation to futuristic remote monitoring system. The current chapter is a review of the relevant literature pertaining to the 3 objectives stated in the previous chapter. This chapter presents the reviewed literature in such a manner as to contextualize this study topic, build a knowledge base and highlight any gaps in the current understanding of this research topic. Therefore the chapter starts with a review of the background literature, building on from the corresponding section in the previous chapter. Thereafter a systematic analysis of the relevant literature, regarding each objective of this research topic, is performed. The current chapter ends with a summary of key insights from the reviewed literature.

2.2. Shale gas and hydraulic fracturing

Shale gas is a form of natural gas that is contained within the pore spaces of deeply buried organic rich shales (Kuhn and Umbach, 2011). Shale gas has been grouped with coal bed methane and tight sands as unconventional reservoirs due to their extremely low hydraulic conductivities (shale formations will typically have a permeability of between 0.0001-0.000001 millidarcies) (King, 2012).

Hydraulic fracturing is an artificial enhanced stimulation technique used to extract natural resources, such as shale gas, from below the earth's surface more efficiently (U.S. Environmental Protection Agency, 2017). In its most modern form involves the fracturing of subsurface geological formation, using high pressure fluids (most commonly water) or gasses (Gandossi, 2013).

Hydraulic fracturing coupled with horizontal drilling was first used in the Barnett shale in 1991 (Robbins, 2013), and with the arrival of "slick-water" technologies, saw a massive boom in shale gas production in America (Gallegos and Varela, 2015). During 2015 the annual total production in the U.S.A. from the various shale gas plays was more than 15 Trillion Cubic Feet of natural gas (U.S. Energy Information Administration, 2016b). This represents approximately 50% of total natural gas production in the U.S.A., compared to only 1% in 2000 (Stevens, 2012; U.S. Energy Information Administration, 2017).

Shale gas, which is a cleaner energy source to burn (De Wit, 2011; MacKay and Stone, 2013), may be a viable option for South Africa's energy mix. Indeed if shale gas development is to proceed in South Africa, it is most likely that the combination of horizontal drilling and "slick-water" hydraulic fracturing will be used to extract the natural gas. Figure 2.1 illustrates the general process of shale gas extraction. This involves the injection of hydraulic fracturing fluid into horizontal boreholes at pressure. The increase in pressure fractures the shale formations, liberating the natural gas contained in the shale formation.

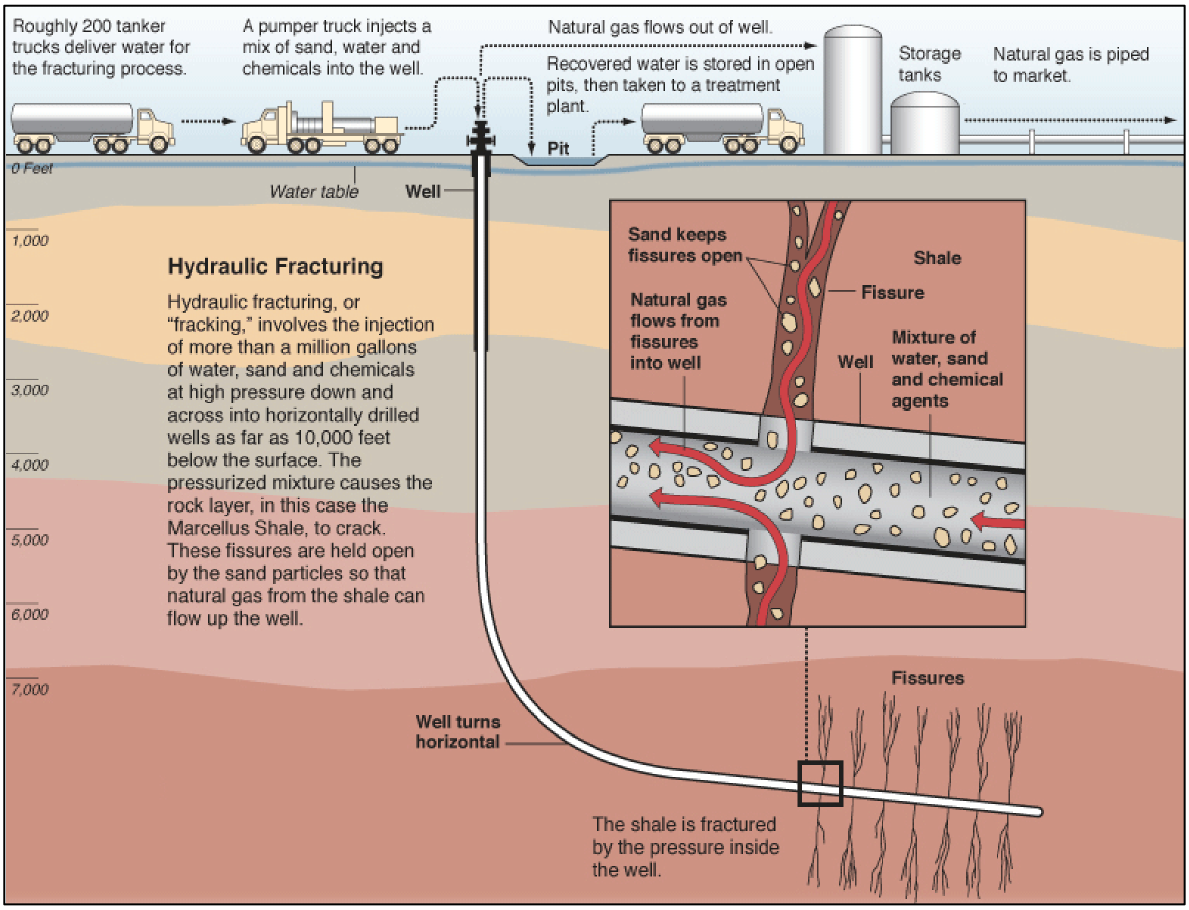


Figure 2.1. Infographic illustrating the typical application of shale gas extraction using hydraulic fracturing (ProPublica, 2017).

2.3. Hydraulic fracturing and contamination

Scientific investigations into the environmental impact of hydraulic fracturing and shale gas development have been numerous in the last few years, mostly due to public criticism. Studies have highlighted the general concerns with shale gas extraction on the shallow aquifers: contamination of stray methane gas, contamination and toxicity of hydraulic fracturing and drilling fluids on the environment, contamination by N.O.R.M's, and the infiltration of saline brines (De Wit, 2011; Steyl and Van Tonder, 2013).

One of the most publicized investigations attempting to connect hydraulic fracturing and groundwater contamination was conducted by Osborn *et al.* (2011). The analysis of 68 groundwater samples from private water wells in the Appalachian Basin reveals a higher concentration of thermogenic methane closer to active gas production wells. Compared to a more biogenic signature for methane further away. This would suggest the influence of hydraulic fracturing and shale gas development on the concentration of methane in the shallow aquifer system. Further investigations by Jackson *et al.* (2013) confirms that indeed thermogenic methane was more prominent closer to shale gas wells. Critics, however, argue that this does not indicate a direct link to hydraulic fracturing, and that other mechanism of stray gas migration need to be considered (Davies, 2011). It would appear wise to monitoring closer to production boreholes, which could act like point sources of pollution (see the following paragraph). However, it is not yet known where production borehole will be placed, and so cannot be accounted for in the present research.

Artificially induced fractures created by the hydraulic fracturing process is one possible mechanism that could promote stray gas migration, but these tend to be no greater than 600m in length (Davies *et al.*, 2012; Flewelling and Sharma, 2014). In deeply buried shale formations that are separated from the shallow groundwater systems by thousands of metres of sediments, chances of hydraulic fracturing creating pathways for contaminant migration is extremely low (Kuhn and Kempka, 2013). Therefore, this research does not consider this feature in the design of the present groundwater monitoring network.

Other mechanisms, and of most concern, is the possibility that stray gas can leak into shallow aquifer system from faulty well bore barriers, or through natural fractures, faults and dykes, among others (Davies, 2011; Osborn *et al.*, 2011; Warner *et al.*, 2013; Jackson *et al.*, 2014). Studies into well integrity failure have been completed by Considine *et al.* (2013), Vidic *et al.* (2013), Davies *et al.* (2014), Ingraffea *et al.* (2014). Their results indicated that faulty well bore barriers are major concern during shale gas development and hydraulic fracturing. In addition, the Karoo Basin is intruded by numerous dolerite dykes, that are known to provide pathways for groundwater movement between shallow formations and deeper formations (thermal springs) (Woodford and Chevallier, 2002). However the movement of groundwater in the deep geology is still poorly understood (Rosewarne *et al.*, 2013; Steyl and Van Tonder, 2013; Van Tonder *et al.*, 2013). Hence, the location of these dykes and old deep boreholes is incorporated into the design of the present groundwater monitoring network.

Methane may not be the only the issue with regard to contamination and shale gas development. Hydraulic fracturing fluid and drilling fluids which is used to perform the hydraulic fracturing operation is known to contain a cocktail of compounds that are toxic and or carcinogenic to humans (Colborn *et al.*, 2011; Committee On Energy and Commerce, 2011; National Toxic Network, 2013). These can be introduced through faulty well bore barriers, but are more likely introduced into the shallow aquifers through surface spills of flowback & produced fluids (Olmstead *et al.*, 2013). Hence, this research includes parameters that can identify the presence of hydraulic fracturing fluid in the aquifer.

NORM's are also an environmental concern in shale gas development. Although these elements do occur everywhere throughout nature, they are more concentrated in shale formations (Perry, 2011). NORM's are generally a concern with regard to produced water, where they become concentrated in storage ponds (Perry, 2011). It is during improper handling and storage that these toxic fluids contaminate the groundwater (Olmstead *et al.*, 2013). Similarly produced waters tend to be hypersaline and can severely deteriorate the quality of the groundwater if introduced to the shallow aquifers (Vengosh *et al.*, 2014). NORM may be an even greater concern in the Southern Karoo, with the presence of uranium ore-bodies in the sediments of the Karoo Supergroup (Duane *et al.*, 1989). Hence, in the present research parameters that monitor the radioactivity of the groundwater is included.

What is often not discussed when concerned about the environmental effects of shale gas development is groundwater over-exploitation. In the U.S.A. the acquisition of water for the drilling and hydraulic fracturing comes mainly from surface water, groundwater and hydraulic fracturing waste water reuse (U.S. Environmental Protection Agency, 2016). Though the total consumption has been far less than other industries, which should not affect national water demand (U.S. Environmental Protection Agency, 2016). However in arid or semi-arid environments majority of the water comes from groundwater, and this could cause local groundwater shortages (Vengosh *et al.*, 2014). In the present study, through the monitoring of groundwater levels, over-exploitation is recognised.

Section 2.3. introduces the most critical concerns; such as stray methane gas, toxic hydraulic fracturing fluid, and NORM, with regards to contaminants. They also highlight possible pathways for contaminant movement in the subsurface, such as dolerite dyke and faulty well casings. In addition, over-exploitation of already stressed water resources in the Karoo could also created water shortages in the region. In summary hydraulic fracturing and shale gas development could present the Karoo with serious groundwater related issues such as contamination. Hence, the need for a groundwater monitoring system that could prevent such contamination and degradation of water resources.

2.4. Current literature pertaining to research objectives

2.4.1. Groundwater monitoring network design

Groundwater monitoring systems/network are essential components for effective groundwater management (Vrba, 2001). They promote informed decision making by providing valuable data on groundwater characteristics, such the current state and trends in quality and quantity. As such they are valuable tools in issues of groundwater protections and sustainability (Vrba and Adams, 2008).

However, most monitoring systems are focused primarily on identifying and controlling the effects of groundwater quality deteriorations, and not on preventative protection (Vrba and Adams, 2008). As such detection monitoring incorporated in a conventional groundwater monitoring system is a must for comprehensive protection (Vrba, 2001). Groundwater monitoring warning systems are required to detect changes in the groundwater before widespread degradation occurs and thus allow timely response to alleviate the cause (Vrba, 2001; Gullick *et al.*, 2004). Hence the scope of the groundwater monitoring systems designed in the present study includes the ability to perform detection monitoring on a regional scale.

The design of a groundwater monitoring network depends mostly on the specific objective of the system (Dutta *et al.*, 1998). Factors such as the hydrogeology, variables of concern, scale of the network, surface-unsaturated-saturated zone interactions, contaminant-hydrogeology interactions, as well as anthropogenic activities must be considered when designing monitoring wells and well locations (Kim *et al.*, 1995; Dutta *et al.*, 1998; Vrba and Adams, 2008). Hence, due to the varied nature of hydrogeological environments across the world, the transferability of an existing system to another region is generally not applicable (Dutta *et al.*, 1998). Thus, this research focuses on an alternative approach to design the intended groundwater monitoring system, based on the specific objectives.

In general there are 3 components that constituent a groundwater monitoring system, in order to realise it's objective (Subcommittee on groundwater, 2013). These are: the spatial network of monitoring points (or the monitoring network density), the frequency of monitoring and the various analytes of concern (Subcommittee on groundwater, 2013). The set-up of these components once again depends largely on the objectives required in the design of monitoring. Current trends in the literature focus on the use of optimization techniques of the monitoring network design, most specifically towards network density (Zhou, 1996; Nunes *et al.*, 2004; Theodossiou and Latinopoulos, 2006; Chadalavada and Datta, 2008; Yang *et al.*, 2008; Narany *et al.*, 2015; and many more). For example, geostatistical techniques are commonly used to determine the spatial network of monitoring points in-order to collect non-redundant information from the aquifer (Yang *et al.*, 2008). Optimization in such a manner is applied more readily currently as greater effort is made to limit cost of monitoring network implementation (Theodossiou and Latinopoulos, 2006; Chadalavada and Datta, 2008). Because of the strength of geostatistical techniques, it is adopted into this methodology of this study.

In terms of objectives, this is the main factor influencing the overall design and implementation of the groundwater monitoring network (Loaiciga *et al.*, 1992; Subcommittee on groundwater, 2013). According to Loaiciga *et al.* (1992), Groundwater monitoring systems around the world generally exist to meet one or more of the following criteria: ambient monitoring, detection monitoring, compliance monitoring, or research monitoring. Ambient monitoring (or background monitoring) is concerned with establishing characteristics of regional groundwater systems. Detection monitoring is designed to detect the presence of target parameters as soon they exceed background limits. This is typical applied in and around point and non-point sources of contamination, such as waste-disposal sites. Compliance monitoring network are normally set-up to enforce strict rules pertaining to groundwater quality characteristics after detection. While research monitoring is the spatial and temporal network used to achieve specific research goals (Loaiciga *et al.*, 1992).

According to (Vrba and Adams, 2008) the scale of monitoring networks generally come in two forms: Background monitoring and site-specific monitoring. Background monitoring network are concerned with establishing the natural groundwater characteristics (baseline), and are usually applied at a national/regional scale, but may also be applied at a local level (Vrba and Adams, 2008). Local or site-specific monitoring networks are designed normally to address a specific problem (point source pollution), for example the monitoring of pollutants leached at waste disposal sites (Vrba and Adams, 2008). Both designs should ideally include an form of detection monitoring (early warning system) (Vrba and Adams, 2008). In shale gas application in the Karoo it may be necessary that a combination regional and site-specific (around well pads) system be designed (Ward *et al.*, 2005; Jackson *et al.*, 2013; O'Brien *et al.*, 2013).

One important consideration when designing groundwater monitoring systems is the effects of anthropogenic activities (Dutta *et al.,* 1998). In this regard the concern is mostly with how contaminants are introduced into the groundwater environment (i.e. point source, multi-point source, or a diffuse source, etc) (Vrba and Adams, 2008). In shale gas activity the contaminants can be introduced to the groundwater from various surface and subsurface sources (Jackson *et al.*, 2013). By extension it will also be important to understand the movement and interaction of the contaminants (e.g. stray methane gas) within the shallow subsurface environment (Vrba and Adams, 2008). This would directly affect the construction and location of monitoring wells. Currently the understanding on the contaminant transport movement with regards to shale gas in a South African context is very limited. This research does not aim to address this issue, however possible transport mechanism and potential receptors is considered in this research.

Most of the contaminants of concern in shale gas production have been well researched before in other applications (e.g. brine salts and hydrocarbons). Gorody (2012) describes the fate and transport of hydrocarbon gases in the subsurface. Xu *et al.* (2012) also researched the fate and transport of radioactive elements in Karoo fractured rock aquifers. U.S. Environmental Protection Agency (2011) also presents various articles on the fate and transportation of various contaminants in hydraulic fracturing operations. However, there is limited information on the characteristics of chemical additives (in drilling and hydraulic fracturing fluids) in the subsurface (Jackson *et al.*, 2013). In addition, limited information exist on the fate and transportation of more uncommon compounds, such as ^{226}Ra and Ba, explains Jackson *et al.* (2013).

The early warning function that can be implemented in groundwater monitoring systems can be achieved through the implementation of various techniques: remote sensing techniques, soil-gas surveys, vadose zone monitoring, etc. (Vrba, 2001; Vrba and Adams, 2008). The temporal resolutions (i.e. the observation/sampling frequency) and the monitoring wells network play an important role in realising the early warning function (Khader and McKee, 2014). Indeed much has been written about the optimization of groundwater monitoring networks both temporally and spatially (Loaiciga *et al.*, 1992; Meyer *et al.*, 1994; Zhou, 1996; Wu *et al.*, 2005; Khader and McKee, 2014). However, the lack of data on the source of contamination, in a Karoo context, hinders the implementation of an early warning system.

Regarding the temporal resolution of the monitoring network, conventional sampling and laboratory analysis are often to slow to provide time effective warning (Storey *et al.*, 2011). For more rapid detection real-time on-line systems are employed which can detect (in-situ) contamination and relay information at near real-time to an early warning systems (Storey *et al.*, 2011). However, this is not yet a perfected technology, and the merger of different components of such a systems (i.e. data transmission, sensors, data management systems etc.) still requires more research (Gunatilaka and Dreher, 2003; Storey *et al.*, 2011).

Considering that so many factors contribute to an effective groundwater monitoring system, it is unfortunate that there are so few examples within unconventional resources, from which to extract lesson. In-fact only one such system was found, and is currently being operated as a pilot project in the Denver-Julesburg Basin of North Eastern Colorado (Center for Energy and Water Sustainability, 2014). In this area the exploration and production of unconventional oil and gas has increased significantly in the last few years (Li and Carlson, 2014). Hence the need for a detection monitoring system.

The system consist of a network of monitoring stations that are linked to a real-time system which transmit the data almost instantaneously to a data management centre, where it is analysed by an advance data management software, called CANARY. This software was designed to manage such networks and to detect anomalous changes in the observations compared to background values (McKenna *et al.*, 2010). Similar software is available in the market each with its own advantages and disadvantages (Storey *et al.*, 2011). The use of such software in this study is an important component in order to perform the role of detection monitoring.

In South Africa groundwater monitoring is part of the responsibilities of the Department of Water and Sanitation (DWS), as stated by national legislation (van Wyk, 2010). The DWS operates a national network of water level (approx. 2500 points) and water quality (Approx. 311 active points and approx. 150 inactive points) monitoring stations (van Wyk, 2010; Department of Water & Sanitation, 2017a, 2017b). Both water levels and water quality observations are conducted biannually, in the dry and wet season (Department of Water & Sanitation, 2017a, 2017b).

According to the national framework setup by DWS the water level monitoring network can operate as both a regional background monitoring network and if necessary as a site-specific monitoring network (van Wyk, 2010). However, the water quality monitoring network (ZQM points) generally are only developed for regional background data collection (van Wyk, 2010). Hence, the spatial distribution of monitoring points in the case study area is not adequate to capture the required level of data to establish a statistically sound baseline. Furthermore, there does not exist a groundwater detection monitoring system in the Karoo, for the monitoring and prevention of groundwater quality deterioration. The aim of this study is to address this issue by developing the appropriate groundwater monitoring network.

2.4.2. Towards a groundwater baseline in the South Western Karoo

Osborn *et al.* (2011) was one of the first to investigate the link between hydraulic fracturing and groundwater contamination. Their research implied that hydraulic fracturing and shale gas operations were the most likely cause. However, Davies (2011) makes the comment that hydraulic fracturing cannot conclusively be labelled as the cause without a pre-drilling baseline of shallow groundwater properties, a sentiment echoed by Osborn *et al.* (2011). Similarly, a United States Environmental Protection Agency investigation into groundwater contamination from unconventional gas extraction near Pavillion in Wyoming was inconclusive as to the cause without the presence of pre-extraction baselines of the groundwater (DiGiulio *et al.,* 2011). Baselines data of groundwater attributes, surface-water attributes, air quality and human health is recognized as an important need to interpret possible future impacts of hydraulic fracturing (Jackson *et al.*, 2014).

In South Africa, many authors (De Wit, 2011; O'Brien *et al.,* 2013; Pietersen *et al.*, 2016) also emphasize the need for baseline data pre exploration and production of shale gas in Karoo, however none exists. At least 1 project is currently underway to characterize baseline groundwater properties in the Karoo, focusing on the eastern half of the prospective area (Stroebel and de Wit, 2015).

No peer-reviewed studies have been done on baselines and shale gas in Karoo. However O'Brien *et al.*, (2013) did highlight the necessities and complications associated with developing a baseline in the Karoo. In summary their work states that defensible baselines should include:

1. Information on the natural temporal and spatial variation in groundwater levels and quality;
2. Appropriate threshold limits to indicate anomalous changes in groundwater properties;
3. Monitoring points that take cognisance of the conceptual hydrogeological model to understand the link between receptors, aquifer systems, flow paths and release mechanisms; and
4. Parameters that define general water quality, those that are required to adhere to safe drinking water standards (SANS), those that can act as indicators to possible contamination sources and those that may be introduced during shale gas activities, such as drilling and hydraulic fracturing fluid additives (see table 2.1. for a detailed list of parameters).

Points 1 and 2 focus on the requirements of a groundwater baseline in the context. This is simply the natural conditions of the aquifer through time and space. Baselines as a reference datum require the establishment of threshold limits that define whether observations are within background range or can be considered an anomaly. Setting up and defining what these limits will require the application of rigorous statistical analysis in combination with geospatial understanding of the hydrogeology (Reimann and Garrett, 2005; McQueen, 2006). Common methodology associated with developing a baseline and associated threshold limits have been discussed by Shand *et al.* (2007). Such a statistical analysis has not been attempted before in the South Western Karoo. Hence, the present study is aimed at apply a statistical approach in the analysis of the current data set in the case study area.

The literature regarding the hydrogeology of the shallow aquifer include research by Vogel *et al.* (1980); Seward (1988); Woodford and Chevallier (2002); Department of Water Affairs and Forestry (2009); Brits (2012); Xu *et al.* (2012); Murray *et al.* (2012, 2015); Water Experts Group (2012, 2013); Solomon (2013), among many others. Even with the information presented in the literature, there is still a lack of knowledge regarding the temporal (long and short term) variations in groundwater chemistry, vertical differences in groundwater chemistry within the aquifer, as well as local scale spatial variations in groundwater chemistry (Woodford and Chevallier, 2002). The available data set is also limited in providing an understanding of the knowledge gaps. This may prove to be a hindrance in qualifying any local scale baselines. For example, the vertical variations and the small-scale spatial variations in groundwater chemistry may be important when considering site-specific (well pad location) baselines (O'Brien *et al.*, 2013). However, a greater level of data is first required to address the identified knowledge gaps. Therefore, the groundwater monitoring system developed in this research includes the capacity to capture new baseline data throughout the case study.

The present understanding of the Karoo hydrogeology is continually being revised and improved with new research work. The current knowledge does allow an attempt at the development of a baseline for the Karoo groundwater, specifically with regard to the shallow aquifer system. However, there is still a lack of groundwater related baseline projects in South Africa from which to draw lessons from that may aid in the development of groundwater baselines. Most likely such lessons can be best learnt from examples from shale gas projects beyond South Africa.

Even in international literature the topic of groundwater baselines in unconventional gas context has not been researched extensively. In fact only 4 such projects could be found from the literature that deal with this subject: Sloto (2013), Chapman *et al.,* (2014), Down *et al.* (2015) and Atkins *et al.* (2015). Moreover, none match the framework of O'Brien *et al.* (2013). The British Geological Survey are also in the process of completing a regional scale groundwater baseline regarding methane concentration in the subsurface (Bell *et al.*, 2016).

Sloto (2013) conducted groundwater sampling in North Central Pennsylvania in anticipations of shale gas development. 20 waters samples from 1 sampling programs was analysed for various constituents. Key parameters were included such as, dissolved methane, gross alpha and beta radioactivity, isotopes of carbon and hydrogen in methane, among others. The concentration of nutrients was also included to account for the affects of nutrient rich fertilizer. The results of this study produced only basic descriptive statistics (max, min and median values) of the various concentrations of analytes in the samples. It was intended to be a preliminary documentation of the groundwater baseline before shale gas development proceeded. Unlike the present research, the work by Sloto (2013) does not provide any experience on the statistical requirements of a groundwater baseline as stated by O'Brien *et al.* (2013).

Chapman *et al.* (2014) and Down *et al.* (2015) conducted pre-shale gas exploration groundwater baseline investigations in Central North Carolina. These studies together analyzed a selection of parameters, including dissolved methane, isotopes of carbon and hydrogen in methane, isotopes and radium, and Volatile Organic Compounds (VOC), among others. Down *et al.* (2015) included isotopic analysis of the various samples, in-order to better understand the source and evolution of the water through the subsurface. However their research also was not able to include a statistical analysis as required for a baseline in unconventional gas context.

Atkins *et al.* (2015) conducted groundwater baseline investigation in the Eastern New South Wales region, Australia, related to Coal Bed Methane. Although this is not a shale gas area and was not pre-development, it can still be considered and informs on what is required in unconventional resource development. They analyzed 91 samples for various parameters such as dissolved methane, isotopes of carbon in methane, Dissolved Inorganic Carbon (DIC), Dissolved Organic Carbon (DOC) and major ions, among others. There results highlighted the distribution of the various parameters in different geological units. Their research as well, did not include the relevant statistical analysis as required in the framework of the present research. However it does highlight the importance of providing a baseline for the different geological units in the study area. Unfortunately the available data in the present research does not support such an analysis, and hence is not attempted.

From the examples provided, it can be seen that the application of a statistical sound baseline which adheres to the framework of O’Brien *et al.* (2013) is non-existent. The importance of this research can then be signified by the fact that it has both international and local relevance. No work on groundwater baselines meets the framework that is required for the Karoo. This presents an opportunity for this research to be the first to develop such a functionally comprehensive baseline in an unconventional gas resources related context.

As a matter of interest point 3 and 4 in the framework of O’Brien *et al.* (2013) stated in section 2.4.2. are concerned with objectives 1 and 3 respectively, of the present research.

2.4.3. Parameters of concern in shale gas environment

Point 4 in section 2.4.2 describes the necessary chemical parameters that must be included in the groundwater baseline for shale gas and hydraulic fracturing activities specifically. Besides table 2.1, published by O’Brien *et al.* (2013), the research by Warner *et al.* (2013, 2014);

Darrah *et al.* (2014); and Murray *et al.* (2015), provide useful information regarding chemical parameters of concern. The present research attempts to consolidate and expand on the literature presented in this study, thereby generating a more comprehensive suite of chemical parameters.

The list of key parameters presented by O'Brien *et al.* (2013) concur with those parameters investigated for in the baseline studies of section 2.4.2. This would suggest that these parameters are the meaningful and relevant parameters to measure in connection to hydraulic fracturing and shale gas development. However, it can also be seen that the examples of baseline work presented in section 2.4.2. fails to address all the key features of a baseline as indentified in the framework by O'Brien *et al.* (2013). For example, minimal relation is made between the conceptual hydrogeological model and the groundwater properties. In addition no long term data was analysed to gain and understanding of seasonal and secular trends. Hence, the present study aims to incorporate those parameters identified as meaningful in the literature.

Table 2.1. Proposed baseline monitoring parameters for shallow aquifers, according to O'Brien *et al.* (2013).

Macro parameters	Trace parameters	Shale Tracers	Organics & Gases	Radiochemistry & Isotopes
pH	Zn	Ba	Dissolved Methane	Methane $\delta^{13}C$
EC	Al	Li	Dissolved Ethane	Methane δD
Ca	Sb	Sr	Radon	Water $\delta^{18}O$
Mg	As	Br		Water δD
K	Cd	U	VOCs	DIC $\delta^{13}C$
Na	Cr	B	PAH	
NH_4	Co	Rb	SVOCs	Gross alpha radioactivity
Cl	Cu	Mo	Glycols	Gross beta radioactivity
NO_3	Fe		Alcohols	Ethane $\delta^{13}C$
SO_4	Pb		TPH	Ethane δD
PO_4	Mn			
Alkalinity	Hg			
F	Ni			
DOC	Se			
DIC	V			
TDS				

** The list is designed to meet the criteria stated in point 4 in section 2.4.2. (D = Deuterium)*

Besides common chemical parameters sampled during water quality analysis, indicator parameters are those determinands that provide clues as to the origin or cause of a change in the groundwater environment (Vrba and Adams, 2008). Essentially, this could be a simple field parameter such as electric conductivity or pH, a specific chemical signature or the presence of a unique chemical compound. In shale gas related monitoring, it would be those groundwater determinands that are most susceptible to the risks imposed by hydraulic fracturing and shale gas development (Vengosh *et al.*, 2014; Son and Carlson, 2015). For example, methane from thermogenic sources or increase in NORM concentrations from produced or flow-back fluids (Vengosh *et al.*, 2014). Therefore the present research focuses on highlighting those determinands that can be used as indicators (provide direct or indirect evidence of contamination in the aquifer).

Son and Carlson (2015) provides evidence that even monitoring of field parameters, such as electric conductivity, oxidation-reduction potential and dissolved oxygen can be used as indicators in identifying for hydraulic fracturing and shale gas contamination. Son and Carlson (2015) set up experiments to establish a correlation and response time of various field parameters to the influence of produced water and methane. The experiments showed that Electric Conductivity (EC) has a high correlation with produced water and that EC increased significantly within in a few minutes, even at low contaminant concentrations. Similarly the experiments also showed that Oxygen Reduction Potential (ORP) had a high correlation with methane. Their work illustrates that EC, ORP and to a lesser extent Dissolved Oxygen (DO) can be used to detect contamination from hydraulic fracturing and shale gas activities.

It must be noted that in Son and Carlson (2015), EC changes was indirectly correlated to Total Dissolved Solids (TDS) contamination. Hence changes in EC can be used to infer contamination from high TDS fluids such as produced waters, hydraulic fracturing fluids and drilling muds, etc. (Son and Carlson, 2015). Further it is described that monitoring of field parameters provides a quick and inexpensive means, compared constant sampling and chemical analysis (Son and Carlson, 2015).

It may be possible for shale formation brines to contaminate shallow aquifers during shale gas development in the Karoo, due the presence of vertical extensive flowpaths (e.g. dolerite dykes) (Steyl and Van Tonder, 2013; Warner *et al.*, 2013). It may then be necessary acquire the chemical signature of these formation brines (Ecca shales), for the purpose of matching observations during monitoring (O’Brien *et al.*, 2013; Down *et al.*, 2015). Indeed this could prove to be a useful indicator to identify possible charging of the shallow aquifer with these formation brines (Warner *et al.*, 2013). Unfortunately the present research could not include such an analysis as there is currently no data available on the deep formation brines.

Another environmental concern with shale gas production is the release of produced and flowback fluids during spills into surface water and shallow groundwater systems (Vengosh *et al.*, 2014; Warner *et al.*, 2014). Warner *et al.* (2014) describes various elemental and isotopic signatures (B/Cl, Li/Cl, $\delta^{11}B$ and $\delta^{7}Li$) that are useful in identifying the presence of flowback fluids in the shallow groundwater in the Marcellus region. Warner *et al.* (2014) further makes the assumption that this technique can be universal in application to all shale gas regions across the world. However, this technique would depend on the chemical signature and various concentrations of Li and B in the Ecca shales. Currently there is very little data on the composition of flowback fluids in the Karoo context, which limits its usefulness in this research.

With regard to fugitive gas contamination of shallow aquifers general isotopic techniques [e.g. $\delta^{13}C\text{-}CH_4$, $\delta^{2}H\text{-}CH_4$, or $\Delta^{13}C=(\delta^{13}C\text{-}CH_4$ minus $\delta^{13}C\text{-}C_2H_6)$] can be applied to decipher naturally occurring methane (biogenic) from those introduced through anthropogenic activities (Osborn *et al.*, 2011; Jackson *et al.*, 2013; Darrah *et al.*, 2014). However, Darrah *et al.* (2014) explains that the original chemical signature of these parameter can be susceptible to changes from oxidation and microbial activity. They instead illustrate the use of noble gas geochemistry (^{4}He, ^{20}Ne, ^{36}Ar) plus the hydrocarbon and molecular geochemistry as more robust indicators for fugitive gas migration. Hence, this research includes the use of noble gas geochemistry as a groundwater baseline parameter.

An alternative solution that goes beyond conventional chemical or physical indicators is the application of biomonitoring. The response of aquatic or terrestrial species to changes in water quality can act as effective indicators of pollution (Steube *et al.*, 2009). In many cases systems can be setup with bio-indicators to provide early warning, sometimes in advance of conventional systems (Mikol *et al.*, 2007). Such techniques are widely applied in surface water environments to monitoring for contamination and ecosystem health (Steube *et al.*, 2009). Even though theory predicts that subsurface species living in groundwater environments can provide bio-indication of contamination, there has been limited application of the techniques (Edmands *et al.*, 2001; Steube *et al.*, 2009; Stein *et al.*, 2010). Further from the literature it can be seen that biomonitoring has not been considered before in shale gas projects, even though it may pose a threat to ecosystem and animal health (Bamberger and

Oswald, 2014). One technical drawback is that the response of various organisms to the various pollutants is not well understood in a shale gas context. However, this approach is beyond the scope of this research and thus won't be considered as parameters of concern.

Essentially indicators to elucidate contamination from hydraulic fracturing and shale gas development depends on the chemistry of the various affecting media, such as flowback fluids, stray gas and produced water, among others. Many of the elemental and isotopic signatures of these media are conserved during operations, and can thus serve as indicators. In addition, technology is continually being advanced for more effective solutions. The general trend is toward synthetic tracer tags (e.g. synthetic DNA tags, nanoparticles) that can be added to the fluids and can be subsequently monitored for. These will be most versatile indicators yet developed (Sharma, 2010). This research advocates the use of such tracers, if shale gas development is to proceed.

The last category of monitoring parameters includes those anthropogenic compounds that do not occur naturally in the groundwater environment. Although the exact configuration of hydraulic fracturing fluid are generally kept secret, a list of chemicals commonly used in these fluids have been reported (Committee On Energy and Commerce, 2011). Drilling fluids are also known to contains toxic chemicals (National Toxic Network, 2013). Although most hydraulic fracturing and drilling fluids contain many chemicals that are not naturally occurring in the groundwater, and so cannot be quantified in a pre-exploration baseline, they should still be monitored for during shale gas operations (O'Brien *et al.*, 2013). Therefore, this study includes a selection of anthropogenic chemicals as key indicators.

The literature highlights numerous parameters that need to be included in a comprehensive groundwater baseline of the study area. However, the current dataset does not included majority of the parameters identified in section 2.4.3. Hence, in order to quantify a baseline, new groundwater data of missing parameters will have to be collected. This will require the establishment of a proper monitoring network. Therefore the aim of this research is to develop a groundwater monitoring network for the purpose of collecting baseline data and detection monitoring.

2.5. Key insights from the reviewed literature

Shale gas development and hydraulic fracturing, the artificial stimulation technique, have recently been criticized for being an environmentally hazardous practice. Literature highlights hazards to shallow groundwater resources that include contamination by stray methane gas, infiltration of deep saline brines, contamination by NORM's, introduction of toxic drilling and hydraulic fracturing fluids. The chemical signature of these contaminants is considered key parameters of concern in this research.

To prevent such environmental concern in the Karoo a groundwater monitoring system is envisioned. However in realising this groundwater monitoring system baseline conditions for various groundwater characteristics, such as hydrogeochemistry and water quantity will need to be quantified. This is used to compare observations after shale gas development has taken place. However, it is apparent that to understand the temporal and spatial variations in groundwater characteristics requires a comprehensive understanding of the hydrogeology of the region. Much research has been done in this regard in the region and the hydrogeology of the shallow aquifer systems (>300m) is well understood.

A key function of a groundwater monitoring system of this nature is the ability to identify changes in the groundwater and relate this to a particular source. Indicators are usually employed that serve this function. Essentially, these are chemical signatures that are characteristic of a specific source of contamination. For shale gas development, plenty of work has been done to establish these chemical signatures, such as the use of stable isotopes to identify thermogenic methane, noble gas geochemistry to identify hydraulic fracturing fluids and field parameter signatures to identify stray methane gas. Biomonitoring also appears to work effectively as early warning indicators.

Lastly, the design of a groundwater monitoring system depends on the objective of the system. Location and construction of monitoring wells need to take into account the scale of the network, the hydrogeology, the nature of the pollutant, and the fate and transport of pollutants, among others. Monitoring networks must be optimized so that no redundant measurements are made. Ideally the most efficient tool to realise this function is the use of real-time on-line monitoring systems which relay in-situ data to data management software that predict changes in the groundwater. Such systems are known to be most efficient in preventing widespread contamination.

Chapter 3 : Research design and methods

3.1. Introduction

In Chapter 1 the research topic of this thesis was introduced, which is the development of a groundwater monitoring system, followed by Chapter 2, which dealt with the relevant literature to identify knowledge gaps. Here in Chapter 3 the focus is on describing the methods and tools used in data collection, data analysis, as well as those methods and tools used in Quality Control/Quality Assurance (QC/QA). This should constitute a blueprint for the application of this research in other similar settings. Firstly, section 3.2. provides a description of the study area and the sampling plan. Thereafter section 3.3. details the actual sampling methods employed. Section 3.4. describes the methods used to accomplish the objectives. Finally this chapter concludes with a description of the reliability of this research, validity of the results, ethical considerations of this research, and limitations experienced during this research.

3.2. Research design

The following section presents the overall strategy that was employed to integrate the various components of this research, in-order to address the research problem. From the literature a few questions must be answered to address the research problem: where to monitor, what parameters to monitor, how often to monitor these parameters, and what the pre-existing conditions of the aquifer is? Three objectives were set in section 1.5. in-order to answer these questions? These three objectives represent the basic component of a groundwater monitoring system. Hence, the completion and integration of the three objectives resulted in a groundwater monitoring system. This research can be classified as a descriptive study, which produced a description of an appropriate groundwater monitoring system according to the defined scope.

For objective 1 a quantitative approach was applied that relied on the use of data collected during this research, as well as historical data. Through the use of various techniques objective one described the number and position of monitoring points. Objective 2 follows a quantitative approach. Data collected during this research, as well as historical data was applied to the chosen methodology, to produce the baseline conditions of the aquifer. Objective 3 follows a qualitative approach. Here the literature was consulted in order to devise relevant monitoring parameters and a theoretical understanding of the groundwater flow was used to establish monitoring frequencies. This approach was used to answer the questions stated in the previous paragraph.

3.2.1. General description of case-study area

In section 1.6 of chapter 1, the regional scope of this research was illustrated. This represents the area of concern in the context of shale gas development for this study. However, the extent of this region was too large to apply the methodology in a manageable sense. Hence, a smaller case study area was chosen instead (Fig. 3.1.). This allowed focusing on smaller scale spatial variations, which benefitted the design of the monitoring network with respect to the spatial resolution.

The reasons behind the selection of the case study area related to the facts that the area was underlain by portions of the most prospective shale gas application areas (Petroleum Agency of South Africa, 2013), the local population (70 000 approx.) was dependent on groundwater sources (Statistics South Africa, 2017), and the current monitoring system in the study area was inadequate to provide the necessary level of data in the context of this study. Hence, this area might be susceptible to the risks associated with shale gas development described in section 2.3. Therefore, a comprehensive groundwater monitoring system was needed in the case study area, to protect vital groundwater resources.

The case study area is situated in the South Western Karoo, South Africa. It is considered a semi-arid region, with mean annual rainfall below 450 mm (Le Maitre *et al.*, 2009). Majority of the case study area is situated in the Upper Breede-Gouritz Catchment, with a portion extending into the Fish-Tsitsikamma Catchment. The current monitoring points (active sites) include 34 groundwater level monitoring points and 5 groundwater quality monitoring points. As a note figure 1.2 displays all the data points, however most of these points are inactive in terms of monitoring.

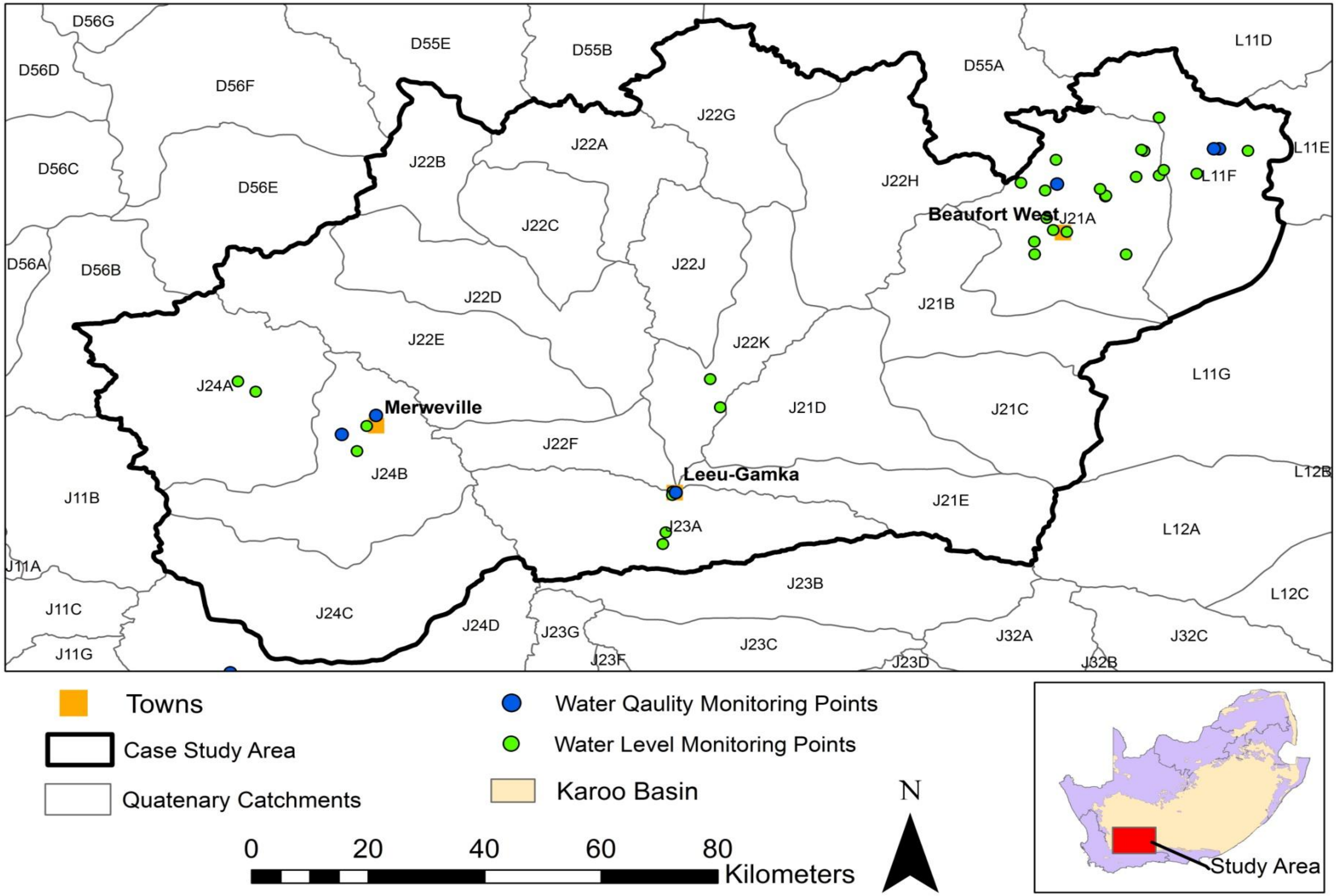


Figure 3.1. Map of the case study area, showing the various catchments as well as the current monitoring network. These points represent the data points used in this study.

3.2.2. General geology and hydrogeology of the case study area

Locally within the study area the geology is fairly simple (Fig. 3.2.). The underlying geology consists of rocks of the Adelaide Subgroup of the Beaufort Group, specifically the Abrahamskraal formation and the Teekloof Formation. The former is concentrated more towards the north and the latter is focused towards the south of the study area. The Abrahamskraal formation consist of alternating units of mudrock and very fine - medium lithofeldspathic sandstone (Woodford and Chevallier, 2002). The Teekloof Formation consists of mudrock alternating with minor laterally accreted lenticular sandstone units (van Wyk and Witthueser, 2011).

According to the Department of Water Affairs and Forestry (2002a, 2002b), in the study there are generally two types of aquifers:

- Fractured hard rock aquifers
- Mixed intergranular/fractured - alluvium & deeply weathered bedrock aquifers

In the study area the fractured hard rock aquifers are represented by the mudstone and fine grained sandstone deposits of the Adelaide Subgroup of the Beaufort Group (van Wyk and Witthueser, 2011). These rocks were deposited in meandering river floodplain system, hence the geometry of aquifers in the Beaufort group are complicated by lateral facies changers brought about by the meandering rivers (Botha *et al.*, 1998; Woodford and Chevallier, 2002). Consequently the aquifers in this study area can be multi-layered, multi-porous displaying anisotropic hydraulic properties both vertically and horizontally (Botha *et al.*, 1998; Woodford and Chevallier, 2002).

The heterogeneity of the hydrogeology of the aquifer can influence the spatial variability of parameters within the aquifer. Hence, this influenced the number and position of monitoring points. This idea was considered during the interpretation of the data in this research. In addition it may be wise to separate the approach based on each formation. However, the variations are manifested on a local scale, and at a regional scale the homogenous can be assumed. Hence, in this thesis the above formations are treated as one geological entity.

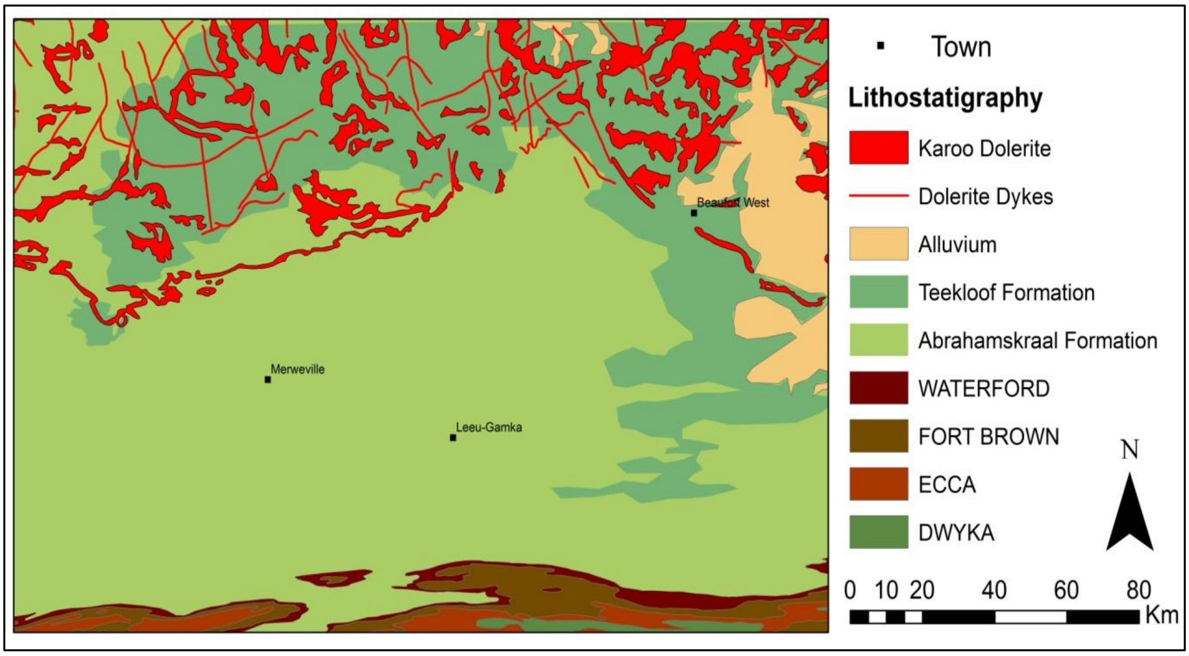


Figure 3.2. Simplified geology of the case study area.

The general perception is that Karoo aquifers are poor in nature and do not provide a sustainable source of fresh water. The primary porosity and permeability of Beaufort Group rocks (and most Karoo rocks), are significantly low, almost completely destroyed by diagenic processes (Woodford and Chevallier, 2002; van Wyk and Witthueser, 2011). Secondary porosity and permeability is provided by a complex network of fractures, due to jointing, folding, faulting & intrusions (Botha *et al.*, 1998; Pacome, 2010; Xu *et al.*, 2012). This fracture network controls the occurrence, flow and perhaps even the storage of groundwater (van Wyk and Witthueser, 2011).

However according to Botha *et al.* (1998) the aperture and areal extent of these fractures are not great enough to store large volumes of water, hence the matrix must be the main storage unit of Karoo aquifers. They main water bearing conduits are bedding-parallel fractures (Botha *et al.*, 1998; Woodford and Chevallier, 2002). Therefore Karoo aquifers have a component of matrix flow and bedding-parallel fracture flow, with the matrix feeding the fracture system (Botha *et al.*, 1998). However due to the small areal extent of the bedding-parallel fractures and the generally poor permeability of the matrix system delineates Karoo aquifers into very localised zones (Botha *et al.*, 1998; Woodford and Chevallier, 2002). This feature was considered during the design of the sampling method, in-order to establish representativeness (see section 3.3.1.).

Alluvial aquifers of significant areal extent are also present in the study area, and exist as large river terrace deposits of Quaternary age (Woodford and Chevallier, 2002; Barrow and Conrad, 2010). Borehole yields (<5 *l*/s) are not as good as those in fracture hard rock aquifers but they do represent large volumes of groundwater (if saturated), and may even provide a source of recharge for the underlying fractured aquifers (Woodford and Chevallier, 2002; Barrow and Conrad, 2010).

The geological description highlights various aquifers in the study area that having different properties. For example, the alluvial aquifers appear to be favourable aquifers for abstraction, while the sediments of the Beaufort group are only high yielding when boreholes intercept favourable fractures. This feature segregates the water resources of the study area into favourable zones in terms of yield and water quality. Favourable zone in this sense was considered as receptors for possible groundwater contamination. Hence, an understanding of this feature was included during the design of the groundwater monitoring system.

One of the most important factors that govern Karoo hydrogeology are the numerous dolerite dykes and sills that have intruded into Karoo Basin (Chevallier *et al.*, 2001). Doleritic material is impermeable in character, unless fractured, hence dykes and sills (especially dykes) act as barriers to groundwater movement (Chevallier *et al.*, 2001). However the dyke-sediment contact zones tend to be highly fractured, hence these form preferable targets for groundwater development. The yields in these zones are generally of a higher yield and often more sustainable than other target aquifers (Woodford and Chevallier, 2002; Murray *et al.*, 2012). Furthermore dyke structures are considered to be vertically extensive features that provide pathways for fluid movement from the deep geological formations to the shallow aquifer systems (Woodford and Chevallier, 2002; Murray *et al.*, 2015). On the other hand the complex variations in the geometry of dolerite sill structures, make them unfavourable features for groundwater development (Chevallier *et al.*, 2001; Woodford and Chevallier, 2002).

Both dolerite dykes and sill structures do not outcrop extensively throughout the study area (Fig. 3.2.). Dolerite intrusions appears to be a major feature north of the Great escarpment (Council for Geoscience, 1979, 1983) Sill formations occur in the north of study area forming cap rock for many of the Nuweveld mountains (Council for Geoscience, 1979, 1983). Dyke structures occur in the locality of Beaufort West, and do influence the groundwater flow considerably. Therefore, the position of monitoring points in relation to dolerite dykes, in particular, was incorporated in this research.

The main zone of recharge in the area appears to be the Nuweveld Mountains in the north of the study, were higher rainfall values are recorded (Department of Water Affairs and Foresry, 2002a, 2002b; Talma and Weaver, 2003; Rose, 2008). Generally Karoo fractured rock aquifers have poor rates of recharge compared to more favourable lithologies (Kirchner *et al.*, 1991; Department of Water Affairs and Forestry, 2006). From there the net regional groundwater flow direction is from north to south, which is believed to mimic the general topographic gradient of the study area (Talma and Weaver, 2003; Rose, 2008; Barrow and Conrad, 2010).

However, locally the groundwater flow directions is significantly affected by groundwater abstraction and the presence of geological structures such as dolerite dykes and sills (Chevallier *et al.*, 2001; Woodford and Chevallier, 2002; Rose, 2008; van Wyk and Witthueser, 2011). Especially the presence of dyke structures, which act as barriers to flow are known to compartmentalize the local Karoo aquifers (Rose, 2008). For example, in the Beaufort West area the groundwater flowing from the north is partitioned into 5 compartments by the presence of 4 dykes (Fig. 3.3). In the vicinity of the dykes groundwater flow is essentially parallel to their strike. The recharge zones are important in terms of baseline conditions as they represent the newest waters in the aquifers. Hence, recharge zones were considered during the design of the groundwater monitoring system. In addition, the movement of groundwater is important in understanding where and when to monitor (Department of Environmental Quality, 2014).

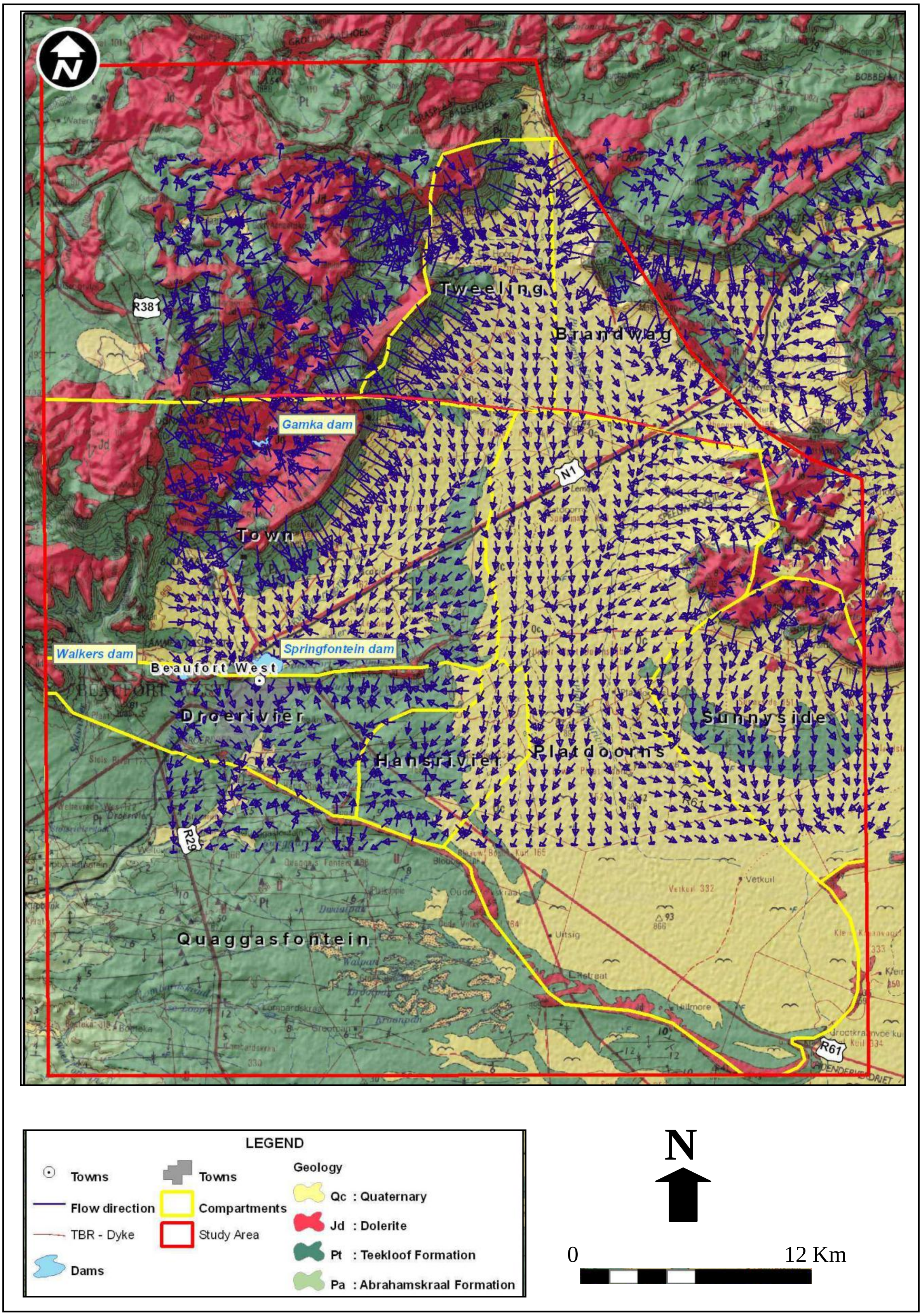


Figure 3.3. Groundwater flow direction and compartmentalization in the Beaufort West area (Rose, 2008).

The conceptual model illustrated in figure 3.4. describes the shallow groundwater system in the study area. Generally the deep groundwater system is separated from the shallow groundwater system by thick packages of impermeable sedimentary layers. However Rose (2008) and Xu *et al.* (2012) with the aid of isotope analysis, explain the possibility of mixing between shallow waters and deep waters along the Town dyke. Evidence is also provided by Murray *et al.* (2015) and Woodford and Chevallier (2002) to suggest mixing of deeper groundwater and the shallow aquifers waters in other localities of the study area, in the form of thermal springs. However this is not conclusive. An explanation of the deep hydrogeology is beyond the scope of this study. For the purpose of this study the environment of concern is the shallow aquifers were majority of the water resources are situated. Since thermal springs indicate instances of deep groundwater moving to the surfaces, they could also act as pathways for movement of contamination. Hence, thermal springs were considered in this research, during the design of the groundwater monitoring network.

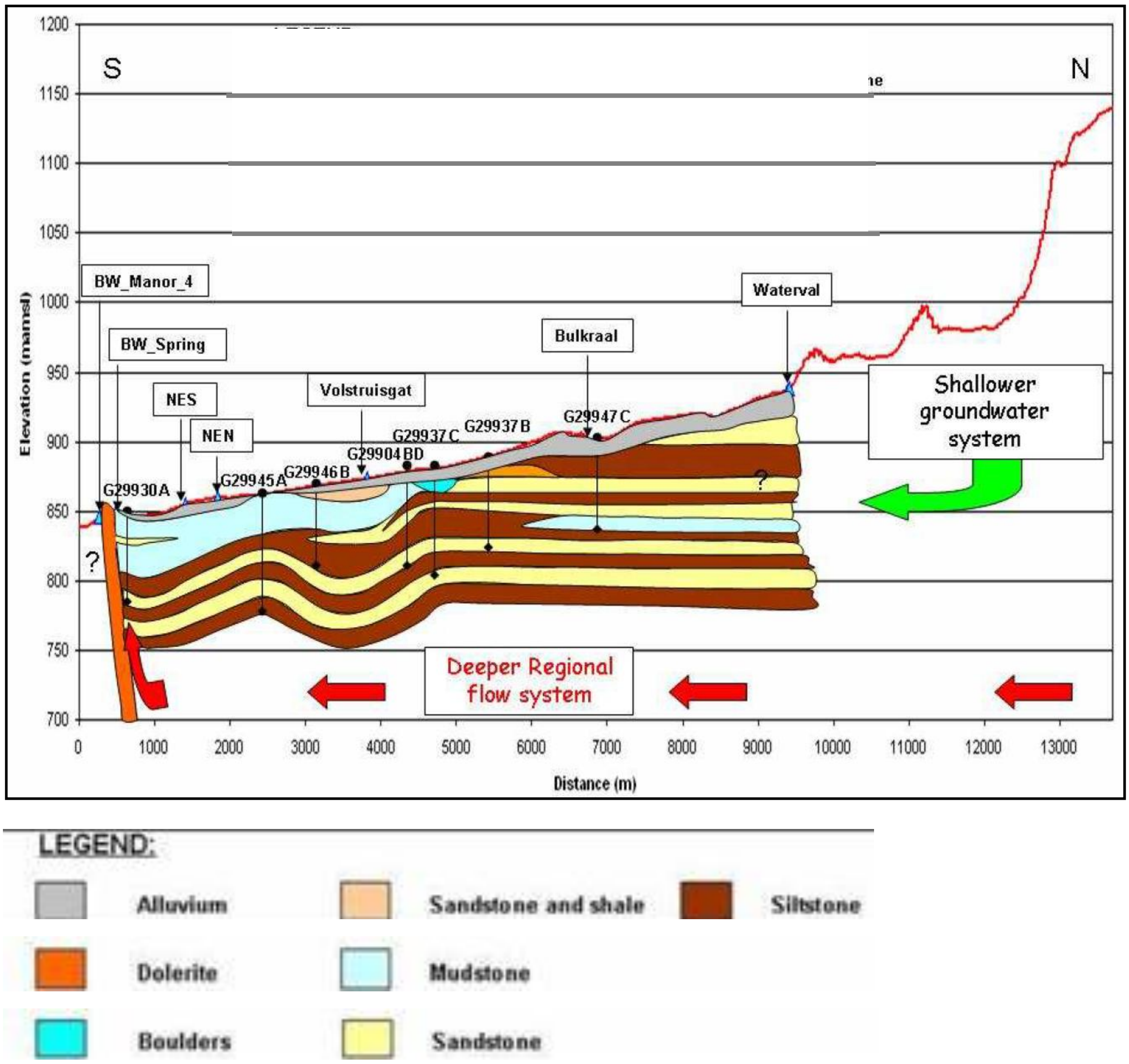


Figure 3.4. Conceptual model of the groundwater system in the Beaufort West area, from north to south (Rose, 2008)

3.2.3. Design of sampling procedure

Figure 3.1. displays the available data points/sample sites in the study area. These points are part of the Department of Water and Sanitation (DWS) monitoring network in the catchment. The monitoring network includes open holes that are maintained by the DWS, while the rest are either municipal or private holes, and 1 spring. A handful of the private and municipal borehole have pumps installed and are utilised as production holes. Majority of the sites are used to measure water levels (a total of 34 are active), while only five points in the case study area are used to measure water quality (Fig. 3.1). These sites can be considered target sampling sites as they intercept the shallow groundwater system and represent the current monitoring network in the region. Furthermore many of these sites have a long history of data attached to them, hence making them ideal for temporal analysis in particular. The historical data used in this research was collected from the monitoring points indicated in figure 3.1.

As part of this research a total of 25 points (only boreholes), as part of the existing network, were sampled over the course of two sampling runs. This was done to supplement the existing historical data, and to perform a field survey of the monitoring points. The first sampling runs were conducted in the summer month of April, while the second sampling run was conducted in the winter months of August through September. The purpose of which is to understand seasonal changes in the groundwater system. However, limitations experienced during the field sampling runs prevented 10 data points from being sampled (see section 3.7)

3.2.4. Parameters of concern during data collection and analysis

The following is list of parameters that were sampled:

1. Field parameters - electric conductivity, temperature & pH
2. Depth to static water level
3. Inorganic chemistry - This includes a selection of major anions and cations, which make up the majority of the compounds dissolved in natural groundwater (see chapter 5)

This list was represents the parameters of concern in order to achieve the objectives of this research. They are also important in terms of their relations to shale gas development, as identified in the literature. For example the chemistry is relevant when discussing the baseline conditions of the aquifer, while the depth to static water level is an important variable when designing a groundwater monitoring network.

3.2.5. Data types and sources

The data used in this research include both primary and secondary quantities data. Primary data was sourced through field sampling programs (see section 3.2.3), and includes the parameters mentioned in section 3.2.4. The historical or secondary data include time series data of the parameters in section 3.2.4. and was sourced from DWS. The DWS conduct regular sampling along the monitoring network indentified in figure 3.1. Relevant GIS and map data were also sourced from the Council for Geological Sciences (geology maps), DWS (hydrogeology maps), various internet websites, and private consultants. This included spatial data sets of auxiliary information used to support the data analysis process. Finally, all the sources of secondary data were acknowledged in this research.

3.3. Data collection methods

In the following section the available data collection methods, which specifically focuses on the groundwater sampling methodology, was detailed. Also the various tools and equipment used to collect the samples will be discussed. Followed by a description of the actual process undertaken in sample collection. For objectives 1 and 2 the following data collection methods apply. The data collected in the following procedure was used in both objective 1 and 2.

3.3.1. Groundwater sampling methods

There are various techniques that can be used to extract a representative sample of groundwater from a borehole. Ideally the choice of technique depends on what part of the groundwater system needs to be sampled. As mentioned earlier the shallow aquifers in the study area have a component of matrix flow and fracture flow, and very often the chemistry in these two features show differences (Murray *et al.*, 2015). These two components make up the majority of the groundwater flux into a borehole. The sampling techniques changed according to which of the component of the aquifer needed to be sampled.

The initial decision to make is to consider whether to purge the borehole or not. It is normally assumed that the water in the borehole has stagnated, hence does not represent the water in the aquifer (Weaver *et al.*, 2007). A well is purged in-order to remove this stagnant water and allow the inflow of fresh water from the aquifer. However, Gomo and Vermeulen (2015) determined that purging is not necessary in certain conditions, such as low yielding boreholes. In this situation, the method is to determine the natural flow zones within the well and pull a sample directly from this zone. In this zone, the water in the well is continually replenished. The flow zone will have to be first determined using downhole profiling techniques (Gomo and Vermeulen, 2015). The sampling equipment is lowered into the flow zones and a sample in extracted.

Natural flow zones in the Karoo can be represented by fractures or less rarely by high permeability sedimentary layers (Murray *et al.*, 2015). Sampling of flow zones will initially not allow the matrix component to be sampled, in fractured rock aquifers (Gomo and Vermeulen, 2015). Another method would be to purge the well, then lower the sampling equipment (in this case a low-flow sampling pump) at most 2 metres below the water level. A sample is then extracted. This technique does not draw a sample from a flow zone; instead the sample is a mixture of all waters entering the well below the pump intake. Essentially this represents both the matrix influx and the fracture flow, which can be considered a bulk sample of the aquifer waters.

There is another often used technique to withdraw a sample of groundwater from a well. This technique requires a passive grab sampler (such as a bailer or a specific depth sampler), and is different than sampling with a low-flow sampling pump. However once again the grab sampler must be lowered directly into a flow zone. The well will normally not be purged in this situation. This technique is often used to investigate vertical stratification of water in the well, and to grab and maintain samples at in-situ conditions.

It was decided to employ a technique that would allow for a bulk sample to be extracted from the wells. This was done due to the fact that the entire aquifer (the entire flux entering the borehole), should be characterized, and not just specific components of the groundwater flow. Also that groundwater resources of the region does not refer specifically to any one component of the system, but is regarded as the entire system at large.

Considering that a bulk sample will be collected, the sampling device that is required is a low-flow sampling pump. There is a wide range of low-flow groundwater sampling devices available on the market today, and each one has unique features that make it ideal for certain situations (Barcelona *et al.*, 1985; Johnston, 2007; Weaver *et al.*, 2007; Sundaram *et al.*, 2009). There is sufficient literature that can aid in the decision of which sampling device to use. It is however crucial to consider the parameters that need to be sampled when choosing a pump (Barcelona *et al.*, 1985; Johnston, 2007; Weaver *et al.*, 2007; Sundaram *et al.*, 2009)

When considering the list of parameters that is required for this research, a pump that allows for sampling a wide spectrum of parameters is needed. According the literature the most versatile sampling pump is a bladder pump. This is a positive displacement type pump, that relies on an inert gas to compress an internal bladder and push a volume of water to the surface (Schalla *et al.,* 2001). The volume of water never comes into contact with the pump material, and the bladder and tubing is composed of inert material, hence no interference. The bladder pump also performs the best when sampling for dissolved gasses and VOC's, as it allows for least amount of degassing (Barcelona *et al.*, 1985; Parker, 1994; Weaver *et al.*, 2007).

When it comes to material of the sampling pump and auxiliary equipment, Barcelona *et al.* (1985) explains that Teflon is the best choice, followed by stainless steel. These materials are the least reactive to sample water, and preserve the chemistry of the sample the most. In contrast to PVC material, which is not ideal for sampling organic parameters (such as VOC's) (Barcelona *et al.*, 1985).

When it comes to other low-flow sampling pumps, their properties are not as comprehensive as that of the bladder pump. There are considered not to be ideal for sampling of volatile compounds. However as mentioned by Parker (1994), more modern designs are proving to be almost as good as bladder pumps. This is especially true for submersible electric pumps.

Lastly a decision must be made whether to field filter samples or not. Field filtering is usually performed on samples that are used for trace metal analysis (Barcelona *et al.*, 1985; Weaver *et al.,* 2007). The idea is that large colloid particles (with metal species attached) and suspended particles are removed from the sample, thus only the dissolved constituents are analysed (Weaver *et al.*, 2007). Each method has its advantages and disadvantages, and choice depends on hydrological conditions, well bore conditions, and target analyte (Saar, 1997). By reducing turbulence during sample collection (low-flow) the amount of suspended particles entrained can be reduced. Also considering that trace metal analysis is not a big part of this thesis, and that laboratories pre-filter samples before analysis, it is not considered a necessary step.

3.3.2. Tools used during groundwater sampling

The following is a list of tools as equipment used during the sampling collection procedure:

1. Electric submersible pump (impeller type) + tubing
2. 12v car battery
3. Generator
4. 500ml LDPE bottle
5. Water level meter
6. Extech EC500 portable water quality meter
7. Martini field meters
8. Cooler box
9. 12v Portable refrigerator
10. GPS
11. Measuring tape
12. buckets of various volume

3.3.3. Procedure followed during groundwater sampling

The following procedure was adapted mainly from (Weaver *et al.*, 2007), with input from other sources such as Wilde (2005), Hackley et al. (2012), Xu *et al.*, (2012), Isotech Laboratories (2014).

Firstly upon arriving at a sample site, the borehole was surveyed. Measurement of the borehole diameter, collar height and borehole depth (where possible) was recorded. The condition and nature (production, monitoring, or private) of the borehole as well as any equipment installed in the borehole was noted. Lastly the elevation and coordinate location of the borehole was recorded, using a GPS. This was done to build a profile of each sampling site. An attempt was made to record all the data on prescribed data logging sheets, for the sake of organization.

There after any equipment (such as water level divers) was removed from the hole. Water level was recorded next. Using the static water level, the diameter of the hole and the depth of the holes, the volume of water in the borehole was calculated. According to Weaver et al. (2007) it is recommended to purge 3 volumes of borehole water, before a sample is taken. Initially it was attempted to purge 3 well volumes during sampling, however it soon became apparent that this would take too long using normally low-flow pumping devices. Instead the practice was to purge the holes for at least 30 min (Murray et al., 2015). Field parameters (EC, pH & Temperature) were continuously monitored during the purging process to establish when they have stabilized. Once the field parameters have stabilized (pH: ± 0.2 units; Temp.: ± 0.2°C; EC: ± 5%, for three consecutive readings) the samples were collected. All meters were calibrated every three days to ensure consistently accurate measurements.

For the purging and sampling phase the pump was installed 2 meters below the water surface, as per the recommendation of Weaver et al. (2007). If during the purging phase the drawdown was large, a recovery period was allowed for inflow of new water.

The first set of samples collected were the inorganic chemistry samples, which required filling a 500ml LDPE bottle. This was done by placing the discharge tube into the bottom of the bottle and allowing bottle to overflow. Hence no headspace was allowed. The samples were not acidified or filtered. All samples were stored on ice while in the field, and transferred to a refrigerator, were it was stored at 4°C.

3.4. Data analysis methods

In this section the various data analysis methods that were used is discussed. Each objective is discussed separately, and focused on the available methods, the software or tools used and the actual procedure employed to achieve the results.

3.4.1. Design of groundwater monitoring network

3.4.1.1. Methods available for groundwater monitoring network design

The objective of this groundwater monitoring network is two-fold, 1) to provide detection monitoring of contamination/deterioration of groundwater resources in the region, and 2) to acquire baseline data of the groundwater quality and quantity. The monitoring network must then be designed such that it could accomplish both these goals. Emphasis must now be placed on competing objectives, which often complicate monitoring network design (Spruill and Candela, 1990; Loaiciga *et al.*, 1992).

Early warning monitoring, which essentially detection is monitoring, is normally applied in and around and point and non-point sources of contamination. Downstream and Upstream monitoring points are typically placed to achieve this. However, currently in the study area there exist uncertainties in the groundwater movement and contaminant transport in the subsurface. Hence the correct placement of wells for early detection of contaminants may not be possible (Meyer *et al.,* 1994; Storck *et al.,* 1995). According to Storck *et al.* (1995) by placing far away from contaminant source an unnecessarily large portion of aquifer may be contaminated before detection. In contrast by placing wells close to contaminant source the contaminant may pass through well network undetected. Contaminant transport models allow us to resolve this situation; however this is beyond the scope of this research, and not to mention the complication caused by the described uncertainties (Meyer *et al.,* 1994; Storck *et al.,* 1995; Junez-Ferreira *et al.*, 2016).

Beyond this, there are generally two other methods for monitoring network design: 1) A hydrogeological approach and 2) a statistics based approach (Loaiciga *et al.*, 1992; Zhou *et al.*, 2013; Junez-Ferreira *et al.*, 2016). A hydrogeological approach relies on the quantitative and qualitative hydrogeological conditions at a site in order to determine the location and number of sampling points, as well as frequency of sampling. This approach is invariably based on a sound understanding of the hydrogeology and the intuition of the designer, which is advantages in that features of interest in the context can be prioritised (Yang *et al.*, 2008). However, the disadvantage is that in a statistical sense, data collection is biased towards the designer's preferences. Hence according the Loaiciga *et al.* (1992), a hydrogeological approach is best suited for site-specific monitoring networks where specific objective needs to be met, such as contaminant plume monitoring. Nonetheless hydrogeological approaches have been used on regional scale monitoring networks by Kim *et al.* (1995) and Zhou *et al.* (2013).

Statistical or geostatistical methods can be further subdivided into 3 categories: 1) simulation based techniques, 2) variance based techniques, and 3) probability based techniques (Loaiciga *et al.*, 1992). A broad discussion on geostatistical methods is beyond the scope of this paper, instead focus was placed on variance based techniques which were utilised in this research. In general all geostatistical techniques focus on understanding the spatial structure of a random (stochastic) regionalized variable (Spruill and Candela, 1990; Loaiciga *et al.*, 1992). Regionalized variables are those that have partially stochastic and partially deterministic character in space and time, such as hydraulic head, and are a result of inherent heterogeneities in the geological formation of concern (Spruill and Candela, 1990; Loaiciga *et al.*, 1992).

The spatial structure in geostatistics refers to the spatial autocorrelation of the data points, which is a measurement of how well neighbouring observations/monitoring points are correlated. The greater the spatial autocorrelation is the lower the degree of variability is for the regionalized variable (Spruill and Candela, 1990). The tool used to determine the spatial autocorrelation is the semivariogram, and is intrinsic to interpolation capabilities of geostatistical methods.

The variance based approach was incorporated in this research to monitoring network design. Many researchers have used this methods to design as well as optimize monitoring network on various scales, such as Olea (1984), Caeiro *et al.* (2003), Theodossiou and Latinopoulos (2006), Yang *et al.* (2008), and Bhat *et al.* (2015). Variance based methods in monitoring network design relies on determining the estimation variance for a particular sampling/monitoring pattern (Olea, 1984; Bhat *et al.*, 2015). The sampling/monitoring pattern with the lowest estimation variance is considered the most efficient (Olea, 1984). In existing monitoring network the addition or subtraction of monitoring points that reduces the estimation variance is considered an improvement in accuracy and thus a gain in information (Loaiciga *et al.*, 1992). Hence in such a way redundant monitoring points can removed, or necessary monitoring points added, to optimize the network to collect the most information for the least amount of monitoring points.

In variance based methods the most recognised tool for such analysis is kriging. Kriging is considered the best linear unbiased estimator, in that it will accurately re-estimate observation points (Theodossiou and Latinopoulos, 2006; Ahmadi and Sedghamiz, 2007). Kriging is used to interpolate stochastic regionalised variables at un-sampled locations (Kumar and Remadevi, 2006; Al-Mussawi, 2008; Chen *et al.*, 2016). Unlike other interpolation techniques, kriging takes into account the spatial structure of the regionalized variable (Kumar and Remadevi, 2006). Kriging applications are also advantageous as they provide the average estimation variance, which is a measure of the estimation error, which represents the accuracy of the interpolation (Kumar and Remadevi, 2006). These advantages are highly favourable to researchers and are used to generate the optimal sampling networks by minimizing the estimation error (Yang *et al.*, 2008; Bhat *et al.*, 2015; Qin *et al.*, 2016).

The empirical variogram model ($\gamma(h)$) is defined as one-half the average squared difference between values of a pairs of random variable separated by a given lag distance (Atkinson and Lloyd, 2007), and is calculated from the following equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

Where N (h) = the number of pairs separated by lag distance h, z (x_i) = measured variable value at point i, and z (x_i+h) = measured variable value at point i+h. A number of predefined mathematical models, such as spherical, Gaussian, exponential and pure nugget effect (linear), can be fitted to the experimental variogram (Ahmadi and Sedghamiz, 2007; Yang *et al.*, 2008). Coefficients of these models are used to assign optimal weights during kriging estimation. The most important coefficient in this application is the spatial autocorrelation range (a). The experimental variogram (3.10) reaches a plateau or sill (C) at a range. Points separated beyond this distance are considered to be spatially uncorrelated (Caeiro *et al.*, 2003).

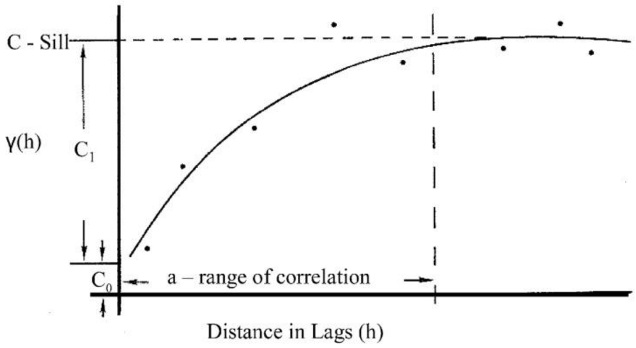


Figure 3.5. Example of experimental variogram, fitted with a corresponding mathematical model. C_0 is the nugget effect, C_1 is the sill range, and a is the spatial autocorrelation range (Caeiro *et al.*, 2003).

However, according to Loaiciga *et al.* (1992) variance-based geostatistical methods are not well equipped to handle design situations upon complex hydrogeological settings. Hence they are applied best to regional scale networks where hydrogeology can be simplified into more homogenous and isotropic blocks.

The simulation approach is based on generating multiple simulations for a regionalized variable, which should yield important statistical information. The overlay of various monitoring networks will allow examination of network efficiency (Loaiciga *et al.*, 1992). This approach works well for contaminant transport modelling, where the probability of detecting contamination can be determined for various network configurations, such as Meyer *et al.* (1994) and Storck et al. (1995). However simulation methods are computationally heavy and are limited in accuracy when limited knowledge of the aquifer is present (Loaiciga *et al.*, 1992). They are furthermore situated to local scale situations dealing in particular with detection monitoring.

In the probability based approach the design of monitoring points frequency of sampling is treated as a mathematical selection problem. Here researchers apply advanced mathematical and statistical problem solving algorithms to select critical monitoring points according to set of selection criteria rules, such as the probability of a monitoring point detecting high levels of a monitoring point (Loaiciga *et al.*, 1992; Li and Chan Hilton, 2005). Hence unlike variance based methods it can take into account both the accuracy (estimation variance) and the magnitude of the observation (Loaiciga *et al.*, 1992). This methodology can be applied to regional scale monitoring systems.

From the information presented in section 3.4.1.1. it is clear that a variance based methods is best suited as a design approach for this research. Simulation based techniques require advanced computational and mathematical skill that is beyond the scope of this study. Similarly probability based techniques require mathematical knowledge that is beyond the scope of this research. It is also apparent that no design approach is complete without the incorporation of a hydrogeological understanding. Therefore the monitoring network design approach utilises in this research follows a novel hybrid geostatistical (kriging based) hydrogeological approach. Similar methodologies have been applied by Olea (1984), Caeiro *et al.* (2003), and Bhat *et al.* (2015).

3.4.1.2. Tools and procedures used to design groundwater monitoring network

Although a groundwater early warning monitoring network would be the best option for detection monitoring, it will not be capable with current knowledge and resources. Instead the following procedure is designed to position (monitoring points) a regional scale monitoring/sampling network, strategically, to meet the objectives of the monitoring system.

Depth to static water level was selected as the regionalized variable, instead of water quality data. This was due to the fact that there are more water level data points in the study area. Sample size is an important factor during many statistical analyses. Historic data collected from the current monitoring network (Fig. 3.1), as well as data collected during this research were collated. The data represents observations for the 2014-2015 summer cycle (October - April). Data points that were deviating significantly from a parametric distribution, trending, or used as abstraction points were similarly removed from the dataset. The final dataset represents average depth to water values for the time period 2014-2015.

ArcGIS 10.3 was used to explore the spatial data structure of the current monitoring network, firstly by understanding the normality of the dataset using histograms and QQ plots. There after any spatial trends in the data was explored. This was done to guide the setup of the kriging tool. Variography was used to calculate the spatial autocorrelation range for the regionalized variable. This range was used to separate monitoring points within a particular sampling pattern. Using the kriging tool in ArcGIS 10.3, variogram models and prediction standard error maps were produced based a 4 combinations of setup parameters (Table 3.2). The setup parameters (lag class, lag interval, anisotropy) were based on recommended guidelines, as well as understanding of the regionalized variable. Thereafter the most accurate model is chosen by comparing the cross-validation results.

The chosen empirical variogram model and corresponding mathematical model is used determine the parameters of a, C and C_0. The spatial autocorrelation range is used to separate sampling points within a hexagonal sampling grid according to the recommendations of Olea (1984) and Caeiro *et al.* (2003).

Table 3.1. Setup parameters for models in the kriging tool.

Model no.	LC*	LCDI*(m)	Application
1	12	2086	Identifies small scale correlations, searches using mean water levels
2	12	1965	Small scale correlations, using log transformed mean water levels
3	15	3164	Average distance between points (recommended setting)
4	10	10 000	Searches for large scale correlations

**LC: number of lag classes used; LCDI: the width of each lag class*

In-order to incorporate a hydrogeological component to the monitoring network design, a list of important hydrogeological features was identified in the literature (Table 3.3). It was attempted to delineate these features in the study area. These features are considered high priority features in the context of the objectives of the monitoring network, such as protecting zones of favourable water resources or monitoring critical contaminant pathways.

Table 3.2. Criteria used for hydrogeological zone mapping.

	Hydrogeological features
1	Zones of high recharge (fresh/new groundwater)
2	Dolerite intrusions (contaminant pathways)
3	Thermal/deep seated spring (contaminant pathway)
4	Deep boreholes (contaminant pathway)
5	Zones of high aquifer yield (water resources)
6	Points of groundwater abstraction (water resources)
7	Zones of favourable groundwater quality (water resources)

The hexagonal sampling grid designed, is overlaid on the hydrogeological features map. Using the map features and the grid a single new monitoring point is located within a hexagon relative to a specific hydrogeological feature, for those hexagons that did not already have monitoring points in them. In such a way a systematic/stratified sampling approach is developed. Thereafter to determine if the new monitoring network provide better coverage and thus an increase in information across the case study area, the kriging estimation error was recalculated for the new network.

The procedure presented in section 3.4.1.2. is intended to develop a network that best incorporates the available resources. There is however another component of a monitoring network that must be discussed as well, the frequency of monitoring. This is to be covered later in the thesis.

3.4.2. Analysis for determining baseline characteristics of aquifer

3.4.2.1. Methods available to determine baseline characteristics

As highlighted by Shand *et al.* (2007), there is no unified approach to achieving a groundwater baseline. Various methods are available such a statistical analyses historical water quality data, comparison to associated areas of pristine water quality, and geochemical modelling, amongst others (Table 3.3.). The choice of methods would depend on the available data and the hydrogeological situation, for example comparing a site to an area of pristine water quality would depend the availability of such areas, as well as similarities in the environmental and hydrogeological systems (Shand *et al.*, 2007). In modern industrial time pristine areas are difficult to find, having all been to some degree effected by anthropogenic factors. In all likelihood an approach that incorporates more than one method appears to be the best solution.

Consideration of the choice of methods depends on the objectives of assessing the baseline characteristics. In exploration and monitoring of contaminants the concept of background, threshold and anomaly is applied to the understanding of the data (Reimann and Garrett, 2005). This allows incorporation into compliance and detection monitoring programmes.

Table 3.3. Methodological approaches used to establish baseline characteristics of an aquifer (Shand *et al.*, 2007).

Method	Application	Limitations
Assessment of historical groundwater quality data	Reveals any temporal trends that may reflect anthropogenic inputs	Long-term monitoring data are rare Range of solutes monitored is limited (more often for nitrate and organic compounds than other determinands) Data may be of variable quality Detection limits may have changed over time and may be too high to be of relevance
Comparison of up-gradient and cross-gradient groundwater quality	Allows spatial distributions in groundwater chemistry to be assessed relative to point sources of pollution	Not appropriate for areas affected by diffuse pollution or areas where pollution sources are not obvious (e.g. septic tanks) Care needed to ensure like-with-like comparison because of heterogeneity in aquifer lithology, geochemistry and groundwater flow regime
Comparison with similar geochemical environments	Comparison with same/similar aquifer in a different region or with confined section of the same aquifer	Regional heterogeneity in aquifer lithology, geochemistry and groundwater flow Comparison of unconfined groundwater chemistry with confined groundwater chemistry has problems because of differences imposed by groundwater residence time, mixing with old groundwater, differences in groundwater flow and redox changes
Geochemical modelling	Can provide thermodynamic support to hypotheses on groundwater chemical evolution and pollution, and quantitative models of the effects of mixing a baseline groundwater with a Pollutant	Requires some geochemical skill Requires sufficient thermodynamic data, understanding of aquifer flow regime and knowledge of end-member compositions to put into the models Biogeochemical reactions not necessarily at thermodynamic equilibrium
Statistics	Can distinguish ‘typical’ from ‘anomalous’ compositions and provides a useful summary of groundwater chemistry data for a given area	Needs to be carried out concurrently with a study of hydrogeochemical processes Needs an awareness that anomalous compositions can result from natural processes as well as pollution

Assessment of historical data is valuable and direct methods, but requires a long records dating back sometimes decades for certain parameters. Comparison to similar hydrogeological environments can be useful, but will depend on similarities in the aquifers. Errors may arise due to employing on assumptions in the complexity and operation of aquifers. Comparison to pristine areas is also relying on understanding of complexities in aquifer proprieties. Hydrogeochemical modelling methods are powerful tools that allow investigation of the spatial evolution of the groundwater chemistry, but require numerous supporting data. Often this method may reveal patterns that are not obvious when scrutinizing the data itself. Statistical methods are invaluable is baseline assessment and are almost always incorporated in such studies. However, an understanding of the intrinsic structure of the dataset is required, such as normality or variability, for proper interpretation.

In dealing with the data, a statistical methodology to analyse spatial and temporal trends in water quality is well suited. It is fundamental to incorporate at least a basic form of statistical analysis in groundwater baseline research especially when dealing with large datasets. Statistical methodologies that are rigorous and comprehensive have been well documented (Naval Facilities Engineering Command, 2003; U.S. Environmental Protection Agency, 2009; Department of Environmental Quality, 2014).

These reports are fairly similar in there methodology. The general process of statistical assessing temporal and spatial groundwater quality data involve the following steps: 1) Determine if dataset is large enough to proceed with methodology and perform quality assessment of the dataset, 2) Perform exploratory data analysis, 3) evaluate the statistical independence of the data set, 4) determine if dataset is parametric or non-parametric, and transform dataset if needed, 5) select tools and procedure appropriate for the distribution, 6) determine the presence of seasonal and long-term trends in the dataset and if needed set the data stationary in time, 8) evaluate for spatial independence in the dataset, & 9) determine the background groundwater quality (Gilbert, 1987; Naval Facilities Engineering Command, 2003; U.S. Environmental Protection Agency, 2009; Department of Environmental Quality, 2014).

In this research a statistical approach to understanding the water quality and water level data is applied. While still relating results to the general hydrogeological understanding. In this we do not neglect to inform the statistics with classical science of hydrogeology. Other methods do not suit the available data such as comparison to pristine areas, or hydrogeochemical modelling.

3.4.2.2. Tools and procedures used to determine baseline characteristics

A statistical methodology was applied according to the strategy followed by The Department of Environmental Quality (2014). Figure 3.7 graphically displays the process. This was intended to bring insight into the aquifer characteristics according to a review of the historic data. Furthermore because the baseline developed here is intended to support the monitoring network in detection monitoring, the procedure is also chosen to provide relevant threshold limits (UCL). This will allow anomalous changes in water quality to be detected during monitoring operations.

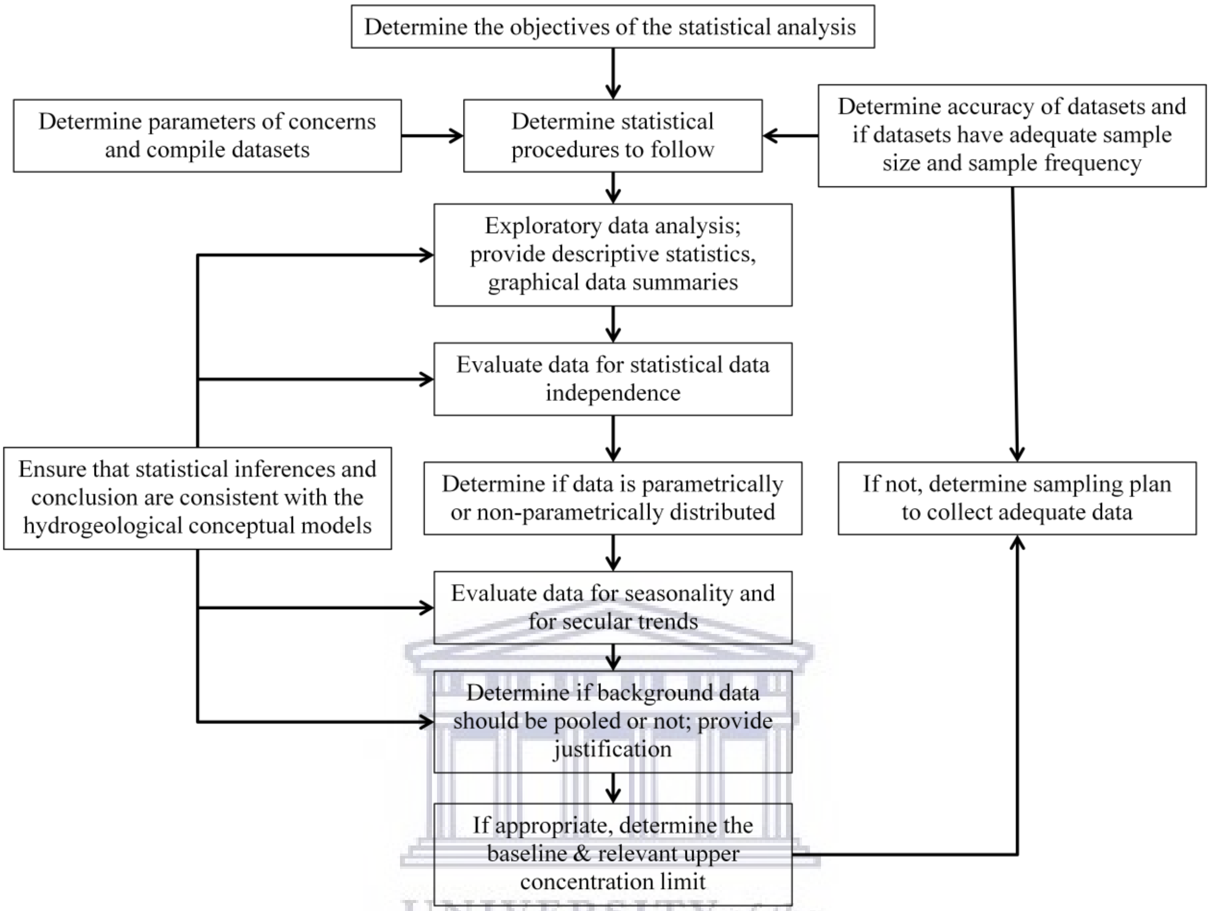


Figure 3.6. Flow-chart process followed to establish the relevant baselines for the study area (adapted from the Department of Environmental Quality, 2014).

Using the approach in figure 3.6. this research proceeded with an assessment of water quality data. Water quality data points are limited in number, and are even fewer for those points that have a long historic record. However, assessing the dataset for adequacy was part of the approach. The initial step in this regard was to determine the sample size as well as the time-scale between samples. The minimum recommended sample size for a parametric data set this is 12 independent samples, however a sample size of 20 - 30 was more ideal. This number greatly increases if the data set was non-parametric. Those points that did not meet requirements necessary to apply the statistical tests in figure 3.6. were not included in the analysis.

Using IBM-SPSS Statistical software, this research started off with an exploratory data analysis. This included a review of descriptive statistics such as mean, median and standard deviation. This allowed an initial interpretation of the dataset and its intrinsic structure.

Graphical summaries of the data were also included, such as histograms, scatter plots, and box and whisker plots.

Thereafter, this research reviewed the datasets for statistical independence, which was considered a critical step during statistical analysis. The basic assumption for temporal independence was that sufficient time has passed for new groundwater to flow past the observation point. Temporal independence between samples was reviewed by using the quarterly-rule-of-thumb, were observation points are collected to close to each other are considered dependant. These observations were removed from the dataset. For spatial independence geostatistical semi-variograms can be calculated, however the results tend to be inaccurate for small sample sizes. Here, simply a review of the hydrogeology and the separation distance between observation points were taken into account to determine spatial independence.

The next step was then to determine the distribution on the datasets that will be used further. This consideration was important to determine the proceeding statistical tools. Using the Shapiro-Wilk test, this research determined if the various datasets are normal or any other theoretical (parametric) distribution. In this procedure it was concluded that the dataset does not have the form of any parametric distribution, if the null hypothesis was rejected at the 90% confidence level. Data sets that were not parametric in any way, were analysed with non-parametric statistical tests during further analysis.

Next it was determined if any of the datasets displayed seasonality. That was to determine if there was any statistical difference in the observations of the various seasons, as delineated. This can provide information on the temporal stationarity of the dataset, which was an important requirement for statistical analysis. This analysis was performed using the Kruskal-Wallis test, which test whether the means of different grouped data (seasons) was different. If any seasonality was found in the dataset, and attempt was made to remove it using a prescribed method.

Once seasonality was accounted for, the long term trends in the data were analysed. This is known as secular trends, and was critical to final assessment of baseline values. If a secular trend exists then setting a baseline value was invalid. The Mann-Kendall test for stationarity, which was part of the EPA ProUCL 5.1 software package, was applied, using deseasonalized data.

The penultimate step involved a decision to pooled data from various data points. This decision was justified using the conceptual model, as well as statistical difference or similarities in the datasets. Hence, therefore summary statistics was again calculated after the above procedures were carried out.

Finally if the dataset displayed was adequate (no seasonal effects or secular trends), an upper concentration limit (UCL) was calculated. This was essentially a decision threshold to determine if contamination or degradation has occurred. For parametric dataset the UCL was set as:

$$\mathrm{UCL} = x + Ks$$

Where x is the mean, K is a constant depending on confidence level and sample size, and s is the standard deviation. For non-parametric datasets the UCL was set at the highest value in the sample set, according to a desired confidence level.

3.4.3. Determining monitoring parameters and frequency

3.4.3.1. Methods for determining monitoring parameters and frequency of monitoring

The first aim of this objective was to identify the parameters that must be sampled by the designed monitoring system. Ideally these parameters should meet the criteria that were highlighted in the literature review. They should be those parameters that are used to classify the aquifer properties (quality and quantity). They should also include parameters that can indicate changes in the aquifer properties and those that can indentify contamination from shale gas activities.

Unfortunately there was no empirical methodology in the literature that could be employed to achieve this goal. Instead an objective approach was utilised which relies on the literature review and a theoretical understanding of hydrogeochemistry.

The second aim of this objective was to determine the required frequency of sampling for baseline and detection monitoring. Fortunately the literature does provide methodologies that could be used to determine the best sampling frequencies used to meet the objectives of the monitoring system. These methodologies were based on the idea that unnecessary sampling is not performed and that non-redundant samples are collected. This saves money and ensures efficiency. Also to low a sampling frequency can result in a loss of information (Zhou, 1996).

Limited literature can be found on the topic of sampling frequency in groundwater monitoring systems. From what was available there are generally two methodologies applied: 1) a statistical approach (Swain and Sonenshein, 1994; Zhou, 1996; Department of Environmental Quality, 2014) 2) a hydrogeological approach (Vrba and Adams, 2008; Subcommittee on groundwater, 2013; Department of Environmental Quality, 2014). For a statistical approach the temporal autocorrelation range for a given time-series is calculated in order to determine temporal redundancy (over-sampling) (Nunes et al., 2004). Hence tools such as kriging and regression analysis are typical used. Sampling within the temporal autocorrelation range is considered redundant. For a hydrogeological approach, factors such as aquifer type, groundwater flow, recharge rates, abstraction rates, and climatic conditions are taken into account. This is based on the idea that a new volume of has passed the sampling point between sampling events (Department of Environmental Quality, 2014).

The statistical approach relies on a long time-series of data, with sufficient temporal resolution. The calculation must be applied to each monitoring points (intrawell), and so can be a tedious process. However this approach may provide the most efficient sampling frequency. The hydrogeological is less time-consuming and relies on a conceptual understanding of the aquifer and groundwater movement. However relies largely on an estimation of the required sampling interval.

However as mentioned previously the most appropriate monitoring frequency for detection monitoring is continuous (real-time) (Gullick et al., 2003; Storey et al., 2011). The on-line real-time approach to this feature has been implemented successfully before (Center for Energy and Water Sustainability, 2014) and allows rapid detection of changes in water quality and quantity. In any case due to the large number of monitoring parameters which are relevant to this study, a cost effective sampling scheme will have to be developed (Subcommittee on groundwater, 2013).

3.4.3.2. Procedure followed to determine monitoring parameters and frequency

In order to identify the monitoring parameters a literature review of relevant studies was carried out. Any chemical compound that was used to delineate groundwater quality, or could act as a tracer in a shale gas environment, was included. In addition, any chemical parameters that were important for water resource protection were also included. A table of parameters (Table 6.1.) was tabulated and categorised according chemical group (e.g. inorganic ions, gasses, isotopes).

For determining the appropriate monitoring frequency, a hydrogeological approach was used. An understanding of the aquifer characteristics and the movement of groundwater was used to determine baseline sampling frequency. For incorporating detection monitoring into a cost effective sampling scheme, different sampling frequencies are envisaged for the various chemical groups depending on the complexity and cost of sampling.

3.5. Quality Control

The following sections deals with the methods used to ensure that the data is of a good quality, and presents and argument on the validity of the results achieved. Furthermore we explore any ethical considerations and document the limitations experienced.

3.5.1. Reliability

Only standard methods were used to collect data. Minor modification made to procedures (where necessary) were appropriate and were applied in previous published peer-reviewed research (Barcelona *et al.*, 1985; Weaver *et al.*, 2007; Murray *et al.*, 2015). Every effort was made to ensure cross contamination did not occur, such as rinsing all equipment before each use, and rinsing of sample containers just prior to sample collection. Internal decontamination of pump equipment could not be performed in the field, however the purge process is considered to rinse the internal mechanics before sample collection. All equipment was calibrated regularity to ensure consistent readings.

Laboratory quality control included duplicate samples (under pseudonyms) in the sample batch. Thereby allowing the reliability of lab protocols to be tested. Each lab ensued standard methods available in the literature were used to analyse samples. Samples were further subjected to Cation/Anion charge balance calculations. Hence the results can be considered to be reliable. All data was checked for errors using statistical techniques, and erroneous values were removed. First, all nil values in the data set were removed. Thereafter, all anomalous readings (those outliers considered abnormal in time compared to closest neighbouring observations) were identified and removed.

3.5.2. Validity

Considering internal validity, the outcome of this research was to determine the configuration of a groundwater monitoring system that would mitigate risk associated with shale gas development. The type of data used in this research, the methods and tools used to collect and analyse this data are consistent with peer reviewed literature pertaining to the objectives of this research. The results achieved in this research do describe the configuration of a monitoring system that can be employed to solve the research problem of this research. Thus it appears that the results have high internal validity. With regard to external validity, the context and environmental setting of this research makes generalization difficult. The results cannot necessarily be validated against the results of other studies on this topic, but a discussion of results may yield similar trends compared to previous studies for individual objectives. Hence, overall the results of this research have low external validity.

3.6. Statement on research integrity

In the execution of this research no harm was done on any part of the environment or any entity therein. The data collection methods are considered to be acceptable according to current standards. Consideration was given to the social implications of shale gas development, by informing relevant parties (land owners, stakeholders, farmers unions etc) about the context of this research and its intentions. No participant was coerced into participating in this research. All participants were treated fairly and with dignity throughout the course of this research. Furthermore, all the information presented in this research has not been fabricated in any way. The information is purely factual and where appropriate, educated assumptions are made. Lastly this research has been conducted independently and in full by the author.

3.7. Limitations of the study

In the completion of this research a few limitations were faced in achieving the desired outcome. Firstly, technical difficulties were faced during the sampling procedure. The sampling equipment was not sufficient to collect groundwater samples from all the available monitoring points. However, the missing data points did not significantly interfere with the analysis. Secondly, ensuring that integrity was maintained by not accessing land where permission was not granted, 10 data points could not be sampled. Here, secondary data was collected from available sources.

Chapter 4 : Design of groundwater monitoring network

4.1. Introduction

Chapter 4 presents and discusses results from objective 1, which focused on the spatial design of monitoring network (i.e. the location of monitoring points). This is one of the components of a holistic design for a competent groundwater monitoring system. Baselines, monitoring parameters and frequency was dealt with in Chapter 5 and 6, respectively.

4.2. Exploratory geostatistical analysis

Figure 4.1 indicates the monitoring points and the corresponding depth to water at each point. A total of 34 monitoring points were used for the analysis. 2 points were removed from dataset. These removed points (boreholes) were either used for water abstraction or the wells had collapsed. Hence, the depth to water observed in these boreholes did not display natural values, and were removed.

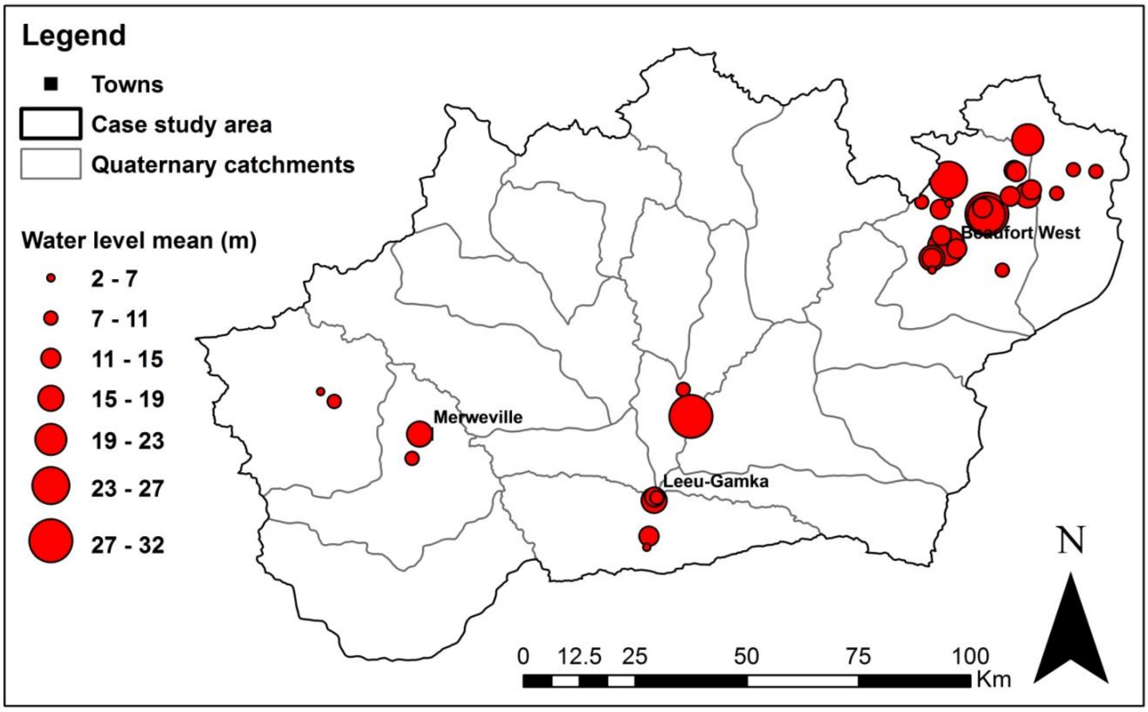


Figure 4.1. Mean depth to water for current monitoring points in the study area.

Exploratory statistical analysis and spatial analysis of the dataset, was critical during the setup and interpretation of geostatistical analysis. For example, the assumption of stationarity within the dataset (constant mean and variance across the study area) was a requirement in-order to reduce the error in the results. Figure 4.1 illustrates mean depth to water for each of the current monitoring point in study area. It was intended to indicate the spatial variability of depth to water across the case study area. The figure illustrates that the mean depth to water varies considerably across the study, from a maximum of 31.26m to a minimum of 2.98m (Table 4.1). Even boreholes in close proximity display completely different depth to water levels, as can be seen in the area surrounding Beaufort West. The global variance of the dataset presented in table 4.1 confirms a highly variable dataset, at 47.45 m^2. In addition, in the north east area around Beaufort West, depths to water levels are on average greater than in the central parts of the study area.

Table 4.1. Summary statistics for depth to water for current monitoring points.

Statistics	Mean depth to water	Natural logarithm
Min (m)	2.98	1.10
Max (m)	31.26	3.44
Range (m)	28.28	2.35
Arithmetic Mean (m)	13.72	2.50
Median (m)	12.82	2.55
Standard Deviation (m)	6.89	0.52
Variance (m^2)	47.45	0.27
Skewness	0.98253	-0.38485
Kurtosis	0.59773	0.51084

In addition, table 4.1. presents standard summary statistics which was used to assess the normality of the dataset. Normally spread datasets are not a requirement for geostatistical analysis; however normal datasets are more favourable during statistical analysis. The range of values presented in the dataset was extreme, and was consequently spread far off the mean, as indicated by a high standard deviation, when compared to the mean. The mean and median deviate from one another, which is indicative of a non-normal dataset. From the histogram in figure 4.2 (A), it can be seen that the spread of the data extends towards the right, which could be a result of outliers. Indeed a positively skewed right dataset was confirmed by the skewness factor in table 4.1. Figure 4.3 (A), which is a Q-Q plot for normality clearly indicates a dataset that was deviating from the normal line. Thus the evidence suggests that the dataset was not normally distributed. Hence, higher error was expected with the use of mean depth to water values.

For comparison the natural logarithm of the dataset was also analysed for normality. It is acceptable to perform the following geostatistical analysis using transformed values. Table 4.1 shows that the natural logarithm values were better distributed, with lower overall variance and a Skewness value closer to zero. This distribution was graphically displayed in figure 4.2. (B). However, from figure 4.3. (B) it can be seen that points do not lie on the normal line; hence the natural logarithm of the dataset was also not normally distributed.

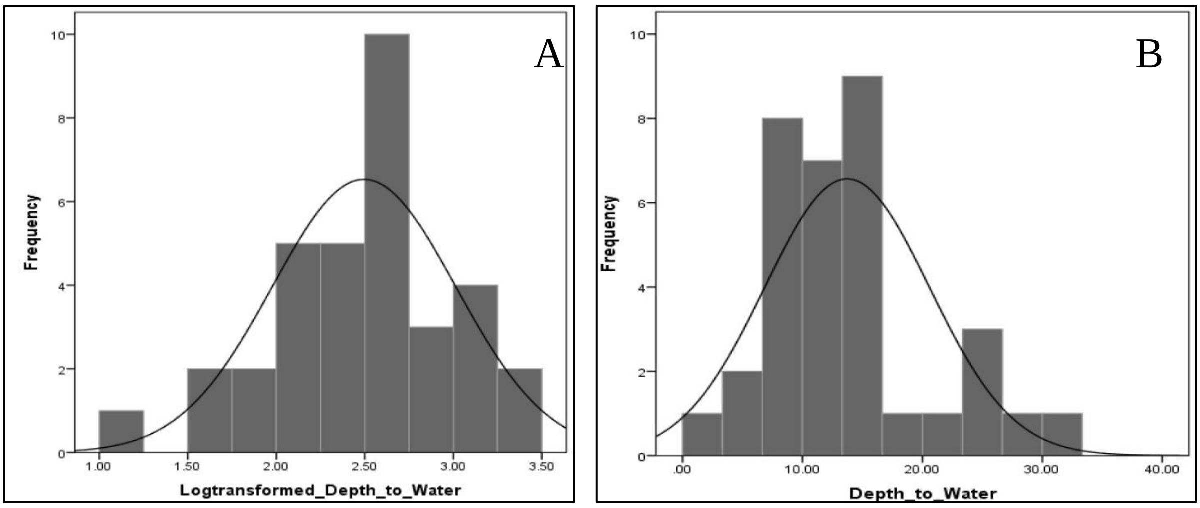


Figure 4.2. Histogram for mean depth to water observations (A) and log-transformed depth to water (B), for the current monitoring system. The normal curve is also show in order to evaluate normality of the dataset.

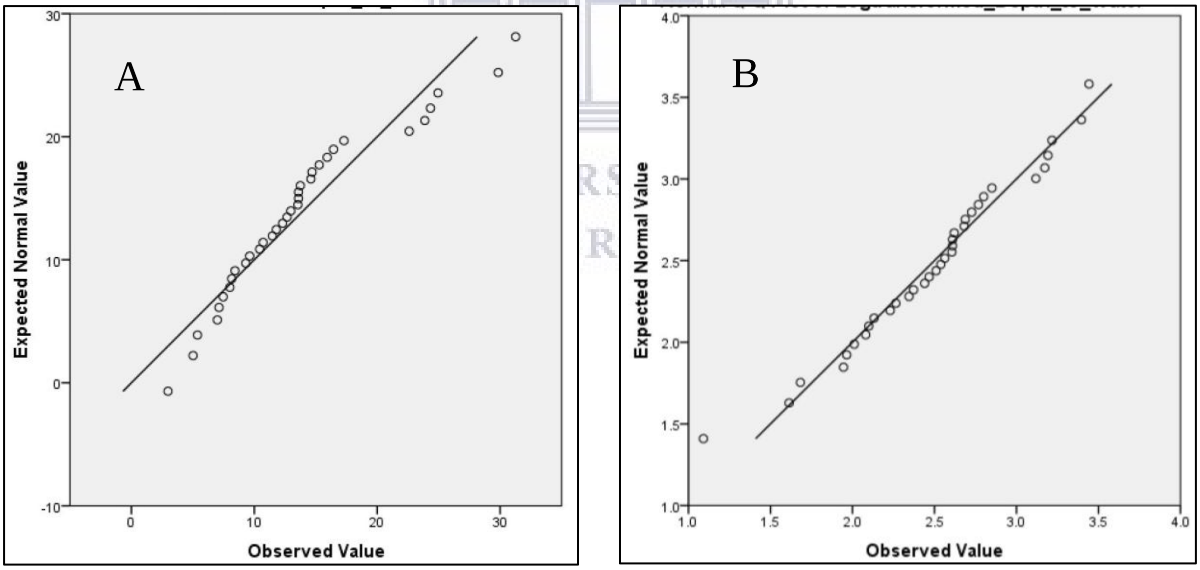


Figure 4.3. Normal Q-Q plots of mean depth to water (A) and log-transformed depth to water (B) values, for the current monitoring network.

Lastly, datasets was investigated for any spatial trends. That is a preferential trend such as increasing or decreasing values from one location to another. Non-trending data is a requirement for the following geostatistical analysis. Figure 4.4 shows the trends analysis block graph for actual values in the dataset, while figure 4.5. shows the trend analysis block graph for natural logarithm of the dataset. The floor of the graphs represents the study area with the data points plotted as spikes. The height of each spike represents the corresponding value. The trend was measured in both the 0^0 directions (A) and the 90^0 direction (B). The data points are mirrored on the side walls and a trend line displayed (green line). The trends line figure 4.4 (A) clearly increases towards the centre, from left to right, then decreases away from the centre. Similarly for the 90^0 direction, figure 4.4 (B - blue line), the trend line increases toward the centre then decreases. This was an indication of a numerical trend in the data. There appears that values were increasing towards a location just right of the centre of the study area. The shape of this trend was mathematically modelled as a second-order polynomial, which was then automatically removed from the dataset.

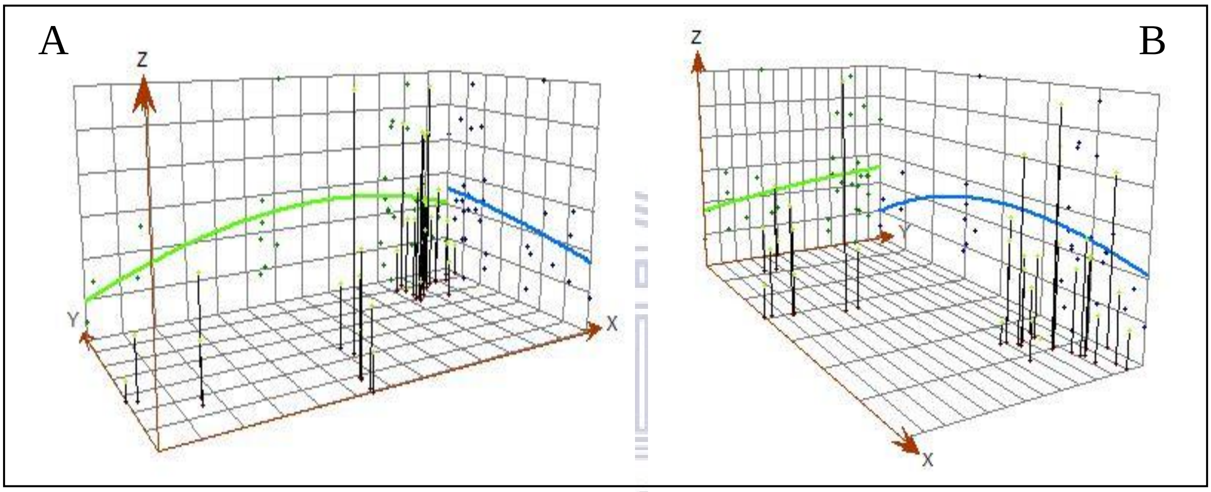


Figure 4.4. Trend analysis graph for mean depth to water observations in the study area. Figure A displays the calculated trend for the 0^0 direction while figure B is rotated 90^0.

Once again the transformed data set was also investigated for trend. As can be seen in figure 4.5 there was a clear trend in the transformed dataset as well. The shapes of A and B figures are almost similar to figure 4.4. but not as pronounced. That indicated an increasing trend towards the centre of the study area for the transformed data. This appeared to confirm the results of the actual dataset.

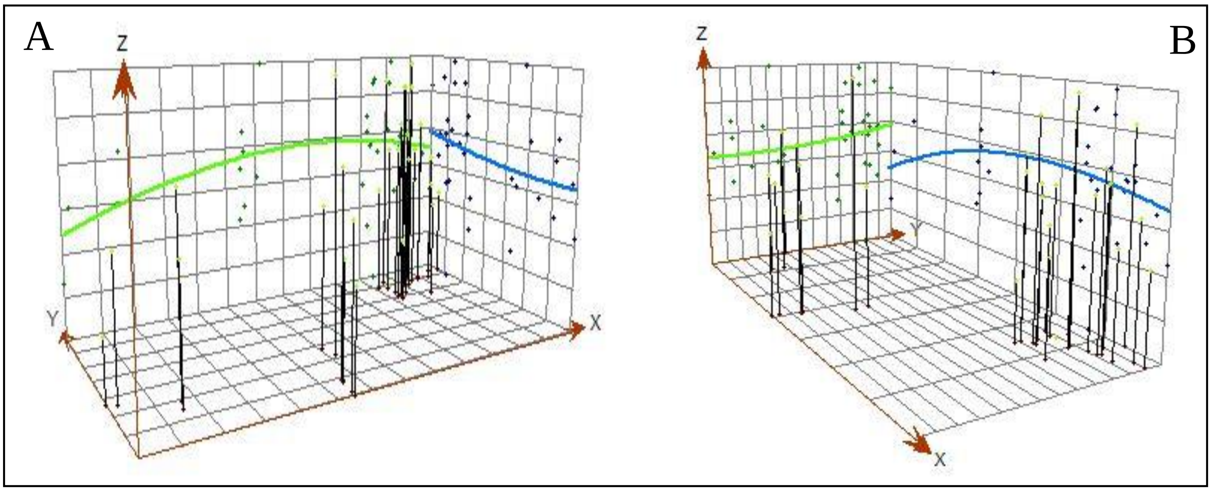


Figure 4.5. Trend analysis graphs for log-transformed depth to water values. Figure A is for the 0^0 direction, while figure B is for the 90^0 direction.

In summary, the dataset showed a high degree of variability, and was neither normally distributed nor log normal for that matter. Consequently, this may have indirectly indicated that the dataset was non-stationary. Both the original dataset and the transformed dataset showed clear evidence of having a second-order polynomial trend. The proceeding kriging analysis was then performed at low confidence and with high error expected to determine the spatial autocorrelation range.

4.3. Determining spatial autocorrelation

Four kriging models were calculated using various combinations of input parameters. As mentioned these parameters govern the number of data points used to estimate the model (i.e. sample size). Table 4.2 show the results of the 4 models. The sill represents the variance of the sample set determined by the input parameters, while the nugget effect represents the small scale variations or even intrinsic error in the dataset. The range value (spatial autocorrelation range) is the target parameter of this analysis, and the greater this value the more similar neighbouring data points are.

Table 4.2. Results of kriging analysis.

Model No.	Theoretical model	Nugget effect(C)	Sill ($C+C_0$)	Range (a)
1	Stable	0.0000	50.9317	16688.43 m
2	Stable	0.0138	00.2438	4957.13 m
3	Stable	7.6700	37.0461	6919.98 m
4	Stable	4.7791	41.7443	5998.13 m

Model 1 was designed to investigate the spatial autocorrelation at a local scale. The sill value of model 1 was higher than the global variance of the dataset. This indicated that at local scale the variability is higher than for the entire study area. Model 1 produces an autocorrelation range of 16688.43 m. This model indirectly suggested a high correlation between neighbouring data points. Model 2 was performed using transformed mean depth to water level values, and was set-up to investigate the autocorrelation at local scale. This model produces a nugget effect which was small and can be neglected. The variance of model 2 was smaller than the global variance of transformed dataset. Model 2 also produces a small autocorrelation range which suggests limited correlation between neighbouring data points.

Model 3 was set-up to investigate the autocorrelation at a local scale, however, using the average distance between monitoring points as an in-put parameter. Model 3 produced a large nugget effect, which could indicate a high variability at a very small scale. The model variance was significantly smaller than the global variance. While the autocorrelation range was considerable smaller than model 1, which concurs well with rest of the model. For model 4 the in-put parameters were set to investigate the autocorrelation at a large scale. The nugget effect produced by this model was lower than model 3, which was expected when working at large scales. The variance of model 4 was closer to the global variance which was expected as well, as the model take into account almost all the data points. The autocorrelation range was similar to models 2 and 3.

In order to determine the most appropriate model to use in the design of the monitoring network, cross validation techniques were applied. Cross validation is a simple technique that was used to determine the accuracy of the kriging model. The software performs this check by removing an individual data point from the model and then comparing the estimated value at that point to the real value. This process was repeated for all data points, and if estimated values are equal to real values then the model was said to be accurate.

Table 4.3 shows the results of the cross validation analysis performed on the kriging models. To judge the accuracy of a kriging model:

- the predictions must be unbiased indicated by a mean prediction error close to zero
- the standard errors must be accurate, indicated by a root-mean-square standardized prediction error close to 1, and
- the predictions do not deviate much from the real value, indicated by a root-mean-square error and average standard error that are as small as possible

Table 4.3. Results of the cross-validation analysis.

Model No.	Mean	Root-Mean-Square	Mean Standardized	Root-Mean-Square Standardized	Average Standard Error
1	0.2830	7.0457	0.0259	1.0283	6.7665
2	0.7854	7.4563	-0.0520	1.0030	7.7581
3	0.4723	7.2474	0.0471	1.0798	6.5688
4	0.5027	7.1719	0.0488	1.0619	6.6516

Evaluating the models according to the above criteria, model 1 had a significantly lower mean prediction error compared to other models. Model 2 produced the best root-mean-square standardized error. However, the average standard error for model 2 was substantially higher than the other models. Model 3 produced the best average standard error. It should be noted that all the models produced high average standard errors, which could be related to the statistical weakness in the dataset. Model 1 also produced the lowest root-mean-square error. The ability to accurately estimate unknown data values was a vital criterion; and so on this basis model 2 can be eliminated. By the same convention model 3 would then be the most ideal. However, the other criteria considerable favour model 1. Considering that the average standard error for model 1 was only slightly above model 3 and that the other criteria favour model 1, model 1 appeared to be the most appropriate.

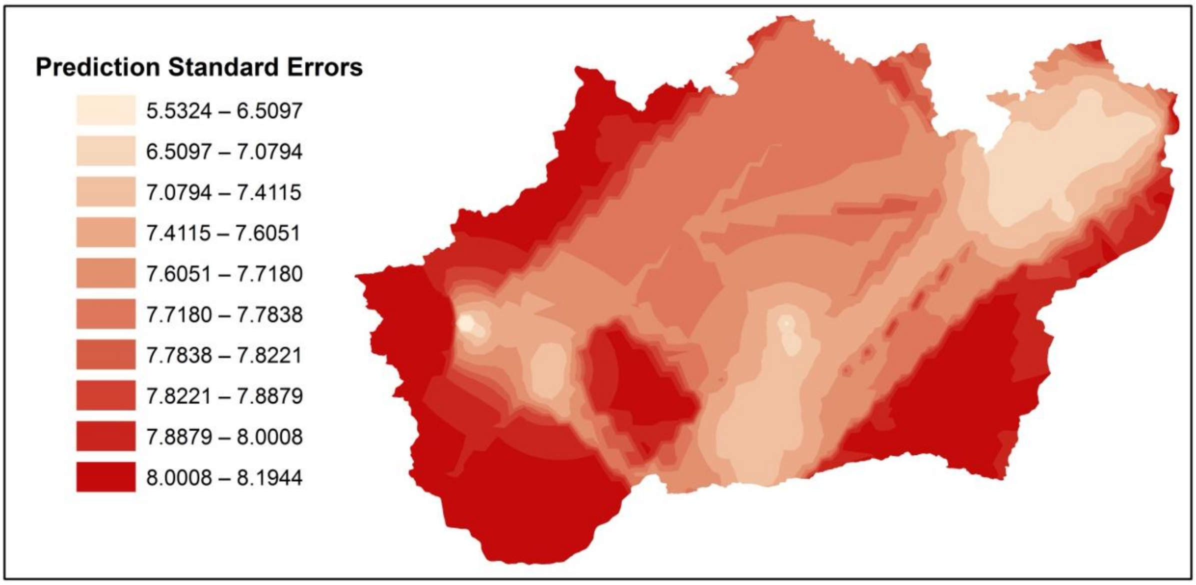


Figure 4.6. Map of the prediction standard errors according to the kriging model 1.

From the results of model 1, a prediction standard error map was generated (Fig. 4.6.). The zones with the highest error occurred at the boundaries of the study area. This was a result of a lack of data points at the boundaries. However, this was expected in any kriging analysis.

The lowest was by contrast, at location were monitoring points are clustered. For example, in the north-east area of the map, surrounding Beaufort West. In addition, between clusters of monitoring points the error increases. New monitoring network design should also focus on these zones of high error.

4.4. Design of the sampling grid

Following the geostatistical analysis in the previous section, model 1 was chosen as the most appropriate to use in monitoring network design. From model 1 the spatial autocorrelation range (16688.43m) was used to construct a hexagonal sampling grid (Fig. 4.7). According to Olea (1984) a hexagonal sampling grid provides the lowest overall standard prediction error of any sampling grid design. The grid was constructed so that from the centre of one hexagon to the next is a distance slightly greater than the spatial autocorrelation distance. Hence, by positioning new monitoring points within cells, according to a hydrogeological criteria, a monitoring network was developed that was optimized in terms of spatial density.

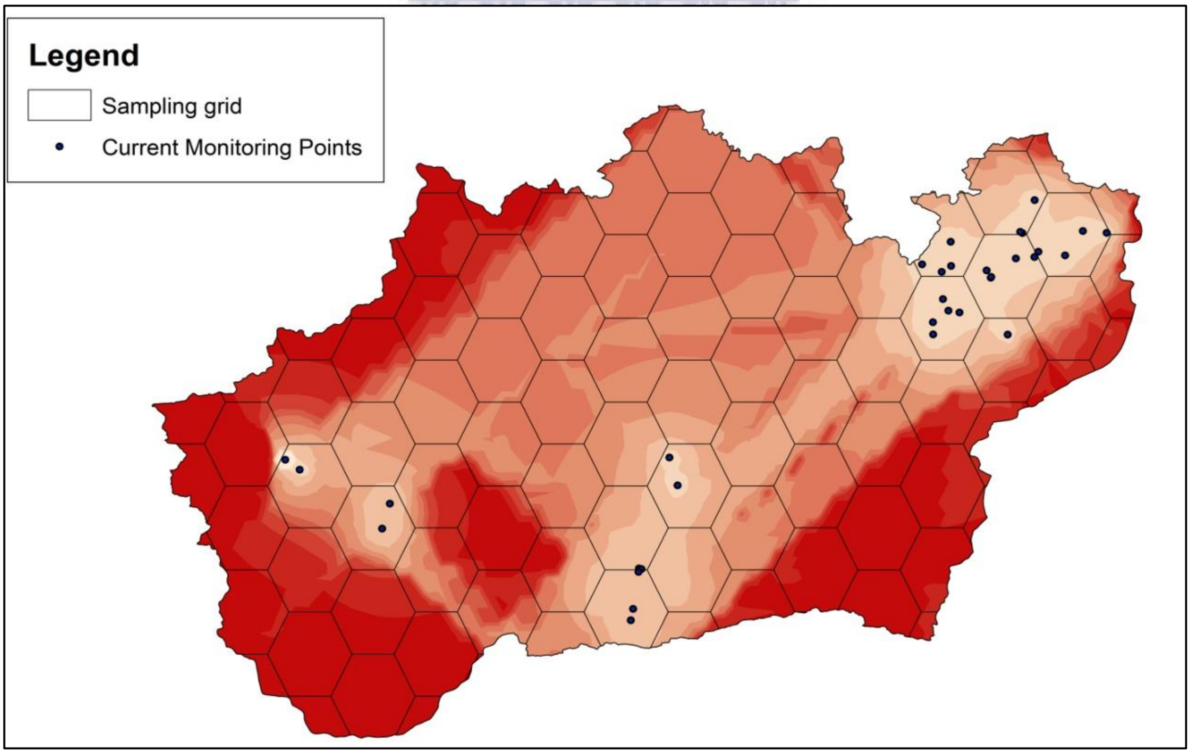


Figure 4.7. Map of the designed hexagon sampling grid, showing the current monitoring points as well as the prediction standard errors.

4.5. Hydrogeological placement of monitoring points

According to a set of hydrogeological criteria defined in chapter 3, five maps were generated that illustrated these criteria (Fig. 4.8. - 4.12.). The chosen criteria relate to key features within the hydrogeology of the study. In addition, in the context of shale gas development and monitoring network design, these features have been deemed important in various literature. This list of hydrogeological criteria used in the following analysis was not meant to be comprehensive, but only represents those features that data was available to generate maps for.

Figure 4.8 illustrates the rate of recharge for the study area. This map was intended to highlight areas of high recharge (recharge zones). The recharge zones represent the area of infiltration of new fresh waters, before mixing and chemical alteration takes place down-gradient. Monitoring of up-gradient wells was important in understanding baseline characteristic of the groundwater. Hence, new baseline monitoring points were positioned within high recharge zones, as indicated by the green points in figure 4.8.

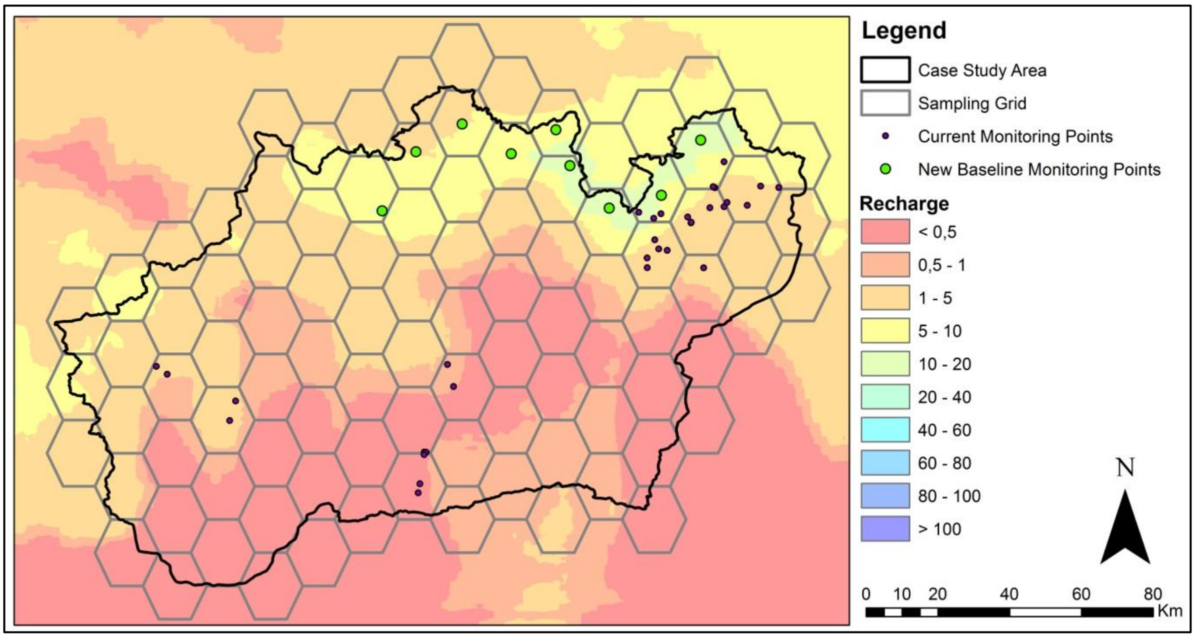


Figure 4.8. Map depicting the rate of recharge for the study area.

Dolerite intrusions, especially dolerite dykes have been discussed intensely in the literature with regard to hydraulic fracturing. These structures represent possible migration pathways for hazardous chemicals to move to the shallow aquifers. The presence of thermal springs indicated the circulation of deeper groundwater to the surface, and hence had been included as a contaminant pathway. In addition, the deep Soekor wells are known to intersect deep groundwater formation, and can provide a link between the deep and shallow groundwater formations. Monitoring of these features could provide early warning of potential contamination. Hence, new monitoring points were positioned within proximity to these pathways, as indicated in figure 4.9.

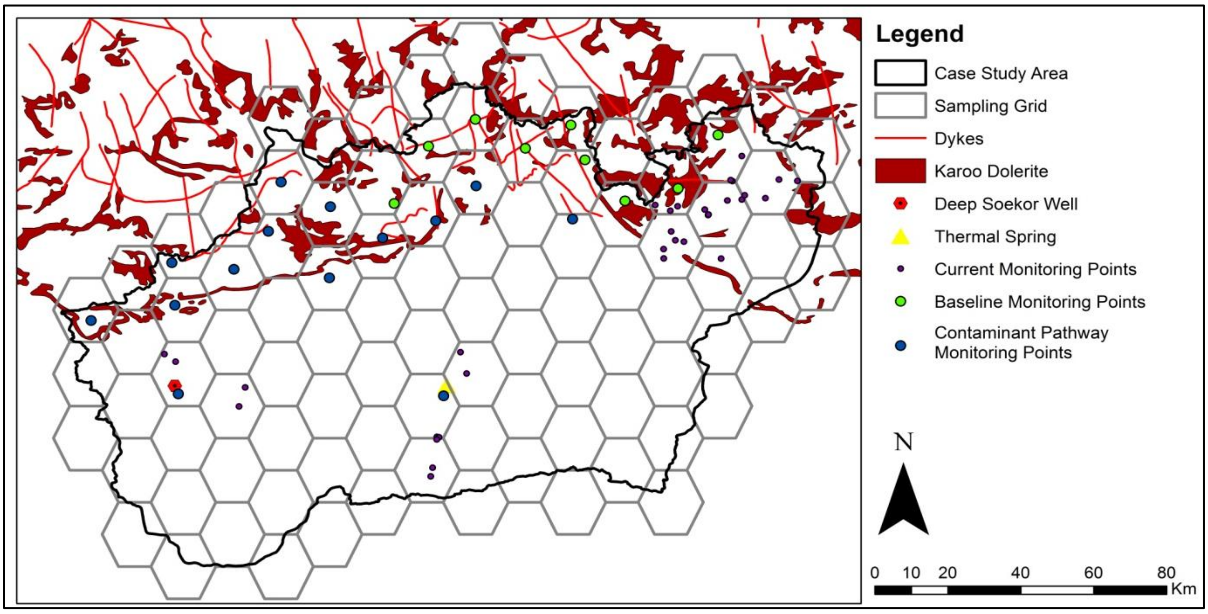


Figure 4.9. Map depicting the outcropping of dolerite intrusion within the study area, including dikes.

One of the objectives of the designed monitoring network was to protect valuable groundwater resources of the region. In this thesis water resources can be thought of as those zones that exhibit a high aquifer yield. High yielding zones represent preferred areas for groundwater abstraction and are so prioritised in this design of this monitoring network. Figure 4.10 indicates the aquifer yield for the study area. As can be seen there was generally only one zone of substantial yield, located in the north east of the study area. The rest of the study area generally exhibits low yield. Only one new monitoring point was positioned in this high yielding zone as the current monitoring network is well densified in this area.

In the same context, isolated high yielding abstraction points were included. These are vital water resource points used for domestic or agricultural purposes. Figure 4.10 displays the abstraction points in the study area (black triangles). There are only few high yielding abstractions points that are beyond the high yielding zone in the north-east. At these points new monitoring points were positioned at close proximity to these abstraction points. These new monitoring points were intended to provide early warning, by detecting changes in groundwater before being abstracted.

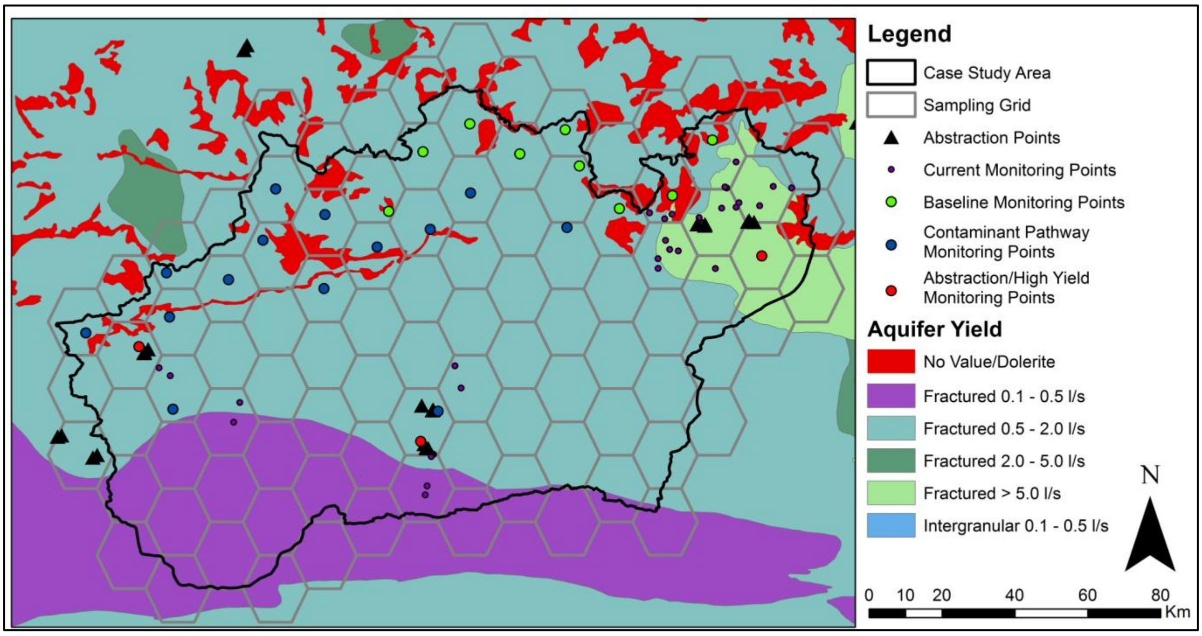


Figure 4.10. Map of the aquifer yield and aquifer type in the study area.

The next hydrogeological criteria defined were zones of favourable water quality. Monitoring of these zones will provide protection for current and future water demand (water resource protection). In figure 4.11 there was only one zone of high water quality located to the south east of the study area. In figure 4.11 the mapped variable was actually electric conductivity. The zone represents an area of low electric conductivity. Here new monitoring points were positioned within the zone of favourable groundwater quality.

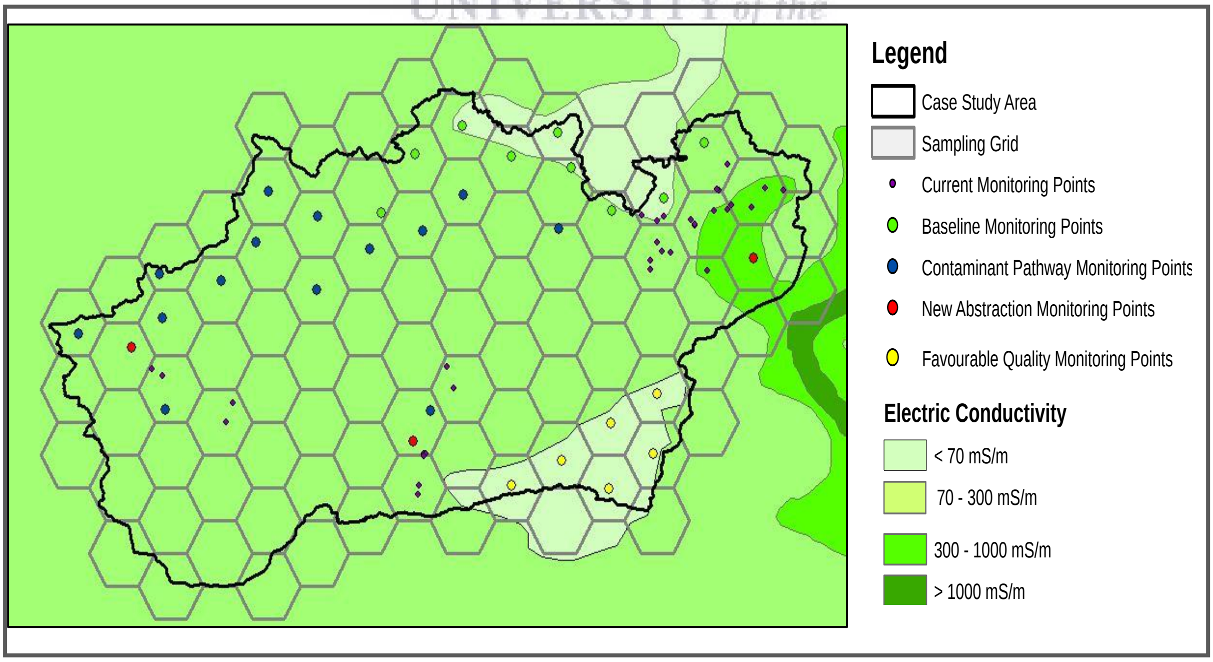


Figure 4.11. Map depicting areas of low electric conductivity (i.e. favourable water quality).

Lastly, the remainder of the hexagon cells that did not have monitoring points within them were included. In order to maintain the required density and provide coverage of all un-sampled locations general monitoring points were positioned at the centre of empty hexagon cells (Fig. 4.12). This further increased the density of the designed monitoring network to what was required to account for the spatial variability of the study area.

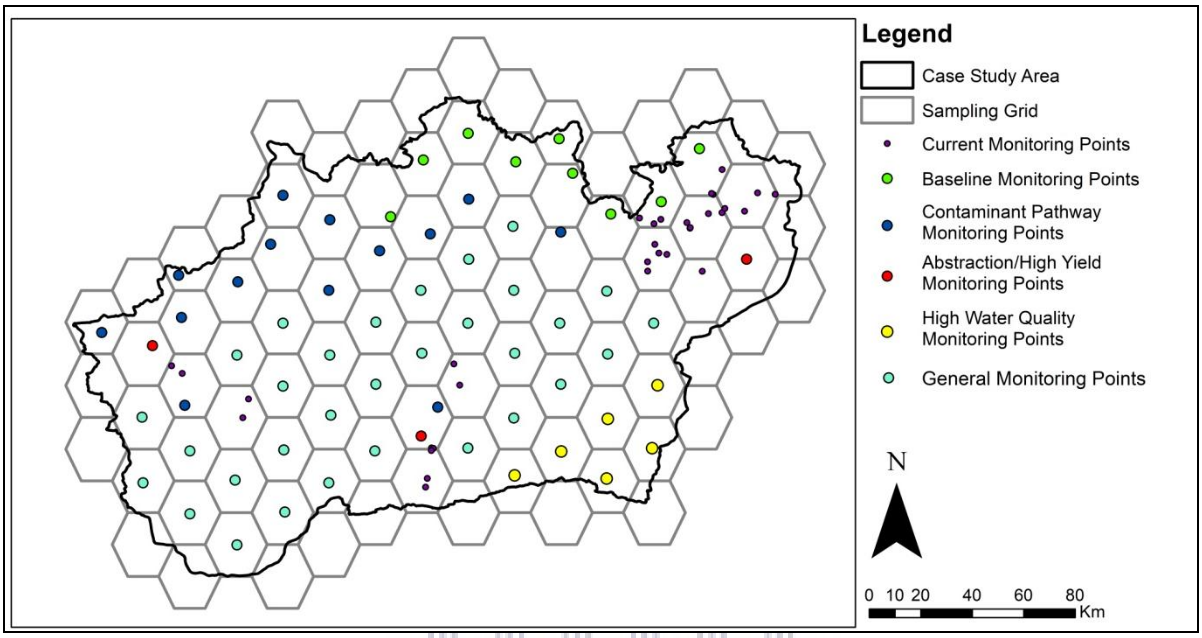


Figure 4.12. Map of all new monitoring points. Existing monitoring points are including in the newly designed network.

The procedure followed in this chapter was performed in order to develop a monitoring system that could be used to 1) determine the baseline groundwater characteristics of the study area, and 2) provide detection monitoring against contamination from shale gas development, as was possible. The later objective was related to trend monitoring during shale gas production. The study area exhibits high variability, and the adapted procedure followed in this chapter was designed to determine the required number of monitoring points to account for this variability. Using kriging, the required number of monitoring points was determined, as well as the spatial density of the network. In such a manner the new network was optimized. Thereafter using a hydrogeological approach, new monitoring points were positioned at key location in the context of shale gas development. The current monitoring network was preserved as the data from these monitoring points are important in trend analysis. Under this guidance, 61 new monitoring points were created. Including the current monitoring points, a total of 95 points exist within the newly designed network.

4.6. Re-evaluation of designed network

Figure 4.13 illustrates the prediction standard error analysis for the newly designed network, using predicted values at new points. Within the centre of the study area the error was relatively low. The prediction standard error was even lower were monitoring points are more dense such as in the north east section of the map. Only on the boundaries of the study area was the prediction standard error relatively high. The error also increases concentrically along the edges. Compared to the previous prediction standard error map for the current network only (Fig. 4.7), this map showed a lower average standard error across the study area. As well as showing smoother contours.

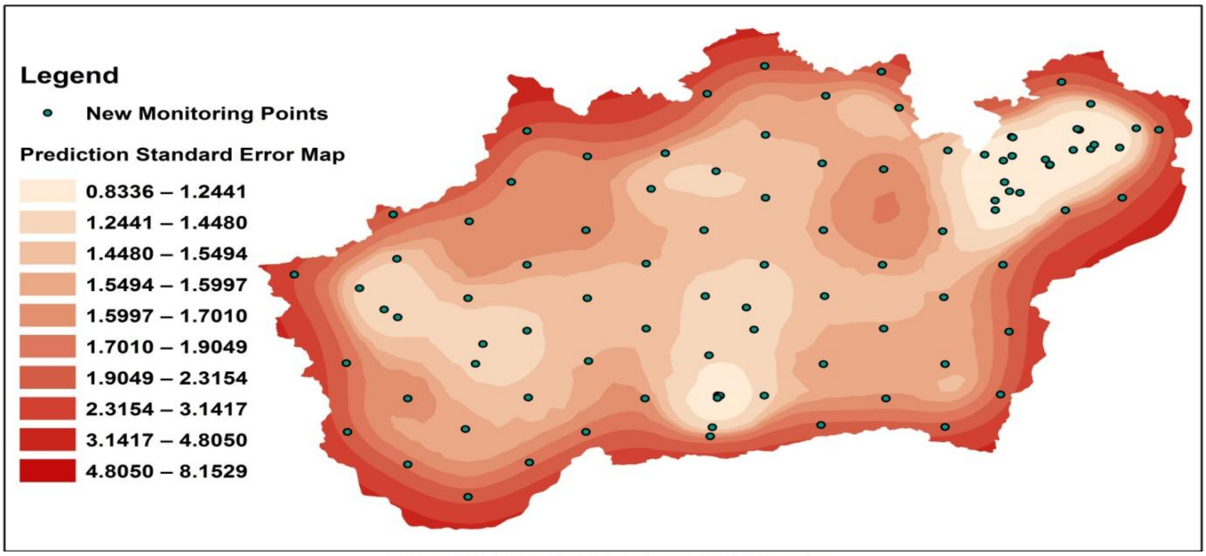


Figure 4.13. Prediction standard error map for the newly designed monitoring network.

4.7. Discussion on results of designing groundwater monitoring network

In an attempt to develop the most appropriate groundwater monitor system in the context of shale gas development, a hybrid geostatistical hydrogeological approach was followed in this chapter. The particular dataset chosen to perform the analysis was selected as it represented the most recent depth to water level data available. In addition, the data points representing these water levels coincide with the current monitoring system in the study area. In essence, this analysis was intended to be an inspection of the spatial distribution of the current monitoring network.

This research began with a statistical analysis of the chosen data set. The most concerning result from this section was high variability of the dataset. Spatially, this high variability manifested itself as highly erratic observations between neighbouring data points. This would imply that a greater number (or more dense network) of monitoring points was required to account for high variability. Hydrogeologically, one would expect gradual changes in depth to water level, following the topography. However, this erratic nature in depth to water levels could highlight that the aquifer was isolated at the local scale and instead depth to water levels are a manifestation of local water bearing fractures with intercept each borehole at different depths (Botha *et al.*, 1998; Woodford and Chevallier, 2002). In addition, this could imply a very heterogeneous aquifer at the local scale, at least (Woodford and Chevallier, 2002).

Following normality tests, which was not critical in kriging application, trend analysis was performed. As previously stated, trending data was not allowed. Trend analysis revealed an increasing trend towards the centre of the study area. Transformed depth to water level values similar also displayed a trend in the data. The fact that both normal and transformed depth to water levels clearly indicated the presence of a trend in the study area, should offer validation of the natural trend with regards to groundwater levels. However, because of a high variability of 47.45 m^2 this trend was obscured during visual inspection.

In summary the dataset was highly variable, non-parametrically distributed, and showed evidence of a trend. These were not ideal attributes of a datasets to be used in kriging applications. Nonetheless kriging was performed using detrended data, but interpretation was made with these factors in mind.

Firstly, all the experimental variograms models were fitted with a stable theoretical variogram model. This was similar in structure to spherical theoretical models. The fact that all the models fitted this structure, where there was a finite sill, suggested that pairs of data points beyond a certain distance show no correlation. This further suggested hydraulic discontinuity at the local scale, related perhaps to the heterogeneity of the fracture flow paths in the aquifer. However, this needs to be investigated further.

When the nugget effect was inspected, Model 1 produced no nugget effect, while models 2-4 had large nuggets effects. No models produced a pure nugget effect. Intuitively, one would expect the most accurate model to express a nugget effect. This would agree more so with the fact that the spatial variability was high. The only models that produced significant nugget effects were models 3 and 4. On the same principle, it was expected that the most appropriate model would have a small autocorrelation range. This was because the spatial variability was high. However, from the cross-validation analysis the most appropriate model appeared to be model 1. This was unexpected, because of the high spatial autocorrelation range of 16688.43 m.

It was also apparent from the cross-validation analysis that all the models produced high error. This could be a result of the intrinsic attributes of the dataset (non-normal, trending and highly variable), affecting the accuracy. Indeed the high variability could affect the stationarity of the data (to be discussed later). This would significantly affect the confidence in the results and consequently produce high error.

Using the results of model 1 a standard prediction error map was created. The map was intended to show the in-efficiencies in the current monitoring network. As previously stated the generally high prediction error away from the clusters indicated an unevenly distributed system. In addition, clustered monitoring points can possibly be reduced to 1 monitoring point without increasing the prediction error, due to being redundant. However, this research advises against this at the present moment due to the long time-series available from these points. Densifying the monitoring network using a predefined sampling grid was the obvious solution.

The spatial autocorrelation range was used to design the hexagonal grid in figure 4.7. Each grid cell was filled with a monitoring point according to a set of hydrogeological criteria. This approach resulted in the establishment of 61 new monitoring points. In realistic terms this was a large number of new monitoring points that may not be feasible/manageable. However, according to the geostatistical model this was the required number of monitoring points in-order to optimize the system. Inspection of the current monitoring network underneath the new sampling grid shows that current network was highly clustered. This would imply that many of these are redundant. Hence, in order to reduce the number of monitoring points in the network, clustering of these points could be avoided.

The newly designed monitoring network was re-evaluated using kriging, in-order to determine if it reduced the average standard prediction error. This was a fairly common procedure in the literature (Olea, 1984; Theodossiou and Latinopoulos, 2006; Yang *et al.*, 2008; Bhat *et al.*, 2015). The fact that across the study area prediction standard errors had decreased, suggests that the newly designed monitoring network was more efficient at interpolating parameters. This would further suggest that the newly designed monitoring network provided better coverage of the study area. Hence, the sampling grid design and location of new monitoring points appear appropriate and further optimized the network.

In the context of this study, that was under the premise of shale gas development and within semi-arid environments, the literature was almost absent of examples. The fact that the designed monitoring network was intended to meet multiple objectives, made this study a particular novel approach. The hybrid approach used in this study was chosen as this was the best manner to design one monitoring network that can meet multiple objectives. Geostatistical analysis provides us with the number of monitoring points and the network density. While the strategic positioning of the monitoring points was based on hydrogeological criteria.

This was by no means a unique methodology, but instead was adapted and modified from various literature, such as Olea (1984), Kim *et al.* (1995), Caeiro *et al.* (2003), Zhou *et al.* (2013), and Bhat *et al.* (2015). Olea (1984) used kriging techniques to determine the sampling grid shape that produces the lowest overall prediction standard error. Hexagonal grids produced the lowest prediction standard error. However, the other grid shapes might be appropriate according to the geometry of the aquifer (or environment). A case study from Olea (1984) showed similarities to this study, in that a randomly scattered monitoring network was optimized using kriging techniques. Similar to this study Olea (1984) determined that a stratified hexagonal grid was the best way to optimize the network. However, in this case the number of monitoring points was reduced without a loss in information.

Bhat *et al.* (2015) provided an even better comparison. Here, a regional monitoring network was expanded based on a kriging approach. The results achieved in this study appear to agree with those from Bhat *et al.* (2015). Both showed that the random pattern of the existing monitoring network produced high error, and a regular hexagonal pattern was preferred when expanding networks where there was no monitoring points.

Caeiro *et al.* (2003) was unique in this regard in that their work was not based on optimizing an existing network, but establishing a new network were there was none. Although the methodological approach was similar to this study, Caeiro *et al.* (2003) was based on developing a sampling/monitoring network for estuarine sediment sampling. The results obtained by Caeiro *et al.* (2003) thus differed from this study. Random sampling within a regular grid was chosen instead, based solely on the movement of the sampling vessel. As opposed to a systemic approach used in this study.

When discussing the hydrogeological component of this study Zhou *et al.* (2013) and Kim *et al.* (1995) provided some comparison. Kim *et al.* (1995) designed a national groundwater monitoring network based on a set of hydrogeological criteria. They also incorporated a grid sampling system to strategically position new monitoring points. Their results, which indicated key areas to monitor, further corroborate the methodology and results of this research.

Zhou *et al.* (2013) provided a comparative analysis of a hybrid design approach. They designed a regional scale network using a hydrogeological approach to position new monitoring points. Using regime zone mapping they highlighted the most critical areas for new monitoring points. Thereafter, using kriging to show a reduction in prediction standard errors to show the efficiency of the new monitoring network. Procedure in this research was reversed compared to Zhou *et al.* (2013); hence, the results were manifested differently. Network density and location was determined by the relative hydrogeological regime zones, and not by geostatistical analysis. Hence a more random pattern of monitoring points (visually) was obtained by Zhou *et al.* (2013)

The methodological approaches detailed in the literature, and subsequently adopted into this research, demonstrated its applicability. However, theoretically in any geostatistical analysis, the significance of the results depended on meeting the assumption that the dataset was statistically stationary. That was, the mean and variance remains the same across 3-Dimensional space. Currently, there was limited technical guidance to determine stationarity within the dataset; rather reliance was placed on intuitive interpretation. It was clear that the spatial variability was high within the dataset and that standard errors produced by all the kriging models were similarly high. This can allude to the fact the dataset may not have been stationary after all. This indeed reduced the accuracy and thus significance of the results.

The preceding discussion indicated that the current network employed in the study area was not adequate to meet the objectives of a regional groundwater monitoring network. Any interpretation or estimation made using the data from the current network poorly represented the entirety of the study area. Changes within the groundwater quality and quantity would not be detected early enough to protect the groundwater environment. This research showed that if a new monitoring network was not implemented before shale gas development begins the possibility of wide spread contamination within the groundwater may occur. Furthermore, establishment of a statistically sound baseline cannot be realised without the implementation of the newly designed network.

4.8. Summary of results and discussion

In-order to design a regional groundwater monitoring network based on multiple objectives, a geostatistical-hydrogeological approach was applied. This allowed for optimization in terms of density of monitoring network, and takes into account key hydrogeological features of interest when position monitoring points. The analysis revealed the current network, which contained 34 monitoring points, to be irregularly distributed and clustered throughout the case study area. Using kriging techniques a new network density was calculated, where monitoring points were separated by approximately 16.7 km. A systematic sampling approach was applied to a hexagonal sampling grid, which has the potential to reduce the kriging prediction standard error. Using key hydrogeological features such as contaminant transport pathways and water resources zones, 1 new monitoring per grid cell (in most cases) was placed within proximity to these features. In this manner, the groundwater monitoring network was expanded from 34 to 91 monitoring points. The new network showed a decrease in the kriging prediction standard error compared to the existing network, which suggests a gain in information.

Chapter 5 : Determining baseline in South Western Karoo

5.1. Introduction

One of the objectives of this monitoring network was detection or contaminant monitoring, in order to protect vital groundwater resources from contamination. Detection monitoring in general terms operates by monitoring certain parameters of concern against a predefined threshold of those parameters. This introduces the concept of baseline or background, which is the immediate natural state of the aquifer. The necessity for a baseline in context of shale gas development has been well discussed in the literature review. In detection monitoring it was important to have statistically credible baseline to act as a reference point for comparison against future observation. This is the concept of detection monitoring. The following chapter presents the results of analysis for the development of baseline. In essence the results in this chapter are intended to be an analysis of the available dataset to determine its applicability in statistical baseline development.

5.2. Exploratory Data Analysis

There were a total of 22 quality assured water quality data points that exist within the study area. However, only 5 of these points had an independent sample size that was greater than the preferred lower limit. Chemical analytes sampled at these points included a suite of major ions and field parameters (Table 5.2.). A data summary of these 5 points was presented in table 5.1. Supplementary information, such as histograms and time series graphs can be found in annex 1.

Table 5.1. Data summary for 5 ZQM groundwater quality monitoring points in the study area.

	ZQMWVL1	ZQMWVL2	ZQMLUE1	ZQMRSK1	ZQMBWW1
Start date	05/26/1994	05/26/1994	05/26/1994	10/18/1994	10/18/1994
End Date	09/02/2015	04/15/2015	04/15/2015	05/23/2015	06/06/2013
No. Of Samples	38	29	35	42	35
Frequency	Biannual	Biannual	Biannual	Biannual	Biannual
Associated WL	None	None	None	None	None

In statistical terms a sample size between 20 & 30 independent samples was considered to be the minimum lower limit. However the Department of Environmental Quality (2014) does suggest that a sample size of 12 independent samples spanning 3 years is sufficient to perform most statistical analysis. That will require quarterly samples. As can be seen in table

5.1 all of the monitoring points had a large enough sample size and the dataset spanned a much greater bandwidth than was required. Only the frequency of sampling was biannual instead of the recommended quarterly samples. However, this did not significantly affect the results of the following statistical analysis.

Table 5.2. Statistical summary for various chemical analytes measured during sampling of ZQM boreholes.

	ZQMWVL1		**ZQMWVL2**		**ZQMLUE1**		**ZQMRSK1**		**ZQMBWW1**	
Analytes	**Mean**	**¶Std.D.**	**Mean**	**¶Std.D.**	**Mean**	**¶Std.D.**	**Mean**	**¶Std.D.**	**Mean**	**¶Std.D.**
Ca	89.66	11.57	84.90	11.42	131.59	34.20	144.74	29.94	67.51	24.19
Cl	67.89	18.49	60.69	11.00	286.71	126.90	211.05	37.77	16.44	4.53
DMS	685.82	63.46	660.31	53.09	1562.81	391.31	1143.25	96.44	481.61	97.26
F	0.73	0.14	0.71	0.08	2.16	0.58	0.76	0.16	0.40	0.12
K	2.60	0.45	2.84	0.60	7.61	3.29	3.35	0.99	0.96	0.61
Mg	16.87	1.66	17.19	1.59	18.20	10.46	38.36	7.23	20.38	7.39
NH4	*bdl	bdl	bdl	bdl	bdl	bdl	bdl	bdl	bdl	bdl
NO3	2.79	0.68	3.87	0.73	0.39	0.93	2.09	1.70	0.10	0.07
Na	77.05	7.86	71.46	4.32	334.38	92.59	148.59	31.13	34.00	10.22
PO4	bdl	bdl	bdl	bdl	bdl	bdl	bdl	bdl	bdl	bdl
SO4	63.09	13.27	58.92	11.25	450.92	187.13	285.66	101.29	78.02	52.89
Si	11.42	0.84	10.74	0.84	9.29	1.76	10.76	0.97	14.34	1.66
TAL	292.04	27.52	284.99	25.38	295.35	130.16	253.30	39.40	217.86	34.67
Hardness	293.54	32.12	†N/A	N/A	N/A	N/A	481.06	125.79	N/A	N/A
EC	95.70	21.79	89.41	25.01	272.91	70.23	156.55	19.67	58.42	15.32
Temp	22.31	2.70	22.65	2.13	21.86	3.20	20.91	0.94	19.30	1.36
pH	7.12	0.66	7.37	0.47	7.23	0.69	7.15	0.17	7.17	0.23

¶Std.D: Standard deviation
**bdl: below detection limit,*
†N/A: no data.

Table 5.2 presents both the mean and standard deviation of the dataset. This was intended to indicate the spread or distribution of the dataset. For many of the analytes the standard deviation was relatively high. For example, EC for ZQMWVL1 and Na for ZQMBWW1, both show high standard deviations compared to the mean. In fact, none of the analytes display a low standard deviation compared to the mean. This table was also intended to show the spread of the dataset; through the variance (variance is the square of the standard deviation). A high standard deviation illustrated a large spread and consequently a high variance. In annex 1, the histograms corroborated this result, illustrated by large spreads for many of the analytes (see for example figure 1.5. and 1.8. in annex 1).

5.3. Analysis of spatial and temporal data independence

Analyzing the independence of data using technical statistical protocols was more rigorous and tedious than required. Instead the independence of the various dataset was proven using hydrogeological reasoning. For spatial independence, the five monitoring points were separated by great distance, that independence could be assumed. Considering groundwater movement was slow and that the monitoring points were separated by geological boundaries, it was not expected for samples to be collected from the same statistical population in the aquifer. In essence, it was unlikely that the sampling points were connected directly.

In terms of temporal independence, the quarterly rule of thumb was used to assume statistical independence. Most of the sampling events in the time series were separated by more than four months in most cases. Under the current groundwater flow conditions this was sufficient time for groundwater to move past the sampling point. Hence, it was assumed that new waters were sampled on each occasion. Therefore, temporally the dataset was also independent.

5.4. Test to determine normality of dataset

The following section dealt with the results of the normality test. This component of the analysis was of importance, as this determined the protocol for the proceeding steps. That was whether parametric or non-parametric test should be used to analyse the dataset.

The Shapiro-Wilks test was used to determine the normality of datasets for the various analytes for each data point. The results of the Shapiro-Wilks test are shown in table 5.3. This table displays the significance score for test, with a confidence level set at 95%. Hence, any significance score below 0.05 did not indicate a normal distribution at the defined confidence level. Analytes mark with an asterisks highlight all the datasets that had significance scores below 0.05. The results showed that majority of the datasets were not normally distributed. More correctly, 67.5 % of the analytes had non-normal datasets. However, before proceeding with non-parametric statistical procedures, the transformed datasets were also analysed for normality.

Table 5.3. Results of the Shapiro-Wilks test for normality performed on original dataset.

	ZQMWVL1	**ZQMWVL2**	**ZQMLUE1**	**ZQMRSK1**	**ZQMBWW1**
Analytes	**¶Sig. Score**	**¶Sig. Score**	**¶Sig. Score**	**¶Sig. Score**	**¶Sig. Score**
Ca	*0.000	*0.000	0.994	*0.001	*0.001
Cl	*0.001	0.143	*0.000	0.291	*0.035
DMS	0.070	*0.008	*0.036	0.303	0.088
F	*0.000	*0.001	*0.003	*0.000	*0.002
K	*0.006	*0.021	0.321	*0.000	*0.000
Mg	0.137	0.295	*0.002	*0.002	*0.002
NH4					
NO3	0.855	0.177	*0.000	*0.004	*0.001
Na	*0.001	0.402	*0.008	*0.000	*0.000
PO4					
SO4	*0.000	*0.001	*0.001	*0.000	*0.000
Si	*0.002	*0.012	*0.000	0.557	0.335
TAL	*0.000	*0.000	*0.002	*0.009	0.907
Hardness	*0.004			0.474	
EC	*0.013	0.152	*0.001	0.105	0.618
Temp	*0.001	*0.001	*0.012	*0.049	*0.006
pH	0.160	*0.014	*0.029	*0.014	*0.019

**denotes those analytes that are above significance level (i.e. non-parametric datasets)*

¶Significance score

Other parametric distributions such as log normal or gamma distribution were also accepted in terms of normal distributions (U.S. Environmental Protection Agency, 2009). Hence, all datasets were transformed to their natural logarithm, and tested for normality. Figure 5.4 represents the results of this analysis, using a 95 % confidence level. The analytes marked with an asterisks indicate those that were not parametrically distributed. Only 36.1 % of analytes displayed parametric distributions. Considering that so few analytes are normally distributed in both real values or transformed values, it was necessary to continue the analysis using only non-parametric statistical techniques.

Table 5.4. Results of the Shapiro-Wilks test for normality performed on transformed values.

	ZQMWVL1	ZQMWVL2	ZQMLUE1	ZQMRSK1	ZQMBWW1
Analytes	**¶Sig. Score**	**¶Sig. Score**	**¶Sig. Score**	**¶Sig. Score**	**¶Sig. Score**
Ln-Ca	*0.000	*0.000	0.084	*0.000	0.115
Ln-Cl	0.067	0.421	*0.000	0.132	0.712
Ln-DMS	*0.017	*0.001	*0.000	0.061	0.505
Ln-F	*0.000	*0.005	*0.000	*0.004	*0.017
Ln-K	*0.022	0.332	*0.000	*0.000	*0.000
Ln-Mg	0.443	0.268	0.701	*0.000	0.155
Ln-NH4					
Ln-NO3	0.204	*0.011	*0.000	*0.000	0.072
Ln-Na	*0.003	0.542	*0.000	*0.000	*0.005
Ln-PO4					
Ln-SO4	*0.000	*0.023	*0.000	*0.000	*0.007
Ln-Si	*0.011	0.070	*0.007	0.412	0.500
Ln-TAL	*0.000	*0.000	0.050	*0.006	0.291
Ln-Hardness	*0.000			0.405	
Ln-EC	0.322	*0.001	*0.013	0.457	0.992
Ln-Temp	*0.000	*0.000	*0.003	0.081	*0.001
Ln-pH	0.089	*0.016	*0.003	*0.017	*0.037

**denotes those analytes that are below detection limit (i.e. non-parametric datasets)*

¶Significance score

5.5. Determination of seasonality within the dataset

Determining a steady state in the time series (or temporal stationary) was a key consideration. Hydrogeologically, this phenomenon is known as seasonality. Seasonality is where there is a difference in measurements at different times of the year (e.g. wet or dry season). Occurrence of seasonality introduces more requirements when determining the groundwater baseline statistically. Figure 5.5 presents the results of a Kruskal-Wallis test for seasonality.

The average winter rank (April to August) was compared to the average summer rank (September to March) using the Kruskal-Wallis test (Table 5.5.). The Kruskal-Wallis test was performed at a 95% confidence level. In table 5.5. any significance score below 0.05 represented an analyte that showed seasonality in its dataset (see analytes marked with an asterisks). The table 5.5. illustrated that using the current dataset, at the determined confidence level, only 6 analytes showed seasonality. Many of the analytes display minimal difference between groups, according to the average rank. Only in instances where there was a vast difference between average ranks, was the significance score low enough to indicate a statistical difference. For example, Ca in ZQMWVL2 or Si in ZQMWVL1.

Table 5.5. Results of Kruskal-Wallis test using seasonality variables

	ZQMWVL1			ZQMWVL2			ZQMLUE1			ZQMRSK1			ZQMBWW1		
Analytes	**¶Sum. Mean**	**¶Win. Mean**	**¶Sig. Score**	**¶Sum. Mean**	**¶Win. Mean**	**¶Sig. Score**	**¶Sum. Mean**	**¶Win. Mean**	**¶Sig. Score**	**¶Sum. Mean**	**¶Win. Mean**	**¶Sig. Score**	**¶Sum. Mean**	**¶Win. Mean**	**¶Sig. Score**
Ca	17.8	19.375	0.656	11.867	18.357	*0.04	17.313	17.667	0.918	19.975	21.976	0.593	16.765	19.167	0.488
Cl	20.19	17.438	0.443	14.4	15.643	0.694	18.875	16.278	0.448	22.25	20.818	0.706	16.294	18.706	0.48
DMS	16	17.143	0.443	10.167	15.615	0.064	17.867	15.294	0.439	19.706	18.4	0.715	15.125	18.765	0.28
F	16.45	19	0.462	13.643	14.385	0.808	14.533	17.375	0.385	20.222	19.81	0.91	17.375	17.611	0.945
K	17.375	18.833	0.677	14.714	13.231	0.627	18.219	15.853	0.482	20.895	20.143	0.839	17.618	18.361	0.83
Mg	18.548	19.594	0.771	14.9	15.107	0.948	17.719	17.306	0.904	22.25	20.818	0.706	14.294	20.706	0.06
NH4															
NO3	17.05	20.313	0.356	13.071	15.929	0.358	19.25	14.882	0.186	21	21	1	19.563	15.667	0.254
Na	17.725	18.367	0.855	10.929	17.308	*0.037	19	15.118	0.249	21.158	19.905	0.735	16.294	19.611	0.338
PO4															
SO4	18.929	19.094	0.963	13.733	16.357	0.407	16.938	18	0.756	22.65	20.455	0.562	17.882	17.118	0.823
Si	14.238	25.25	*0.002	15	15	1	14.969	18.912	0.242	21.4	21.591	0.96	13.765	22	*0.017
TAL	17.85	19.313	0.679	11.857	17.143	0.089	17.375	16.647	0.829	20.65	21.333	0.855	13.647	21.353	*0.024
Hardness	17.8	19.375	0.656							8.429	7.625	0.728			
EC	12.656	16.958	0.171	9.55	11.45	0.472	13	12.077	0.75	22.825	20.295	0.505	16.714	17.211	0.884
Temp	10.429	17.083	*0.027	9.85	11.15	0.615	14.682	10.654	0.162	19.95	22.909	0.431	14.036	19.184	0.125
pH	14.813	12.818	0.521	9.8	11.2	0.596	10.227	12.773	0.358	16.563	20.857	0.23	16.167	15.056	0.733

** Denotes those analytes that are below significance level (i.e. display seasonality within the dataset)*.

¶*Sum.Mean: average rank for summer season; Win.Mean: average rank for winter season; Sig.Score: significance test score).*

5.6. Determination of secular trends within datasets

Another important consideration was testing for steady state which was determining whether secular trends existed in the dataset. The presence of secular trends meant baselines could not be defined according to statistics. Unlike seasonal trends which can be removed. Alternative methods would have to be used if secular trends existed.

The table 5.6 represents the results of a Mann-Kendall test for trends in the datasets, performed at the 95% confidence level. The Mann-Kendall test score represents the sum of scores from comparing earlier observations to later observations in the time series. Positive test scores implied an increasing trend. A negative test score implied a decreasing trend. A test score of 0 implied not trend in the time-series. A significance score less than 0.05 confirms the presence of a statistically significant trend (positive or negative). The results shown in table 5.6 indicate that a large percentage of analytes display long term data trends in the time series. Approximately 37 % of analytes display a secular trend (analytes marked with an asterisks). Other analytes such as Ca in ZQMWVL1 or SO4 in ZQMWVL1, display a minor decreasing trend over time, but are not significant at the current level of confidence.

Table 5.6. Results of the Mann-Kendall test for secular trend.

	ZQMWVL1		**ZQMWVL2**		**ZQMLUE1**		**ZQMRSK1**		**ZQMBWW1**	
Analytes	**¶Test. Score**	**¶Sig. Score**	**¶Test. Score**	**¶Sig. Score**	**¶Test. Score**	**¶Sig. Score**	**¶Test. Score**	**¶Sig. Score**	**¶Test. Score**	**¶Sig. Score**
Ca	-74	0.170	-26	†0.320	11	0.441	-209	*0.010	-45	0.266
Cl	15	0.430	-2	0.493	127	*0.031	67	0.237	117	*0.043
DMS	-110	*0.046	-22	0.312	86	0.084	10	0.453	10	0.445
F	22	0.383	11	0.417	121	*0.021	161	*0.026	51	0.229
K	53	0.239	104	*0.016	-17	0.402	178	*0.020	205	*0.002
Mg	78	0.166	7	0.455	-38	0.292	-83	0.187	39	0.287
NH4										
NO3	-44	0.287	-280	*0.000	12	0.431	-139	0.061	-253	*0.000
Na	-183	*0.007	57	†0.122	152	*0.010	222	*0.005	22	0.383
PO4										
SO4	-111	0.083	12	0.418	219	*0.001	29	0.381	-115	*0.046
Si	63	†0.218	-70	0.098	-113	*0.041	-252	*0.003	47	†0.257
TAL	-136	*0.039	-12	0.414	-160	*0.007	-126	0.080	-7	†0.465
Hardness	-52	0.252					-45	*0.015		
EC	-83	0.053	-7	0.423	2	0.490	-118	0.102	65	0.161
Temp	-92	†0.022	-66	*0.015	-48	0.120	-308	*0.000	-134	*0.018
pH	-7	0.450	-61	*0.026	-70	*0.026	-25	0.376	-3	0.486

**denotes those analytes that a below significance level (i.e. trending data)*

†*Denotes those analytes that display seasonality*

¶ *Test.Score: Mann-Kendal test score; Sig.Score: Significance test score*

5.7. Determining the need to pool data

The penultimate decision in this analysis was to decide if data pooling should occur. This would increase sample size, which would in-turn increase statistical confidence in the final results. The consideration to pool data from each monitoring point together, was taken from a statistical as well as hydrogeological viewpoint.

Considering the hydrogeology, it was not advisable to pool data between all the monitoring points. Monitoring points in the north east of the study area were underlain by a different geology (see chapter 3). Aquifer properties in this zone were different to the other monitoring points. Under the current hydrogeological conditions it was only advisable to consider pooling points that are not separated by great distance and were in the same geological formation. ZQMWVL1 and ZQMVWL2 were thus realistic dataset to pool. Hence, statistical analysis was performed to determine similarities in variance and mean between the datasets.

Table 5.7 shows the comparison between variances for each analyte between ZQMWVL2 and ZQMWVL1. The ANOVAs test was performed at the 95% confidence level. Analytes with a significance score below 0.05 did not have similar variances. The results showed that all the significance scores for the analytes were above 0.05. This indicated that the variances were the same between ZQMWVL1 and ZQMWVL2.

Table 5.7. Results for ANOVAs test for homogeneity of variances between ZQMWVL1 and ZQMWVL2.

Analytes	[¶]Sig.Score	Analytes	[¶]Sig.Score
Ca	0.310	**NO3**	0.961
Cl	0.415	**Na**	0.608
DMS	0.833	**SO4**	0.079
F	0.617	**Si**	0.734
K	0.369	**TAL**	0.282
Mg	0.465		

[¶] *Sig.Score: Significance score*

Table 5.8 shows the results for a Kruskal-Wallis test, to determine similar means between ZQMWVL1 and ZQMWVL2. The test was performed at the 95% confidence level. The analytes marked with an asterisks indicated those that did not display similar means. Hence, these analytes could not be pooled. Considering that a few datasets did not have similar means, the data was not pooled, as it was not statistically advisable.

Table 5.8. Results of the Kruskal-Wallis test for similar means between ZQMWVL1 and ZQMWVL2.

Analytes	[¶]Sig.Score	Analytes	[¶]Sig.Score
Ca	*0.029	**NO3**	*0.000
Cl	0.136	**Na**	*0.003
DMS	0.108	**SO4**	0.097
F	0.155	**Si**	*0.000
K	0.053	**TAL**	0.126
Mg	0.369		

**denotes those analytes that are below significance level (i.e. do not have similar means*

[¶]*Sig.Score: Significance score*

5.8. Discussion on the determination of baseline in South Western Karoo

The following is a discussion of the results obtained during the analysis in chapter 5. It was previously stated that the point of the analysis in chapter 5 was to 1) determine the appropriateness of the current dataset for statistical baseline analysis, and if possible 2) determine with confidence upper decision thresholds for the various analytes. The later objective was important for detection monitoring.

During exploratory data analysis, it was determined that there were only 5 data points with proper time series of data. The number and distribution of these points limits the hydrogeological inferences that could be made from it. This implied that any baseline or background developed would not be representative of the entire study area. In addition, the time series which spanned more than 20 years only had a temporal resolution of 1 sample rough every six month. This was not ideal for seasonality analysis.

The analysis began with determining the descriptive statistics of the datasets. The importance of this section illustrated the general statistical characteristics of the various datasets. However, discussing every parameter was beyond the scope of the study, instead the discussion focused on understanding the statistical distribution. A full exploratory data analysis can be found in annex 1 (see for example table 1.1. in annex 1). The ranges of values for most datasets were large and were spread far away from the mean according to the standard deviation. This implied the presence of extreme values (outliers) in the dataset. Box and Wisker plots appeared to confirm this observation. At this stage it was not advised to remove the outliers as they could represent real values, albeit extreme. The presence of multiple outliers within a dataset did suggest this. A high standard deviation corresponds to a relatively high variance within the dataset, which suggested that values measured were significantly different from one observation period to the next.

These characteristics alluded to possibility of a non-normal dataset. Certainly the fact the majority of the datasets are either positively or negatively skewed (i.e. non symmetrical), strongly suggested non-normal datasets. Histograms in the annex 1 (see figure 1.13. in annex 1) corroborated this hypothesis, in addition Shapiro-Wilks test for normality showed that majority of the datasets were non-parametrically distributed. This decreased the confidence in any future upper detection limits set using this data. It was hence not ideal to have a dataset that was not normally distributed, to determine baselines in a statistical context.

Analysis of the seasonality revealed inconclusive seasonal changes in the data, which could inform the hydrogeological understanding of the region. The statistical test revealed that there was generally limited variation in concentration between seasons within the datasets. When the mean values for winter and summer months were compared, it could be seen that there was a no common trend that one group was higher than the other. In essence, seasonality needs to be confirmed with more data, and should also extend to correlate ground water-levels with chemistry data. Hydrogeologically, it appeared that seasonal changes in the physiological environment had little effect on the chemistry of the groundwater, at least with the current level of data.

It must be noted that the datasets were not truly ideal to determine seasonality. An ideal dataset would have had consisted of quarterly samples with each sampling event having taken place at the same month every year. The fact that so few datasets were non-seasonal could have been a consequence of the irregular seasonal sampling. It may be possible that there was seasonality presents but with the current data it was inconclusive.

Secular trends or long-term trending data was a critical consideration in this analysis. Having such a large portion of the datasets exhibiting a long-term trend, effectively limited the ability to determine a baseline via statistical means. It may be possible to remove secular trends from the datasets; however at this stage more investigation into this phenomenon should take place. Long-term trending data suggest that the aquifer was not in a steady state. Indeed from the time-series graph it could be seen that the measurements were highly erratic (variable) from one observation to the next. This could be a consequence of changing water levels, but this requires further investigation.

Of particular caution when dealing with trending data, was positively trending data. That was values that were increasing over time. The concept was that if a threshold (upper detection limit) was determined and set, then at some point in the future observations will have exceeded this threshold. Hence, setting thresholds for positively trending data was not advised. It was more preferred to wait for the aquifer to reach a steady state, if applicable. Future research was needed to understand these secular trends spatially, temporally and within a hydrogeological framework.

Data-pooling was not advised on the basis that an intrawell analysis was preferred, instead of estimating one baseline dataset for the entire study area. Hydrogeological this was more appropriate due to the spatial variation of groundwater chemistry across the study area. Thus in a monitoring sense only future observation within a specific well could be compared to the baseline of that well. The baseline of a specific well cannot be extended to other monitoring points.

Summarizing the results, firstly there were only 5 data points that had large enough sample size which presented an extended time-series. This spatial distribution of these 5 points across the study area was not ideal to represent the entire study area. The datasets of analytes sampled for were highly variable and express a large range of values. Extreme outliers skewed the dataset which in-turn affected the normality of these datasets. Thus majority of the datasets were non-parametrically distributed. Surprisingly seasonal changes in the datasets were not evident. However, long-term trends in the groundwater chemistry were evident.

Limited research had been carried out on pre-shale gas development baselines. In fact in the context of this study, this research could be considered unique. For instance Sloto (2013) was the only research aimed at developing a baseline, pre-shale gas development. However, there study was only focused on characterising and describing the groundwater quality, not on statistically determining the baseline and threshold concentrations. In addition, their study falls short in regards to understanding long term trends in groundwater chemistry. This was a key consideration in the present research. In essence, this research provides a more comprehensive analysis of how to establish a groundwater baseline compared to Sloto (2013).

Edmunds *et al.* (2003) and Shand *et al.* (2007) perhaps provided a better comparison. Their research focused on establishing the baseline conditions for groundwater chemistry across the United Kingdom. They used a statistical approach, such as descriptive statistics and cumulative frequency plots to determine the natural background concentrations as well as upper concentration thresholds. However, the results of this research could not truly be compared with Edmunds *et al.* (2003) and Shand *et al.* (2007), as the statistical approaches differed. What could be said was that the statistical component of baseline development was critical, especially when focused on future trend and contaminant monitoring.

Baseline in a statistical sense could not be considered one specific value (for example the mean or median of the dataset); instead it must rather be considered an entire dataset. That was, a range of values that defines the population of possible values that can be observed from the aquifer. The threshold (upper and lower) then defined the boundary of the most likely values that will be observed during monitoring. Observations that are repeatedly beyond the threshold could be considered a statistically significant change. This approach and understanding was best suited to meeting the objectives of this research. Establishing a reference point to incorporate in the monitoring system could not be achieved without statistical analysis.

In analysing the available dataset, it was not currently advisable to establish upper decision thresholds. The high variability in the time-series as well as the prominent secular trends within a large portion of the datasets emphasised the need to study these phenomenon further. It would also be fruitful to research and understand these phenomenon's in a hydrogeological sense. The number of groundwater monitoring points as well as the annual frequency of sampling will have to be increased to better understand the system. Hence, it was suggested that the prescribed monitoring network designed in this thesis be implemented and that at least 3 years of quarterly samples be collected before shale gas development proceeds.

5.9. Summary of results and discussion

To establish the natural baseline, the historical and current groundwater chemistry data was analyzed using statistical methods. Only 5 groundwater chemistry points exist in the case study area with a sample size large enough to perform the statistical review. The exploratory data analysis revealed that for most analytes the distribution was spread far beyond the mean, with high variability (skewed heavily by the presence of outliers). This would imply a large population of expected values during sampling, which suggests that the aquifer is not in a steady state. All datasets were considered to be spatially and temporally independent. A total of 67.5 % of the analytes had non-normal distribution, while the natural logarithm of the datasets was only marginally better at 63.9 %. Seasonality was not present in the dataset, but more frequent annual sampling is required to accurately determine seasonality with greater confidence. Long term trends on the other hand are clearly evident within the time series of 37% of the analytes. This indicates that the aquifer is not in a steady state. Finally data should not be pooled due to differences in mean and variability between the data points. Instead an intrawell analysis should proceed, which will allow site-specific interpretations. From the results presented in this thesis the current level of data is not sufficient to statistically determine the baseline conditions of the aquifer. The high variability and secular trends should be investigated from a hydrogeological point of view.

Chapter 6 : Parameters and Frequency of Monitoring

6.1. Introduction

In order to implement a groundwater monitoring system the parameters that need to be measured, as well as the frequency of monitoring need to be considered. This was an important consideration in order to realise the purpose of the monitoring system. The following chapter presented the results of objective 3, which details what parameters need to be monitored for and the frequency of monitoring. The results were based on analysis of relevant literature and a general understanding of hydrogeochemical properties.

6.2. Parameters of concern

Table 6.1. illustrates a comprehensive list of chemical parameters that should be monitored in the context of shale gas development. Chemical parameters that are marked with an asterisks in table 6.1. are those that were most important during detection/ compliance monitoring. From the literature review (section 2.3.), the general concerns in terms of sources of contamination include: stray gas migration, infiltration of hydraulic fracturing fluids, infiltration of drilling fluids and infiltration of deep formation fluids. These chemical parameters provided a direct and indirect tracer towards possible sources of contamination.

For baseline monitoring, macro chemical constituents, such as those in column 2 (Table 6.1.), were most important. These parameters provided information on the natural background chemistry within the aquifer. They were easy and inexpensive to analyze for, hence their inclusions. In addition, macro constituents can also be used as indicator parameters, such as Cl, DOC and TDS. Abnormally high values of these elements could indicate the infiltration of hydraulic fracturing fluid or even deep seated formation waters (Warner *et al.*, 2014). It should also be noted that physical parameters (column 1) could act as surrogate indicators for changes in the aquifers brought about by most sources of contamination from shale gas development (Son and Carlson, 2015). For this reason, the present study recommended the real time monitoring of field parameters.

Trace elements had also been included. These elements exist at relatively low concentrations within the study area. Hence, in terms of understanding baseline conditions they are not critical. Trace elements however, could act as indicators for contamination by hydraulic fracturing fluids, as they are common additives. In addition, trace elements such as Li and B (amongst others) tend to be more concentrated in fluids originating from deeply buried shale formations. Hence, they can also act as indicators for infiltration of deep formation fluids. Thus, they were included but should be sampled for at lower frequencies as they were generally more expensive to analyze and provided the same level of information as those in the macro constituents category.

Table 6.1. Target chemical parameters to include during monitoring (Modified from O'Brien et al., 2013).

Physical chemistry	Macro elements	Trace elements	Anthropogenic compounds	Dissolved gasses	Radiochemistry and isotopes
*pH	Ca	Zn	*VOCs	*Dissolved Methane	*Water $\delta^{18}O$
*EC	Mg	Al	*PAH	*Dissolved Ethane	Water δD
*TDS	K	Sb	*SVOCs	Radon	DIC $\delta^{13}C$
*DO	Na	As	*Glycols	*Methane $\delta^{13}C$	
*ORP	NH4	Cd	*Alcohols	*Methane δD	Gross alpha radioactivity
Temp	*Cl	Cr	*TPH	*Ethane $\delta^{13}C$	Gross beta radioactivity
	NO3	Co		*Ethane δD	^{4}He
	SO4	Cu			^{20}Ne
	PO4	Fe			^{36}Ar
	F	Pb			
	Alkalinity	Mn			
	*DOC	Hg			
	DIC	Ni			
		Se			
		V			
		Ba			
		*Li			
		Sr			
		Br			
		U			
		*B			
		Rb			
		Mo			

* *denotes elements and compounds that are critical in understanding sources of contamination*

The next category of parameters to monitor anthropogenic compounds. This category includes compounds such as volatile organic compounds, alcohol compounds, amongst others (analytes marked with an asterisks). These group of chemicals were anthropogenic (i.e. they do not occur naturally within the aquifer). They were also one of the main components in hydraulic fracturing fluid. Hence, their presence in sampled groundwater could only be assumed to be as a result of contamination by hydraulic fracturing fluid. These compounds were thus considered important indicator parameters.

Dissolved gasses make up the next category of chemical compounds. Considering the possible occurrence of stray gas migration, this was a critical component to monitor. For example, the ratio of dissolved methane to dissolved ethane in the groundwater was a clear indication of contamination by stray shale gas (Jackson *et al.*, 2013). Isotopic analysis of dissolved gasses is also critical to understand the difference between shallow biogenic methane and thermogenic methane (Talma and Esterhuyse, 2013).

The last category of compounds dealt with isotope hydrogeochemistry, and also included radiochemistry analysis. Isotope chemistry could act as indicator parameters. Isotopic signatures of compounds such as water and DIC could reveal changes in the groundwater composition. This could provide indication of infiltration of formation fluids (Warner *et al.*, 2014). Radiochemistry analysis (Beta and Gamma radioactivity) was included in the list to account for the large degree of uranium mineralisation that occurs in this region of the Karoo. Similarly the contamination by NORMs, which were more concentrated in deep formation fluids, could be indicated by radiochemistry analysis (Perry, 2011).

6.3. Frequency of monitoring

In terms of frequency of sampling, both the tasks of baseline monitoring and detection monitoring should be considered. Multiple tasks required from one monitoring system often competed with each other; hence two phases of monitoring should be undertaken. Pre-shale gas development sampling for baseline development, and later detection (trend) monitoring during shale gas production.

During pre-development sampling phase, in-order to establish statistically sound baselines, it was advisable to collect at least 3 years of quarterly samples (Department of Environmental Quality, 2014). Groundwater flow within the study area was not expected to be fast (Botha *et al.*, 1998; Woodford and Chevallier, 2002). Hence, sampling more frequently, for example monthly, was not advisable as this may yield statistically redundant samples. Based on the literature, parameters that should be sampled must include macro constituents, dissolved methane and ethane, anthropogenic chemicals, and gross alpha and beta radioactivity (Table 6.1.)

During the detection monitoring phase, a tiered approach was best suited for this task. Figure 6.1 illustrates the concept of the tiered approach. Here parameters from table 6.1. were grouped into three tiers based on their sampling complexity and analysis complexity. Based on the literature real-time online monitoring provided the best means of early warning detection of contamination (Vrba and Adams, 2008; Storey *et al.*, 2011). Hence, tier 1 was composed of field chemical parameters and other compounds that could easily be monitored in field, and provide indication of possible contamination (Son and Carlson, 2015). Only if threshold limits were exceeded in preceding tiers, should sampling proceed in the next tier. This mechanism of tiered approached to sampling for detection monitoring was designed to limit unnecessary data collection and thereby reduce cost.

If there was a positive exceedance in any of the threshold limits of tier 1, field sampling of tier 2 parameters should proceed. Tier 2 parameters comprised macro chemicals constituents and other compounds that could not be easily analysed in field. However, this tier include included parameters that provided a more direct line of evidence towards possible contamination, such as TOC and chloride (U.S. Environmental Protection Agency, 2011). Only if there was an exceedance in key parameters, should tier 3 sampling be undertaken. This tier warrants a comprehensive analysis and included definitive indicators of shale gas contamination, such as anthropogenic chemicals. The parameters in this tier required specialised analysis, and were thus costly to analyze.

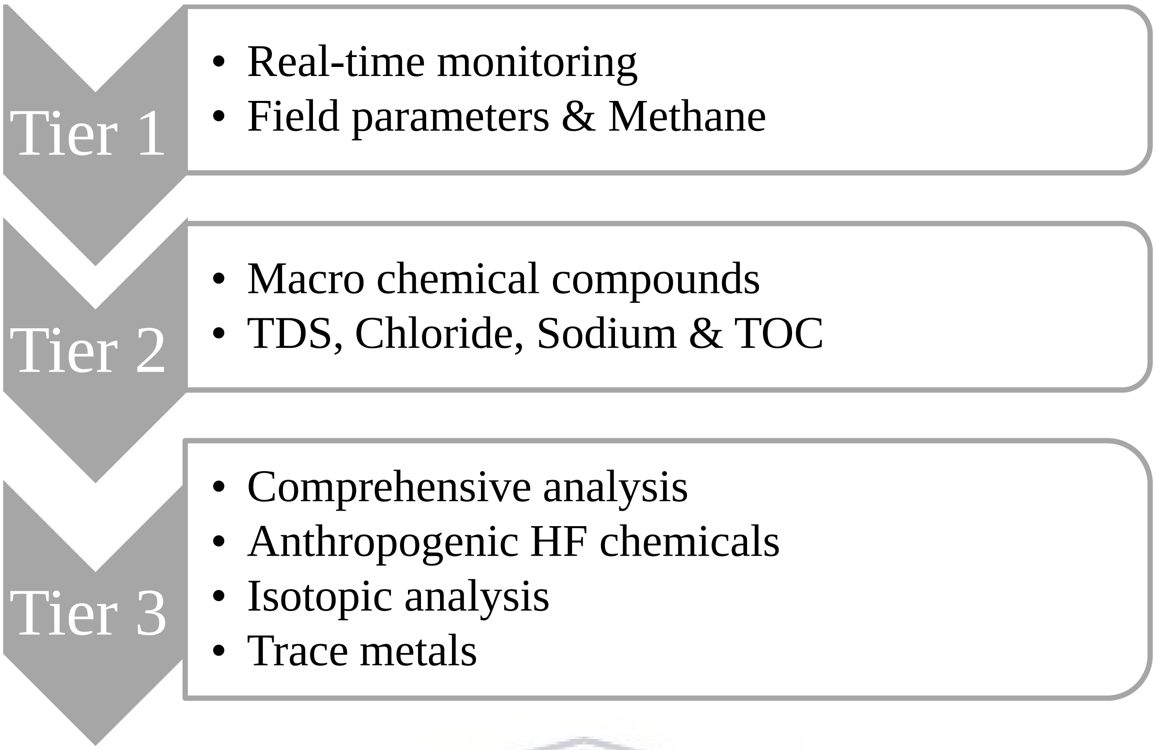


Figure 6.1. Diagram showing the tiered approach to monitoring frequency and parameters. Successive tiers are only activated when there is an exceedance beyond the threshold limit.

6.4. Discussion on parameters and frequency of monitoring

While it was important to determine monitoring network (chapter 4) and baseline (chapter 5), it was equally important to determine the target parameters and frequency of monitoring. The correct choice of monitoring parameters could facilitate identification of possible groundwater contamination due to shale gas development. In addition, an effective monitoring strategy in terms of frequency and monitoring parameters allowed for the early detection of contamination. It was of course the ultimate task of this monitoring system, to protect the groundwater environment.

The list of parameters identified in the table 6.1. was not intended to be exhaustive by any means. They represented only those that were identified in the literature. In addition, no single parameter could truly identify a source of contamination, in a shale gas contamination situation. In most cases multiple lines of evidence (multiple parameters) needed to be reviewed in-order to be conclusive. This was however a question for the operator or regulator of the monitoring system. The strongest evidence of contamination was however, provided by those anthropogenic chemicals identified in table 6.1.

From the results of chapter 5 it was determined that the current dataset did not include all of the parameters identified in table 6.1., neither was the level of data conducive to a statistical baseline analysis. Hence, in this chapter it was determined that a baseline sampling phase should be undertaken before shale gas development commences. During shale gas development a detection monitoring phase must be implemented to monitoring for potential changes in aquifer properties due to possible contamination. In the context of the study this appeared to be the best implementation strategy, in order to collect the necessary level of data. The idea was that the data collected during the baseline phase was used as a reference level during the detection monitoring phase. Thus, the multiple tasks required of the monitoring system could be achieved.

In terms of frequency, during the baseline phase the collection of quarterly samples for at least 3 years prior to shale gas development, seemed appropriated due to slow movement of groundwater. During the detection monitoring phase, it was determined that regular monitoring of a full spectrum of target parameters would not facilitate early warning detection. Thus, a more realistic and achievable tiered approach was suggested. This should make sampling a more manageable task. This type of monitoring strategy would ensure the ultimate task of protecting the groundwater resources through early warning, was realised.

Following on from O’Brien *et al.* (2013), which this part of the research was largely based on, the conclusions of this research did agree with those of (O’Brien *et al.*, 2013). Similarly, both this work and O’Brien *et al.* (2013) discuss the nature of indicator parameters and the importance of it during monitoring and baseline development. However, this work provided an extended an understanding of key parameters towards implementation within a regional groundwater monitoring system.

From a methodological point of view, the literature provided no empirical methods to determine the most appropriate monitoring parameters. Hence, the only approach was a qualitative analysis of the literature to identify those parameters that were most appropriate in terms of groundwater baseline and indicators. Applicability of these analytes as indicators had been suggested based on a theoretical understanding. There was limited evidence to suggest that these indicator parameters actually provided useful lines of evidence (U.S. Environmental Protection Agency, 2011). For example, theoretically there should not naturally be any anthropogenic chemicals in the groundwater. However, influence of non shale gas sources of pollution, such as landfills and fuel tanks, may introduce these compounds into the groundwater. Hence, these compounds should be included in the baseline phase sampling.

When discussing the monitoring frequency, Zhou (1996) showed the used of statistical means to estimate the optimized sampling frequency for a local scale monitoring system in the Netherlands. However, such an approach required a time-series with a high resolution. Papapetridis and Paleologos (2012) showed that monthly sampling was best for a wide range of hydrogeological conditions. They used contaminant transport modelling to predict the best frequency of sampling, best on plume movement and groundwater flow rates. The lack of data limited the application of these two approaches. Hence, other researchers such as the Department of Environmental Quality (2014) advised on using conceptual understanding of groundwater movement to determine monitoring frequency. This approach was followed in this research. Groundwater flow in the study area was expected to be slow and not pervasive, hence quarterly sampling seemed more appropriate. In addition, this was ideal for seasonality analysis.

The proposed set-up of monitoring parameters and sampling frequency described in this chapter appeared to be the best manner to achieve the objectives of the monitoring system. Real time monitoring was a must for timely detection of contamination, while optimized distribution of the monitoring points as well as regular sampling could help establish a baseline. This would mean that data management software was needed to be employed to manage the network. Advanced data management software packages were indeed available that can receive, analyze, interpret and output real-time data. In such a way the aquifers could be monitored remotely and allow timely intervention to take place. Furthermore, baseline sampling should continue through the detection monitoring phase, in order to maintain a statistically robust baseline.

6.5. Summary of results and discussion

In objective 3, the literature was reviewed in order to develop a list of parameters that must be sampled during monitoring. This resulted in 6 classes of chemical parameters that include macro inorganic chemical constituents, trace elements, inorganic compounds and dissolved gasses, and radiochemistry and isotopes. Anthropogenic chemicals commonly used in hydraulic fracturing fluid, were also included. A pre-development baseline monitoring phase should be undertaken with no more than quarterly sampling, due to slow groundwater movement. A detection monitoring phase should follow during shale gas development. During detection monitoring phase, in order to manage the large list of parameters, the sampling frequency plan followed a tiered approach. Only exceedance of thresholds set per tier warranted sampling in the next tier. Nevertheless, it was determined that no more than quarterly samples should be collected, based on groundwater flow rates.

Chapter 7 : Conclusions & Recommendations

7.1. Introduction

In the context of shale gas development, groundwater may be vulnerable to contamination. It was determined that the current monitoring system was not adequate to provide the necessary level of data to protect vital groundwater resources. Thus, this research was focused on understanding what the best monitoring system could be, that could protect groundwater against shale gas development. This aim required the designed monitoring system to be able to collect pre-development baseline data and provide detection monitoring during shale gas development. In chapters 4, 5 and 6, the design of an applicable groundwater monitoring system was laid out. Chapter 4 (objective 1) dealt with designing the configuration of the monitoring network. Chapter 5 (objective 2) dealt with understanding the current dataset in-order to develop a statistically sound baseline. Lastly, chapter 6 (objective 3) highlighted the target parameters and the frequency of monitoring. The following chapter summarizes the results achieved in this study and provides future recommendations.

7.2. Conclusions

Objective 1 was concerned with the design of a groundwater monitoring network, in terms of number and location of monitoring points. The focus was thus, to determine the most appropriate spatial configuration of monitoring points according to the aim of this research. A hybrid geostatistical hydrogeological methodology was used, as it allowed optimization of the number and location of monitoring points. Depth to water level measurements for the month of October - April (2014-2015) were collected from the current monitoring network. This dataset was used in a kriging analysis to determine the spatial autocorrelation range, which defines the density of the monitoring network. Based on the results, the required density (number of monitoring points) was approximately 1 monitoring point every 16688 m. Using this density and various hydrogeological features, such as contaminant transport pathways and potential receptors, new monitoring points were positioned across the study area. Using this approach, the current monitoring network was expanded to a total of 95 monitoring. The new monitoring network increased the density of monitoring points in areas that had minimal coverage. Revaluating the newly designed network using the kriging prediction error maps, showed a decrease in the kriging prediction error. This indicated that the newly designed monitoring network performed better at providing the necessary level of data.

For objective 2, the purpose was to determine the baseline conditions of the aquifer with regards to groundwater chemistry. In-order to determine a baseline for the groundwater chemistry, a statistical approach was used to analyse the existing dataset. Groundwater chemistry data was collected from the current groundwater quality monitoring points through sampling and historical groundwater data collection. The results of this analyses indicated that there were only 5 data points in the study area that had groundwater chemistry data with a time series more than 20 independent samples. The analysis of these 5 points revealed that only 32.5% of the analytes were normally distributed. The key findings in a hydrogeological sense were the fact that no seasonal variations occurred within the dataset; however, 37% of the analytes displayed a secular or long-term trend. This possibly meant that the aquifer was in a non-steady state. All analysis were performed at the 95 percent confidence level, which alludes to the reliability of the data used. Hence, under these conditions it was advisable not to establish a baseline using the existing dataset.

Objective 3 focused on determining a monitoring strategy in terms of monitoring parameters and the frequency of monitoring, in order to collect baseline data and provide detection monitoring. Therefore, it was necessary to determine which parameters could be used as indicators for shale gas contamination. In-order to determine the suite of monitoring parameters needed, an analysis of the literature was undertaken. This method resulted in the identification of a comprehensive table of chemical parameters, such as macro chemical constituents, trace elements, shale tracers, organic compounds, dissolved gasses, anthropogenic chemicals, radiochemistry and isotopes. In addition, many of the chemical parameters identified in this study could be used to as indicators/tracers for shale gas contamination, such as Cl, TDS, TOC, thermogenic methane, and anthropogenic chemical, amongst others. However, it was determined that no single indicator parameter can conclusively identify shale gas contamination, based on possible multiple sources for specific indicators. Rather multiple lines of evidence must be used to determine contamination.

In terms of frequency, it was determined sampling should be divided into a pre-development baseline monitoring phase, and a detection monitoring phase during shale gas development. For the baseline monitoring phase, the groundwater flow rate was used to determine that quarterly samples were sufficient. However, during the detection monitoring phase, a tiered approached was best suited to the goal of detection monitoring. In this strategy, real-time monitoring of EC, pH, and methane can provide early warning. Only if there were anomalous recordings in this tier, should sampling proceed to the next tiers, which included a comprehensive analysis of all parameters identified in table 6.1. It was determined that such an implementation strategy would facilitate the collection of baseline data and provide the necessary level of protection to groundwater resources.

7.3. Recommendations

During the design of the groundwater monitoring network, the possibility of non-stationary dataset might have affected the validity of the results. Based on such finding, the current research recommends that further optimization to the network take place as more robust data become available. Thus, research recommended that at least 3 years of quarterly samples be collected across the case study prior to shale gas development to improve the current data set. This sampling should continue through the detection monitoring phase. In addition, the trending nature of the data and seasonality within the data must be investigated from a hydrogeological view point as well. Lastly, in-order to provide a truly robust monitoring system the fate and transport of shale gas contaminants within Karoo geology needs to be understood.

7.4. Concluding remarks

The results of the 3 objectives have been summarised in sections 7.2. These three objectives represented the 3 general components of a groundwater monitoring system in its entirety. Thus, the results achieved in this research represent the most applicable monitoring system within the scope of this research, which could protect valuable groundwater resources in the South Western Karoo. Furthermore, in the future more advance methodology such as contaminant transport modelling can be employed to develop an early warning system. Finally, it was suggested that the monitoring system designed in this research be implemented at least three years prior to shale gas development in-order to collect the necessary baseline data.

References

Ahmadi, S. H. and Sedghamiz, A. (2007). Geostatistical analysis of spatial and temporal variations of groundwater level. *Environmental Monitoring and Assessment*. 129(1–3) pp. 277–294.

Al-Mussawi, W. H. (2008). Kriging of Groundwater Level - A Case Study of Dibdiba Aquifer in Area of Karballa-Najaf. *Journal of Kerbala University*. 6(1) pp. 170–182.

Atkins, M. L., Santos, I. R. and Maher, D. T. (2015). Groundwater methane in a potential coal seam gas extraction region. *Journal of Hydrology: Regional Studies*. Elsevier B.V. 4pp. 452–471.

Atkinson, P. M. and Lloyd, C. D. (2007). Non-stationary variogram models for geostatistical sampling optimisation: An empirical investigation using elevation data. *Computers and Geosciences*. 33(10) pp. 1285–1300.

Bamberger, M. and Oswald, R. E. (2014). Unconventional oil and gas extraction and animal health. *Environmental Science: Processes & Impacts*. pp. 1860–1865.

Barcelona, M., Gibb, J. P., Helfrich, J. A. and Garske, E. E. (1985). Practical guide for ground-water sampling. *ISWS Contract Report 374*. Champaign, Illinois: Illinois State Water Survey p. 103.

Barrow, D. and Conrad, J. (2010). Groundwater exploration and development - Beaufort West.pdf. Stellenbosch: Geohydrological and Spatial Solutions International - GEOSS p. 174.

Bell, R. A., Darling, W. G., Manamsa, K. and Dochartaigh, B. E. O. (2016). The Baseline Concentrations of Methane in Great British Groundwater - the National Methane Baseline Survey. Nottingham: British Geological Survey.

Bhat, S., Motz, L. H., Pathak, C. and Kuebler, L. (2015). Geostatistics-based groundwater-level monitoring network design and its application to the Upper Floridan aquifer, USA. *Environmental monitoring and assessment*. 187(1) pp. 4183–4197.

Botha, J. F., Verwey, J. P., van der Voort, I., Viviers, J. J. P., Buys, J., Collistion, W. P. and Loock, J. C. (1998). Karoo Aquifers: Their Geology, Geometry and Physical Properties. *Water Research Commission*. Pretoria: Water Research Commission p. 212.

Boyer, C., Lewis, R., Miller, C. K. and Clark, B. (2011). Shale Gas : A Global Resource. *Oilfield Review*. 23(3) pp. 28–39.

Brits, J. (2012). Merweville groundwater situation and future monitoring. Pretoria: Department of Water & Sanitation p. 109.

Caeiro, S., Painho, M., Goovaerts, P., Costa, H. and Sousa, S. (2003). Spatial sampling design for sediment quality assessment in estuaries. *Environmental Modelling & Software*. 18(10) pp. 853–859.

Center for Energy and Water Sustainability (2014). *Colorado Water Watch*. Available at: http://cewc.colostate.edu/research/cews-cww/ (Accessed: 1 August 2017).

Chadalavada, S. and Datta, B. (2008). Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers. *Water Resources Management*. 22(6) pp. 651–670.

Chapman, M., Gurley, L. and Fitzgerald, S. (2014). Baseline Well Inventory and Groundwater-Quality Data from a Potential Shale Gas Resource Area in Parts of Lee and Chatham Counties , North Carolina , October 2011 – August 2012 Data Series 861. Reston, Virginia: U.S. Geological Survey p. 22.

Chen, K., Ni, M., Cai, M., Wang, J., Huang, D., Chen, H., Wang, X. and Liu, M. (2016). Optimization of a Coastal Environmental Monitoring Network Based on the Kriging Method: A Case Study of Quanzhou Bay, China. *BioMed Research International*. 2016p. 12.

Chevallier, L., Goedhart, M. and Woodford, A. . (2001). The Influences of Dolerite Sill and Ring Complexes on the Occurrence of Groundwater in Karoo Fractured Aquifers: A Morpho-Tectonic Approach. Pretoria: Water Research Commission p. 165.

Colborn, T., Kwiatkowski, C., Schultz, K. and Bachran, M. (2011). Natural Gas Operations from a Public Health Perspective. *Human and Ecological Risk Assessment*. 17pp. 1039–1056.

Committee On Energy and Commerce (2011). United States House of Representatives. Washington, D.C.: United States House of Representatives p. 32.

Considine, T. J., Watson, R. W., Considine, N. B. and Martin, J. P. (2013). Environmental regulation and compliance of Marcellus Shale gas drilling. *Environmental Geosciences*. 20(1). Available at: http://eg.geoscienceworld.org/content/20/1/1 (Accessed: 31 July 2017).

Council for Geoscience (1979). 1:250 000 Geological series: 3222 Beaufort West. Pretoria: Council for Geoscience.

Council for Geoscience (1983). 1:250 000 Geological series: 3220 Sutherland. Pretoria: Council for Geoscience.

Darrah, T. H., Vengosh, A., Jackson, R. B., Warner, N. R. and Poreda, R. J. (2014). Noble gases identify the mechanisms of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. *Proc Natl Acad Sci U S A*. 111(39) pp. 14076–14081.

Davies, R. J. (2011). Methane contamination of drinking water caused by hydraulic fracturing remains unproven. *Proceedings of the National Academy of Sciences*. 108(43) pp. E871–E871.

Davies, R. J., Almond, S., Ward, R. S., Jackson, R. B., Adams, C., Worrall, F., Herringshaw, L. G., Gluyas, J. G. and Whitehead, M. A. (2014). Oil and gas wells and their integrity: Implications for shale and unconventional resource exploitation. *Marine and Petroleum Geology*. Elsevier Ltd 56pp. 239–254.

Davies, R. J., Mathias, S. A., Moss, J., Hustoft, S. and Newport, L. (2012). Hydraulic fractures: How far can they go? *Marine and Petroleum Geology*. Elsevier Ltd 37(1) pp. 1–6.

Department of Environmental Quality (2014). Statistical guidance for determining background ground water quality and degradation. Boise, Idaho: Idaho Department of Environmental Quality p. 78.

Department of Water & Sanitation (2017a). *Groundwater Level Monitoring Network*. Available at: http://www.dwa.gov.za/Groundwater/data/NGwLMN.aspx (Accessed: 1 August 2017).

Department of Water & Sanitation (2017b). *National Groundwater Quality Monitoring Project (NGwQMP)*. Available at: http://www.dwa.gov.za/Groundwater/NGQMP.aspx (Accessed: 1 August 2017).

Department of Water Affairs and Forestry (2002a). 1:500 000 Hydrogeological map series of the Republic of South Africa: Beaufort West. Pretoria: Department of Water Affairs and Forestry.

Department of Water Affairs and Forestry (2002b). 1: 500 000 Hydrogeological map series of the Republic of South Africa: Calvinia. Pretoria: Department of Water Affairs and Forestry.

Department of Water Affairs and Forestry (2006). Groundwater resource assessment II - task 3aE recharge. *Groundwater Resource Assessment Phase II*. Pretoria: Department of Water Affairs and Forestry p. 85.

Department of Water Affairs and Forestry (2009). Strategy and Guideline Development for National Groundwater Planning Requirements: Review of GRA1, GRA2 and international Assessment Methodologies. Pretoria: Department of Water Affairs and Forestry pp. 1–65.

DiGiulio, D. C., Wilkin, R. T. and Miller, C. (2011). Investigation of Ground Water Contamination near Pavillion, Wyoming. *Draft Report*. Washington, D.C.: United States Environmental Protection Agency p. 121.

Down, A., Schreglmann, K., Plata, D. L., Elsner, M., Warner, N. R., Vengosh, A., Moore, K., Coleman, D. and Jackson, R. B. (2015). Pre-drilling background groundwater quality in the Deep River Triassic Basin of central North Carolina, USA. *Applied Geochemistry*. Elsevier Ltd 60pp. 3–13.

Duane, M. J., Welke, H. J., Allsopp, H. L. and Wilsher, W. A. (1989). U-Pb isotope systematics, ages and genesis of Karoo uranium deposits, South Africa. *South African Journal of Geology*. 92(1) pp. 49–64.

Dutta, D., Das Gupta, A. and Ramnarong, V. (1998). Design and optimization of a ground water monitoring system using GIS and multicriteria Decision Analysis. *Groundwater monitoring & Remediation*. 18(1) pp. 139–147.

Edmands, J. D., Brabander, D. J. and Coleman, D. S. (2001). Uptake and mobility of uranium in black oaks: Implications for biomonitoring depleted uranium-contaminated groundwater. *Chemosphere*. 44(4) pp. 789–795.

Edmunds, W. M., Shand, P., Hart, P. and Ward, R. S. (2003). The natural (baseline) quality of groundwater: A UK pilot study. *Science of the Total Environment*. 310(1–3) pp. 25–35.

Flewelling, S. A. and Sharma, M. (2014). Constraints on upward migration of hydraulic fracturing fluid and brine. *Groundwater*. 52(1) pp. 9–19.

Gallegos, T. J. and Varela, B. a (2015). Trends in Hydraulic Fracturing Distributions and

Treatment Fluids, Additives, Proppants, and Water Volumes Applied to Wells Drilled in the United States from 1947 through 2010— Data Analysis and Comparison to the Literature. Reston, Virginia: United States Geological Survey p. 15.

Gandossi, L. (2013). An overview of hydraulic fracturing and other formation stimulation technologies for shale gas production. *JRC Technical Reports*. Luxembourg: European Commission pp. 1–64.

Geel, C., Schulz, H. M., Booth, P., De Wit, M. and Horsfield, B. (2013). Shale gas characteristics of Permian black shales in South Africa: Results from recent drilling in the Ecca Group (Eastern Cape). *Energy Procedia*. Elsevier B.V. 40pp. 256–265.

Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. New York: Van Nostrand Reinhold Company.

Gomo, M. and Vermeulen, D. (2015). An investigative comparison of purging and non-purging groundwater sampling methods in Karoo aquifer monitoring wells. *Journal of African Earth Sciences*. Elsevier Ltd 103pp. 81–88.

Gorody, A. W. (2012). Factors affecting the variability of stray gas concentration and composition in groundwater. *Environmental Geosciences*. 19(1) pp. 17–31.

Gullick, R. W., Gaffney, L. J., Crockett, C. S., Schulte, J. and Gavin, A. J. (2004). Developing regional early warning systems for US source waters. *Journal of American Water Works Association*. 96(6) pp. 68–82.

Gullick, R. W., Grayman, W. M., Dieninger, R. A. and Males, R. M. (2003). Design of early warning monitoring systems for source waters. *Journal / American Water Works Association*. 95(11) pp. 58–72.

Gunatilaka, A. and Dreher, J. (2003). Use of real-time data in environmental monitoring: current practices. *Water Science and Technology*. 47(2) pp. 53–61. Available at: http://wst.iwaponline.com/content/47/2/53 (Accessed: 1 August 2017).

Hackley, K. C., Coleman, T. M. and Pelphrey, S. (2012). Collection and Analysis of Gas Samples from Groundwater. in *Stray Gas Conference*. Champaign, Illinois: Isotech Laboratories.

Haluszczak, L. O., Rose, A. W. and Kump, L. R. (2013). Geochemical evaluation of flowback brine from Marcellus gas wells in Pennsylvania, USA. *Applied Geochemistry*. Elsevier Ltd 28pp. 55–61.

Ingraffea, A. R., Wells, M. T., Santoro, R. L. and Shonkoff, S. B. C. (2014). Assessment and risk analysis of casing and cement impairment in oil and gas wells in Pennsylvania, 2000-2012. *Proceedings of the National Academy of Sciences*. 111(30) pp. 10955–10960.

International Energy Agency (2017). Natural Gas information : Overview. Paris: International Energy Agency p. 9.

Isotech Laboratories (2014). Collection of Ground Water Samples from Domestic and Municipal Water Wells for Dissolved Gas Analysis Using Gas Bottles. Champaign, Illinois: Isotech Laboratories p. 1.

Jackson, R. B., Vengosh, A., Carey, J. W., Davies, R. J., Darrah, T. H., Sullivan, F. O. and Gabrielle, P. (2014). The Environmental Costs and Benefits of Fracking. *Annual Review of Environmental Resources*. 39pp. 327–362.

Jackson, R. B., Vengosh, A., Darrah, T. H., Warner, N. R., Down, A., Poreda, R. J., Oshborn, S. G., Zhao, K. and Karr, J. D. (2013). Increased stray gas abundance in a subset of drkng water wells. *Proceedings of the National Academy of Sciences of the United States of America*. 110(28) pp. 11250–11255.

Johnston, D. (2007). Regulatory monitoring and testing: Groundwater sampling. Adelaide: South Australia Environmental Protection Authority p. 55.

Junez-Ferreira, H. E., Herrera, G. S., Gonzalez-Hita, L., Cardona, A. and Mora-Rodriguez, J. (2016). Optimal design of monitoring networks for multiple groundwater quality parameters using a Kalman filter: application to the Irapuato-Valle aquifer. *Environmental monitoring and assessment*. 188(1) p. 39.

Khader, A. I. and McKee, M. (2014). Use of a relevance vector machine for groundwater quality monitoring network design under uncertainty. *Environmental Modelling & Software*. Elsevier Ltd 57pp. 115–126.

Kim, N. J., Cho, M. J. and Woo, N. C. (1995). Developing A National Groundwater-Monitoring Network In Korea. *Hydrogeology Journal*. 3(4) pp. 89–94.

King, G. E. (2012). Hydraulic Fracturing 101: What Every Representative, Environmentalist, Regulator, Reporter, Investor, University Researcher, Neighbor and Engineer Should Know About Estimating Frac Risk and Improving Frac Performance in Unconventional Gas and Oil Wells. in *Proceedings of the SPE Hydraulic Fracturing Technology Conference*. Woodlands: Society of Petroleum Engineers p. 80 pp.

Kirchner, J., Van Tonder, G. J. and Lukas, E. (1991). Exploitation Potential of Karoo Aquifers. *Water Research Commission Report No 170/1/91*. Pretoria: Water Research Commission p. 284.

Kuhn, M. and Kempka, T. (2013). Numerical modelling of interactions between deep and shallow groundwater aquifers. in *International Association of Hydrogeologists 40th International Congress*. Perth: International Association of Hydrogeologists p. 267. Available at: http://gfzpublic.gfz-potsdam.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:247909 (Accessed: 31 July 2017).

Kuhn, M. and Umbach, F. (2011). Strategic Perspectives of Unconventional Gas: A Game Changer with Implications for the EU’s Energy Security. *EUCERS Strategy Paper*. London: European Centre for Energy and Resource Security pp. 1–56. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Strategic+Perspectvies+of+Unconventional+Gas:+A+Game+Changer+with+Implications+for+the+EU’s+Energy+Security#0.

Kumar, V. and Remadevi (2006). Kriging of Groundwater Levels – A Case Study. *Journal of Spatial Hydrology*. 6(1) p. 12.

Li, H. and Carlson, K. H. (2014). Distribution and Origin of Groundwater Methane in the Wattenberg Oil and Gas Field of Northern Colorado. *Environmental Science & Technology*. 48(3) pp. 1484–1491.

Li, Y. and Chan Hilton, A. B. (2005). Reducing Spatial Sampling in Long-Term Groundwater Monitoring Networks Using Ant Colony Optimization. *International Journal of Computational Intelligence Research*. 1(1) pp. 19–28.

Loaiciga, H., Charbeneau, R., Everett, L., Fogg, G., Hobbs, B. and Rouhani, S. (1992). Review of ground-water quality monitoring network design. *Journal of Hydraulic Engineering*. 118(1) pp. 11–37. Available at:

http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9429(1992)118:1(11).

MacKay, D. J. and Stone, T. J. (2013). Potential Greenhouse Gas Emissions Associated with Shale Gas Extraction and Use. *URN 13D/230*. London: United Kingdon Department of Energy & Climate Change pp. 1–50.

Le Maitre, D., Colvin, C. and Maherry, A. (2009). Water resources in the Klein Karoo : the challenge of sustainable development in a water-scarce area. *South African Journal of Science*. 105(January/February) pp. 39–48.

Maule, A. L., Makey, C. M., Benson, E. B., Burrows, I. J. and Scammell, M. K. (2013). Disclosure of hydraulic fracturing fluid chemical additives: analysis of regulations. *New solutions : a journal of environmental and occupational health policy : NS*. 23(1) pp. 167–187.

McKenna, S., Hart, D., Klise, K., Koch, M., Vugrin, E. D., Martin, S., Wilson, M., Cruz, V. and Cutler, L. (2010). Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems. *EPA/600/R-10/036*. Washington, D.C.: United States Environmental Protection Agency pp. 1–91.

McQueen, K. G. (2006). Identifying Geochemical Anomalies. Canberra: Department of Earth and Marine Sciences, Australian National University pp. 1–7.

Meyer, P. D., Valocchi, A. J. and Eheart, J. W. (1994). Monitoring network design to provide initial detection of groundwater contamination. *Water Resources Research*. 30(9) pp. 2647–2659.

Mikol, Y., Richardson, W., Van Der Schalie, W., Shedd, T. and Widder, M. (2007). An online real-time biomonitor for contaminant surveillance in water supplies. *Journal of American Water Works Association*. 99(2) pp. 107–115.

Murray, R., Baker, K., Ravenscroft, P., Musekiwa, C. and Dennis, R. (2012). A groundwater-planning toolkit for the Main Karoo Basin: Identifying and quantifying groundwater-development options incorporating the concept of wellfield yields and aquifer firm yields. *Water SA*. 38(3) pp. 407–416.

Murray, R., Swana, K., Miller, J., Talma, S., Tredoux, G., Vengosh, A. and Darrah, T. (2015). The use of chemistry, isotopes and gases as indicators of deeper circulating groundwater in the Main Karoo Basin. Pretoria: Water Research Commission p. 215.

Narany, T. S., Ramli, M. F., Fakharian, K., Aris, A. Z. and Sulaiman, W. N. A. (2015). Multi-objective based approach for groundwater quality monitoring network optimization. *Water Resources Management*. 29(14) pp. 5141–5156.

National Toxic Network (2013). Toxic Chemicals in the Exploration and Production of Gas from Unconventional Sources. *Toxic Chemicals in the Exploration and Production of Gas from Unconventional Sources*. Bangalow, Australia: National Toxics Network p. 12. Available at: http://www.chiefscientist.nsw.gov.au/__data/assets/pdf_file/0006/29877/CSG-Review-Submissions-0039.pdf.

Naval Facilities Engineering Command (2003). Guidance for Environmental Background Analysis - Volume III: Groundwater. Washington, D.C.: Naval Facilities Engineering command p. 203. Available at: http://web.ead.anl.gov/ecorisk/related/documents/ug-2059-bkgrnd-analysis.pdf%5Cnpapers2://publication/uuid/664358DA-BC7A-4610-8826-A471C84FE79F.

Nunes, L. M., Cunha, M. C. and Ribeiro, L. (2004). Groundwater monitoring network optimization with redundancy reduction. *Journal of water resources Planning and Management*. 130(1) pp. 33–43.

O'Brien, R., Talma, S. and Tredoux, G. (2013). The Complexity of Baseline Monitoring in Unconventional Gas Projects. in *13th Biennial Groundwater Division Conference and Exhibition*. Durban: Groundwater Division of the GSSA pp. 33–33. Available at: http://gwd.org.za/sites/gwd.org.za/files/06_R O'Brien _The Complexity of Baseline Monitoring in Unconventional Gas Projects - revised copy.pdf.

Olea, R. A. (1984). Sampling design optimization for spatial functions. *Journal of the International Association for Mathematical Geology*. 16(4) pp. 369–392.

Olmstead, S. M., Muehlenbachs, L. A., Shih, J.-S., Chu, Z. and Krupnick, A. J. (2013). Shale gas development impacts on surface water quality in Pennsylvania. *Proceedings of the National Academy of Sciences*. 110(13) pp. 4962–4967.

Osborn, S. G., Vengosh, A., Warner, N. R. and Jackson, R. B. (2011). Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proceedings of the National Academy of Sciences*. 108(20) pp. 8172–8176.

Pacome, A. D. (2010). *Fracture Characterisation of Karoo Aquifers*. University of the Free State.

Papapetridis, K. and Paleologos, E. K. (2012). Sampling Frequency of Groundwater Monitoring and Remediation Delay at Contaminated Sites. *Water Resources Management*. 26(9) pp. 2673–2688.

Parker, L. (1994). The effects of groundwater sampling devices on the water quality: A literature Review. *Groundwater monitoring & Remediation*. 14(2) pp. 130–141.

Perry, S. A. (2011). Understanding Naturally Occurring Radioactive Material in the Marcellus Shale. *Marcellus shale*. June pp. 1–8.

Petroleum Agency of South Africa (2013). *Unconventional Resources*. Available at: https://www.petroleumagencysa.com/index.php/petroleum-geology-resources/frontier-geology (Accessed: 4 August 2017).

Pietersen, K., Kanyerere, T., Levine, A., Matshini, A. and Beekman, H. E. (2016). An analysis of the challenges for groundwater governance during shale gas development in South Africa. *Water SA*. 42(3) pp. 421–431.

ProPublica (2017). *What Is Hydraulic Fracturing? Fracking*. Available at: https://www.propublica.org/special/hydraulic-fracturing-national (Accessed: 4 August 2017).

Qin, R., Cao, G., Wu, Y., Zhang, C. and Xu, Z. (2016). A method for selecting monitoring wells and measured water-quality characteristics with application to the Liaohe River (China) groundwater system. *Environmental Earth Sciences*. Springer Berlin Heidelberg 75(9) pp. 792–806.

Reimann, C. and Garrett, R. G. (2005). Geochemical background - Concept and reality. *Science of the Total Environment*. 350(1–3) pp. 12–27.

Robbins, K. (2013). Awakening the slumbering giant: How horizontal drilling technology brought the endangered species act to bear on hyraulic fracturing. *Case Western Reserve Law Review*. 63(4) pp. 1142–1166.

Rose, R. (2008). Aquifer characterization and groundwater flow regimes of Karoo aquifers in Beaufort West , Western Cape Province. Pretoria: Water Research Commission p. 52.

Rosewarne, P., Woodford, A., Goes, M., Talma, S., O'Brien, R., Tredoux, G., Esterhuyse, C., Visser, D. and Van Tonder, G. (2013). Recent Developments in Understanding of Karoo Aquifers and the Deeper Underlying Formations. in *13th Groundwater Division Conference*. Durban: Groundwater Division of the GSSA pp. 1–11.

Saar, R. (1997). Filtration of groundwater samples: A review of industry practice. *Groundwater monitoring & Remediation*. 17(1) pp. 56–62.

Schalla, R., Webber, W. D. and Smith, R. M. (2001). Selection of Sampling Pumps Used for Groundwater Monitoring at the Hanford Site. *PNNL-13690*. Washington, D.C.: U.S. Department of Energy p. 87.

Seward, P. (1988). A re-evaluation of the geohydrology of the Beaufort West area. Pretoria: Department of Water & Sanitation p. 28.

Shand, P., Edmunds, W. M., Lawrence, A. R., Smedley, P. L. and Burke, S. (2007). The natural (baseline) quality of groundwater in England and Wales. *BGS Research Report RR/07/06*. Nottingham: British Geological Survey p. 72.

Sharma, A. (2010). *DNA-Based hydrological tracers: proof of concept*. Cornell University.

Sloto, R. a (2013). Baseline Groundwater Quality from 20 Domestic Wells in Sullivan County, Pennsylvania, 2012. *Scientific Investigations Report 2013-5085*. Reston, Virginia: U.S. Geological Survey p. 27.

Solomon, H. G. (2013). *Application of multivariate statistics and Geographic Information Systems (GIS) to map groundwater quality in the Beaufort West area , Western Cape , South Africa*. University of the Western Cape.

Son, J.-H. and Carlson, K. H. (2015). Real-time surrogate analysis for potential oil and gas contamination of drinking water resources. *Applied Water Science*. 5(3) pp. 283–289.

Spruill, T. B. and Candela, L. (1990). Two approaches to design of monitoring networks. *Ground Water*. 28(3) pp. 430–442.

Statistics South Africa (2017). *Statistics by place | Statistics South Africa*. Available at: http://www.statssa.gov.za/?page_id=964 (Accessed: 2 August 2017).

Stein, H., Kellermann, C., Schmidt, S. I., Brielmann, H., Steube, C., Berkhoff, S. E., Fuchs, A., Hahn, H. J., Thulin, B. and Griebler, C. (2010). The potential use of fauna and bacteria as ecological indicators for the assessment of groundwater quality. *Journal of environmental monitoring*. 12(1) pp. 242–254.

Steube, C., Richter, S. and Griebler, C. (2009). First attempts towards an integrative concept for the ecological assessment of groundwater ecosystems. *Hydrogeology Journal*. 17(1) pp. 23–35.

Stevens, P. (2012). The ‘Shale Gas Revolution’: Developments and Changes. *EERG BP 2012/04*. London: The Royal Institute of International Affairs p. 12.

Steyl, G. and Van Tonder, G. J. (2013). *Hydrochemical and Hydrogeological Impact of Hydraulic Fracturing in the Karoo, South Africa*. *Effective and Sustainable Hydraulic Fracturing*. Edited by Bunger, Andrew P, J. McLennan, and R. Jeffrey. Rijeka, Croatia: IntechOpem.

Storck, P., Eheart, J. W. and Valocchi, A. J. (1995). A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogenous aquifers. *Water Resources Research*. 33(9) p. 2081.

Storey, M. V., van der Gaag, B. and Burns, B. P. (2011). Advances in on-line drinking water quality monitoring and early warning systems. *Water Research*. Elsevier Ltd 45(2) pp. 741–747.

Stroebel, D. and de Wit, M. (2015). A Baseline Study of Scarce Groundwater Resources in anticipation of Fracking in the Karoo. in *14th Biennial Goundwater Division Conference and Exhibition*. Muldersdrift: The Groundwater Division of the GSSA p. 13.

Subcommittee on groundwater (2013). A National Framework for Ground-Water Monitoring in the United States. Reston, Virginia: The Advisory Committee on Water Information p. 168.

Sundaram, B., Feitz, J. A., Caritat, D. P., Plazinska, A., Brodie, S. R., Coram, J. and Ransley, T. (2009). Groundweater Sampling and Analysis-A Field Guide. Canberra: Geoscience Australia p. 95.

Swain, E. and Sonenshein, R. (1994). Spatial and Temporal Statistical Analysis of Ground-water Level Network, Broward County, Florida.pdf. Reston, Virginia: U.S. Geological

Survey p. 48.

Talma, A. and Esterhuyse, C. (2013). Natural Methane in the Karoo: Its Occurrence and Isotope Clues to its Origin. in *13th Biennial Groundwater Division Conference and Exhibition*. Durban: Groundwater Division of the GSSA p. 12. Available at: http://gwd.org.za/sites/gwd.org.za/files/03_S Talma Esterhuyzen_Natural Methane in the Karoo.pdf.

Talma, A. S. and Weaver, J. M. C. (2003). Evaluation of Groundwater Flow Patterns in Fractured Rock Aquifers Using CFCS and Isotopes. Pretoria: Water Research Commission p. 128.

Theodossiou, N. and Latinopoulos, P. (2006). Evaluation and optimisation of groundwater observation networks using the Kriging methodology. *Environmental Modelling and Software*. 21(7) pp. 991–1000.

Van Tonder, G. J., de Lange, F., Steyl, G. and Vermeulen, D. (2013). Potential Impacts of Fracking on Groundwater in The Karoo Basin of South Africa. in *13th Groundwater Division Conference*. Durban: Groundwater Division of the GSSA pp. 31–31.

U.S. Energy Information Administration (2011). *The geology of natural gas resources*. *Today In Energy*. Available at: https://www.eia.gov/todayinenergy/detail.php?id=110 (Accessed: 4 August 2017).

U.S. Energy Information Administration (2016a). *Shale gas production drives world natural gas production growth*. *Today In Energy*. Available at: https://www.eia.gov/todayinenergy/detail.php?id=27512 (Accessed: 4 August 2017).

U.S. Energy Information Administration (2016b). *U.S. Shale Production*. *Natural Gas*. Available at: https://www.eia.gov/dnav/ng/hist/res_epg0_r5302_nus_bcfa.htm (Accessed: 4 August 2017).

U.S. Energy Information Administration (2017). *How much shale gas is produced in the United States? Frequently asked Questions*. Available at: https://www.eia.gov/tools/faqs/faq.php?id=907&t=8 (Accessed: 4 August 2017).

U.S. Environmental Protection Agency (2009). Statistical Analysis of Groundwater Data at RCRA Facilities—Unified Guidance. *EPA 530/R-09-007*. Washington, D.C.: United States

Environmental Protection Agency p. 884.

U.S. Environmental Protection Agency (2011). Proceedings of the Technical Workshops for the Hydraulic Fracturing Study : Fate and Transport. *EPA 600/R-11/047*. Washington, D.C.: United States Environmental Protection Agency pp. 1–122. Available at: www.usepa.gov.

U.S. Environmental Protection Agency (2016). Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States. Washington, D.C.: United States Environmental Protection Agency p. 666.

U.S. Environmental Protection Agency (2017). *The Process of Hydraulic Fracturing*. Available at: https://www.epa.gov/hydraulicfracturing/process-hydraulic-fracturing (Accessed: 4 August 2017).

Vengosh, A., Jackson, R. B., Warner, N., Darrah, T. H. and Kondash, A. (2014). A Critical Review of the Risks to Water Resources from Shale Gas Development and Hydraulic Fracturing in the United States. *Environmental Science & Technology*. 16(4) p. 6838.

Vidic, R. D., Brantley, S. L., Vandenbossche, J. M., Yoxtheimer, D. and Abad, J. D. (2013). Impact of shale gas development on regional water quality. *Science*. 340(6134) p. 1235009.

Vogel, J., Talma, A. S. and Heaton, T. (1980). The isotopic, chemical and dissolved gas concentrations in groundwater near Beaufort West. Pretoria: Council for Scientific and Industrial Research p. 34.

Vrba, J. (2001). Early warning groundwater quality monitoring. in *Workshop on the protection of groundwaters used as a source of drinking water supply*. Budapest pp. 1–2.

Vrba, J. and Adams, B. (2008). Groundwater early warning monitoring strategy: a methodological guide. Paris: UNESCO.

Wait, R. and Rossouw, R. (2014). A comparative assessment of the economic benefits from shale gas extraction in the Karoo, South Africa. *Southern African Business Review*. 18(2) pp. 1–34. Available at: http://www.ajol.info/index.php/sabr/article/view/107279.

Ward, R. S., Streetly, M. J., Singleton, A. J. and Sears, R. (2005). A framework for monitoring regional groundwater quality. *Quarterly Journal of Engineering Geology and Hydrogeology*. 37(4) pp. 271–281. Available at: http://qjegh.lyellcollection.org/content/37/4/271.short (Accessed: 1 August 2017).

Warner, N. R., Darrah, T. H., Jackson, R. B., Millot, R., Kloppmann, W. and Vengosh, A. (2014). New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environmental Science and Technology*. 48(21) pp. 12552–12560.

Warner, N. R., Kresse, T. M., Hays, P. D., Down, A., Karr, J. D., Jackson, R. B. and Vengosh, A. (2013). Geochemical and isotopic variations in shallow groundwater in areas of the Fayetteville Shale development, north-central Arkansas. *Applied Geochemistry*. Elsevier Ltd 35pp. 207–220.

Water Experts Group (2012). Karoo Groundwater Atlas: Volume 1. Cape Town: SRK Consulting.

Water Experts Group (2013). Karoo Groundwater Atlas: Volume 2. Cape Town: SRK Consulting p. 43. Available at: http://gwd.org.za/sites/gwd.org.za/files/KGEG_Karoo Groundwater Atlas Volume 2_Final_12Aug13_opt_opt.pdf.

Weaver, J. M. C., Cave, L. and Talma, A. (2007). Groundwater sampling. *WRC Report No TT 303/07*. Pretoria: Water Research Commission p. 168.

Wilde, F. (2005). National Field Manual for the Collection of Water-Quality Data. *Techniques of Water-Resources Investigations*. Reston, Virginia: U.S. Geological Survey p. 1539.

De Wit, M. J. (2011). The great shale debate in the Karoo. *South African Journal of Science*. 107(7–8) pp. 1–9.

Woodford, A. C. and Chevallier, L. (2002). Hydrogeology of the Main Karoo Basin : Current Knowledge and Future Research Needs. *WRC Report No. TT 179/02*. Pretoria: Water Research Commission p. 482.

Wu, J., Zheng, C. and Chien, C. C. (2005). Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions. *Journal of Contaminant Hydrology*. 77(1–2) pp. 41–65.

van Wyk, E. (2010). *Estimation of episodic groundwater recharge in semi-arid fractured hard rock aquifers*. University of the Free State.

van Wyk, Y. and Witthueser, K. (2011). A forced-gradient tracer test on the Hansrivier Dyke: Beaufort West, South Africa. *Water SA*. 37(4) pp. 437–444.

Xu, Y., Mahed, G., van Wyk, Y., Lin, L., Sun, X., Xiao, L. and Talma, S. (2012). Towards a Sampling and Monitoring Protocol of Radioactive Elements in Fractured Rock Aquifers for Groundwater Resource Security in Towards a Sampling and Monitoring Protocol of Radioactive Elements in Fractured Rock Aquifers for Groundwater Resource Secur. Pretoria: Water Research Commission p. 100.

Yang, F., Cao, S., Liu, X. and Yang, K. (2008). Design of groundwater level monitoring network with ordinary kriging. *Journal of Hydrodynamics, Ser. B*. 20(3) pp. 339–346.

Zhou, Y. (1996). Sampling frequency for monitoring the actual state of groundwater systems. *Journal of Hydrology*. 180(1–4) pp. 301–318.

Zhou, Y., Dong, D., Liu, J. and Li, W. (2013). Upgrading a regional groundwater level monitoring network for Beijing Plain, China. *Geoscience Frontiers*. Elsevier Ltd 4(1) pp. 127–138.

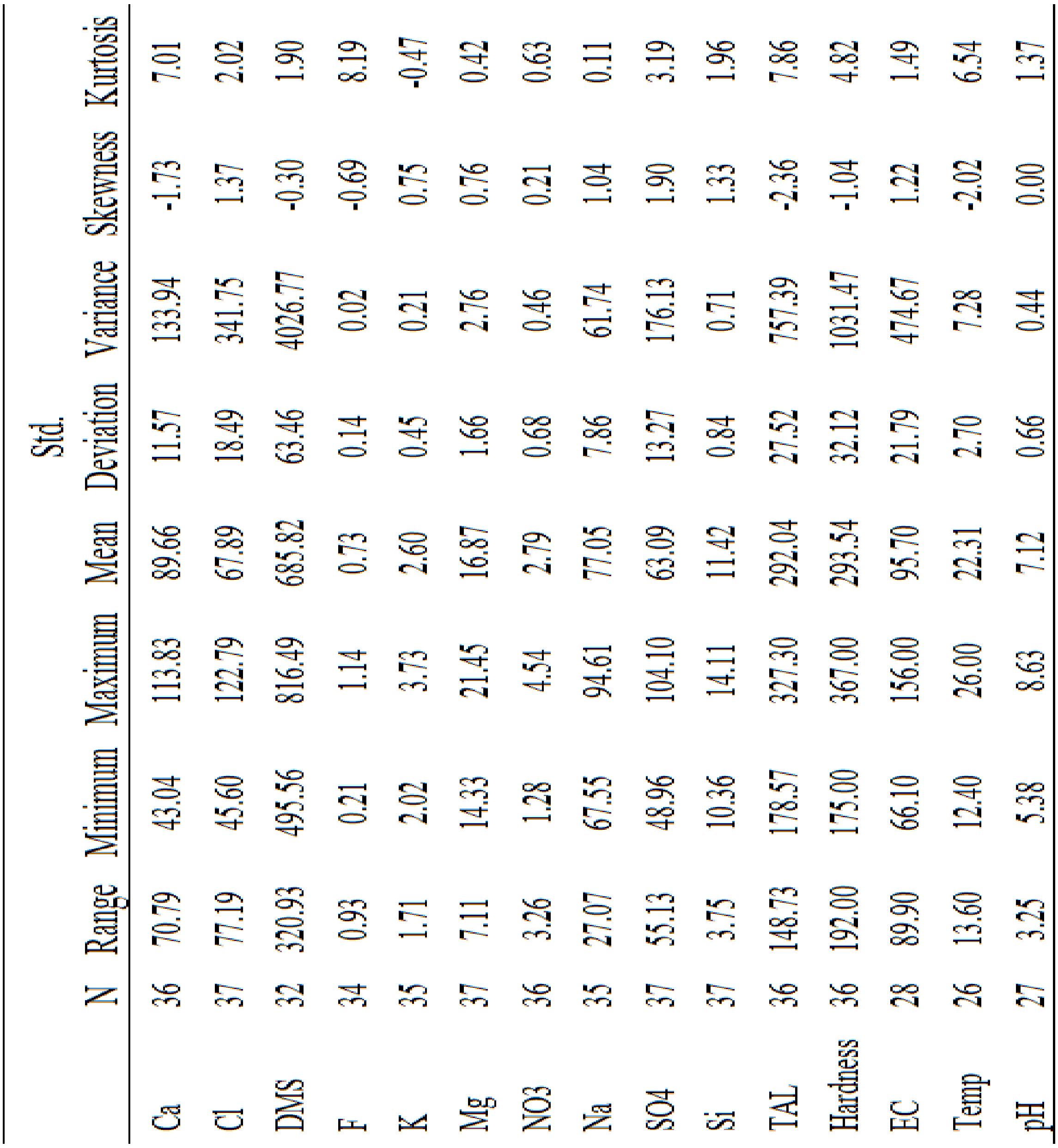
Annex 1: Supplementary statistical information

1.1. Introduction

The following annex presents supplementary information regarding statistical analysis of the available dataset. It is intended to support information already present in chapter 5 of this thesis.

1.2.1. Exploratory statistical analysis for ZQMWVL1

Table 1.1. Descriptive statistics for ZQMWVL1



	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Ca	36	70.79	43.04	113.83	89.66	11.57	133.94	-1.73	7.01
Cl	37	77.19	45.60	122.79	67.89	18.49	341.75	1.37	2.02
DMS	32	320.93	495.56	816.49	685.82	63.46	4026.77	-0.30	1.90
F	34	0.93	0.21	1.14	0.73	0.14	0.02	-0.69	8.19
K	35	1.71	2.02	3.73	2.60	0.45	0.21	0.75	-0.47
Mg	37	7.11	14.33	21.45	16.87	1.66	2.76	0.76	0.42
NO3	36	3.26	1.28	4.54	2.79	0.68	0.46	0.21	0.63
Na	35	27.07	67.55	94.61	77.05	7.86	61.74	1.04	0.11
SO4	37	55.13	48.96	104.10	63.09	13.27	176.13	1.90	3.19
Si	37	3.75	10.36	14.11	11.42	0.84	0.71	1.33	1.96
TAL	36	148.73	178.57	327.30	292.04	27.52	757.39	-2.36	7.86
Hardness	36	192.00	175.00	367.00	293.54	32.12	1031.47	-1.04	4.82
EC	28	89.90	66.10	156.00	95.70	21.79	474.67	1.22	1.49
Temp	26	13.60	12.40	26.00	22.31	2.70	7.28	-2.02	6.54
pH	27	3.25	5.38	8.63	7.12	0.66	0.44	0.00	1.37

1.2.2. Histograms for ZQMWVL1

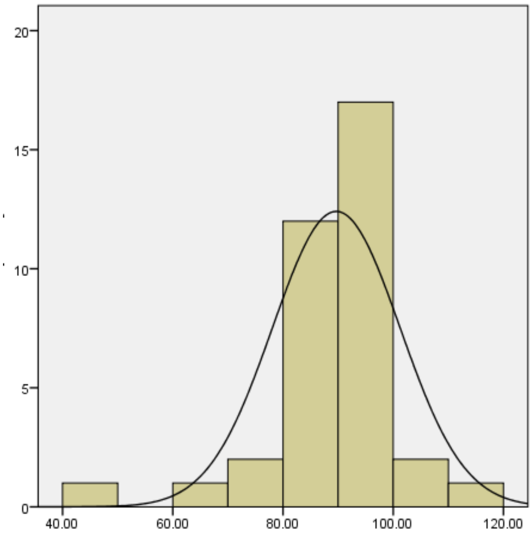


Figure 1.0.1. Histogram for Ca concentrations

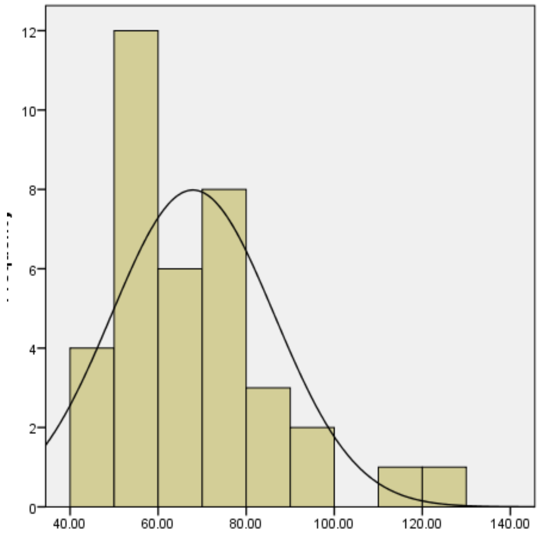


Figure 1.0.2. Histogram for Cl concentrations

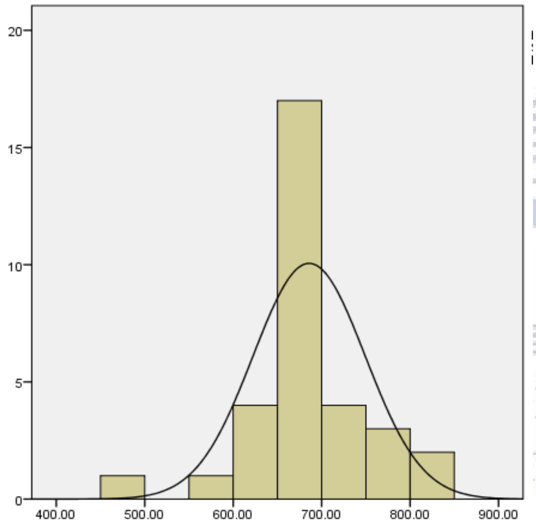


Figure 1.3. Histogram for DMS concentrations

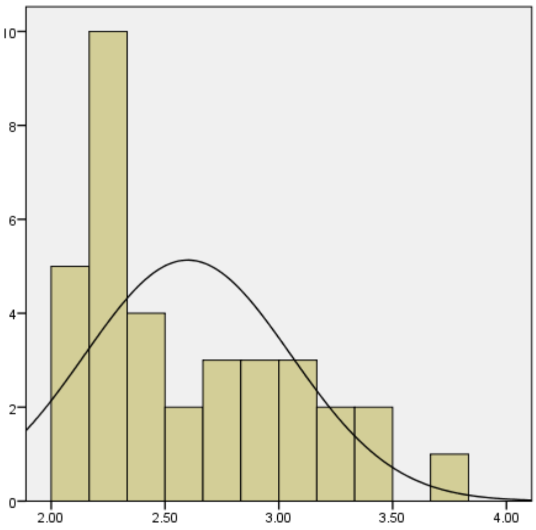


Figure 1.5. Histogram for K concentrations

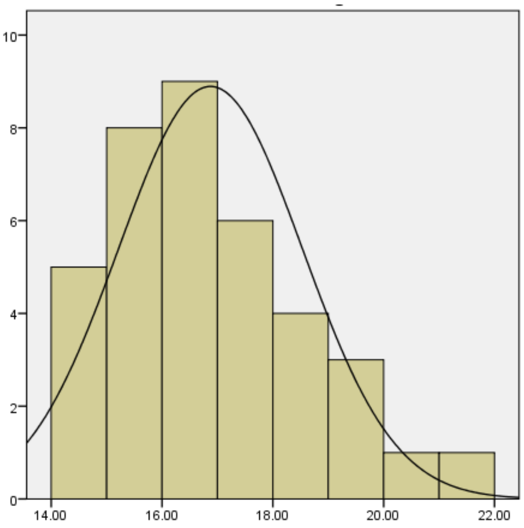


Figure 1.6. Histogram for Mg concentrations

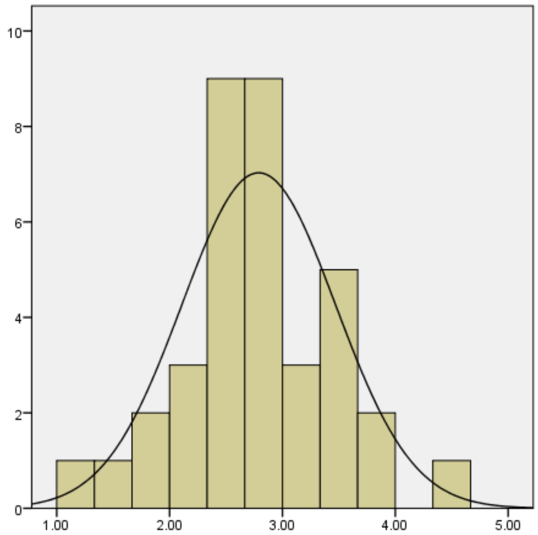


Figure 1.7. Histogram for NO_3 concentrations

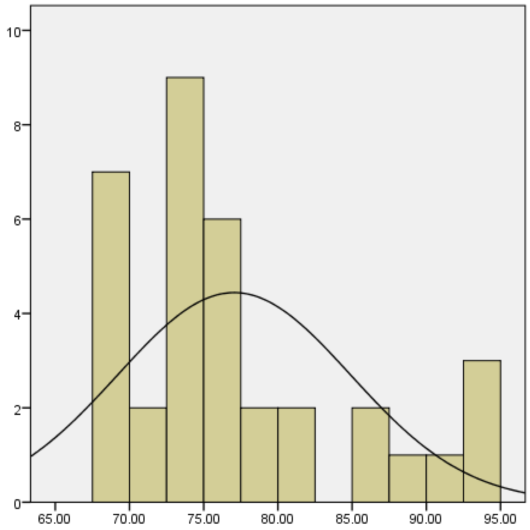


Figure 1.8. Histogram for Na concentrations

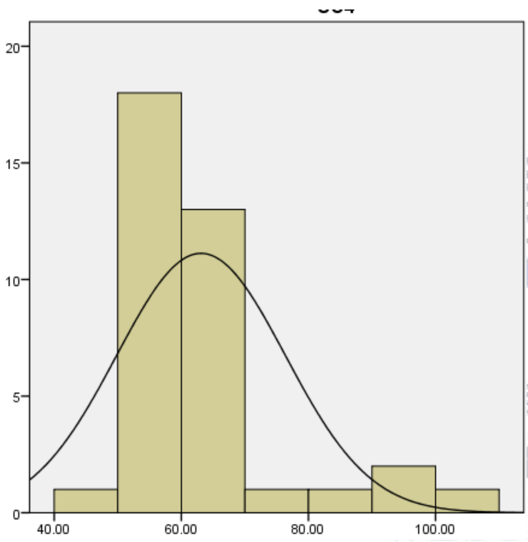


Figure 1.9. Histogram for SO_4 concentrations

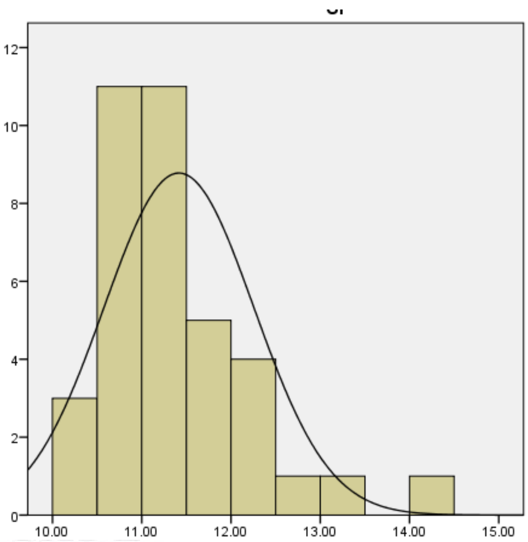


Figure 1.10. Histogram for Si concentrations

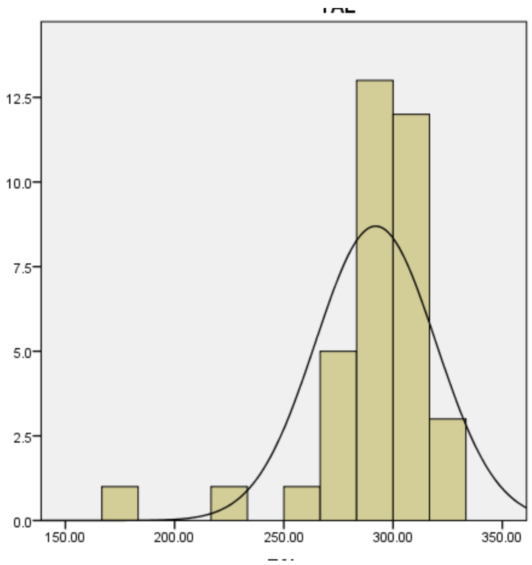


Figure 1.11. Histogram for TAL concentrations

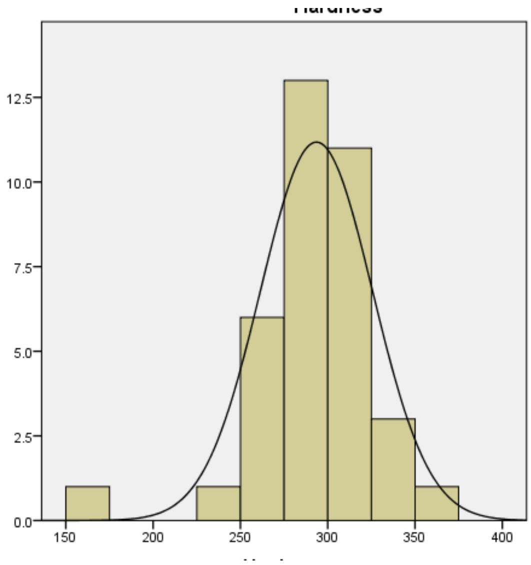


Figure 1.12. Histogram for Hardness concentrations

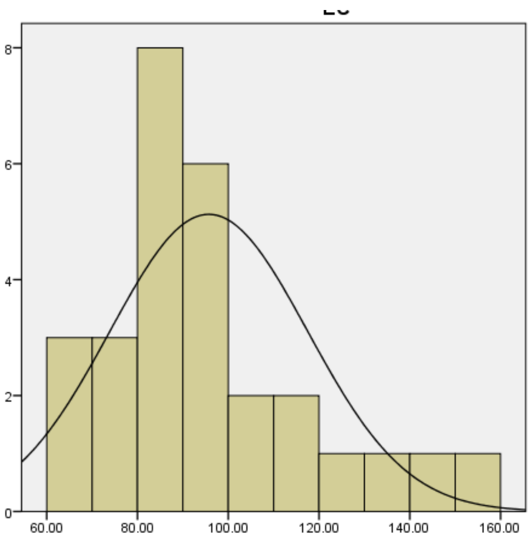


Figure 1.13. Histogram for EC observations

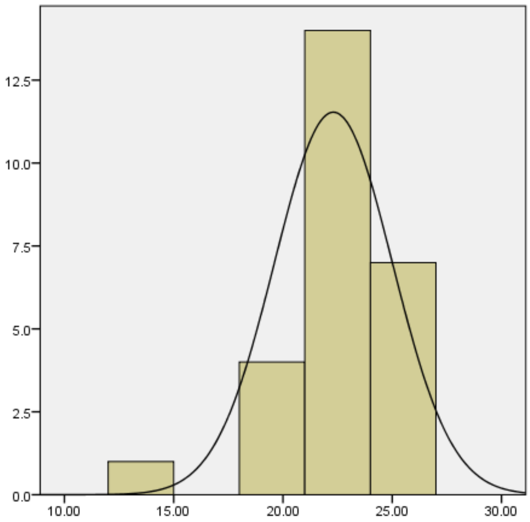


Figure 1.14. Histogram for Temp observations

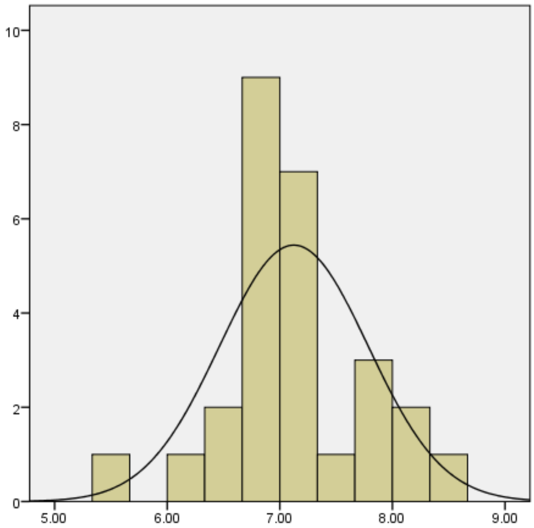


Figure 1.15. Histogram for pH observations

1.2.3. Box and Whisker plots for ZQMWVL1

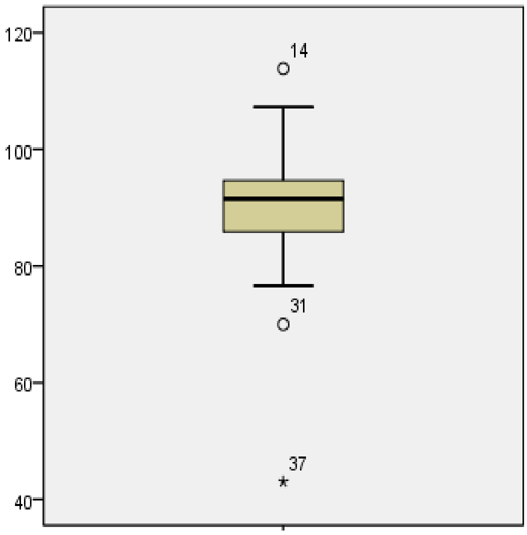


Figure 2.1. Box and whisker plot for Ca

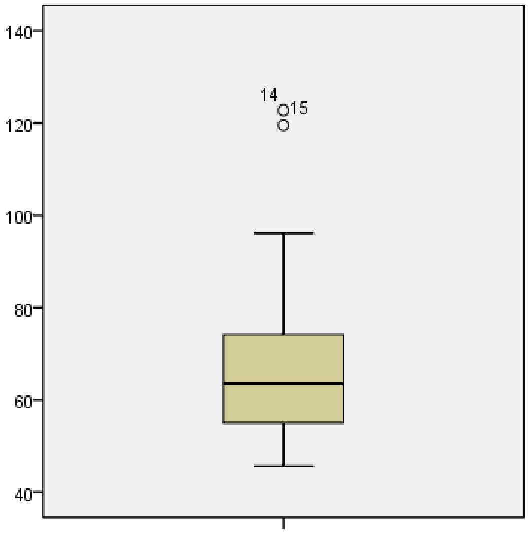


Figure 2.2. Box and whisker plot for Cl

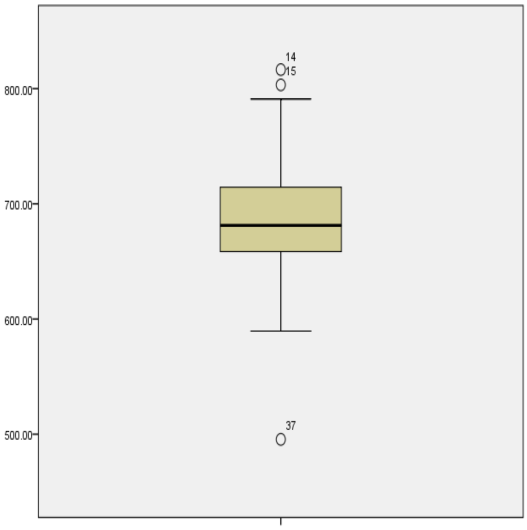


Figure 2.3. Box and whisker plot for DMS

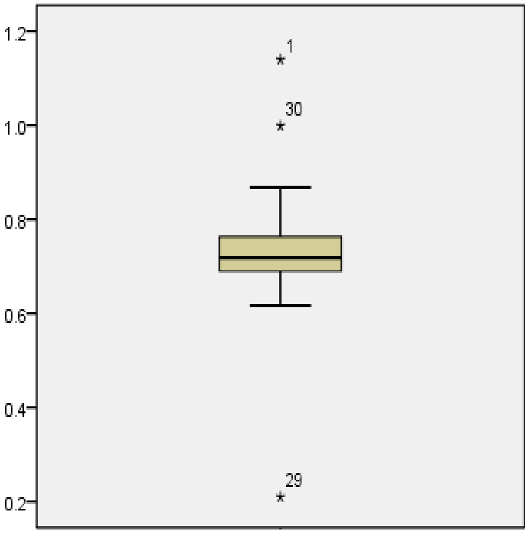


Figure 2.4. Box and whisker plot for F

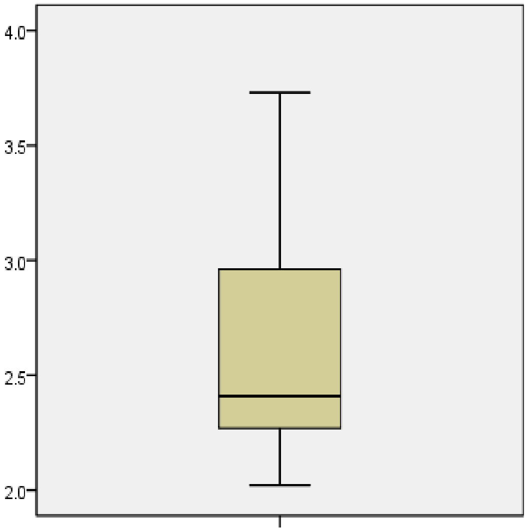


Figure 2.5. Box and whisker plot for K

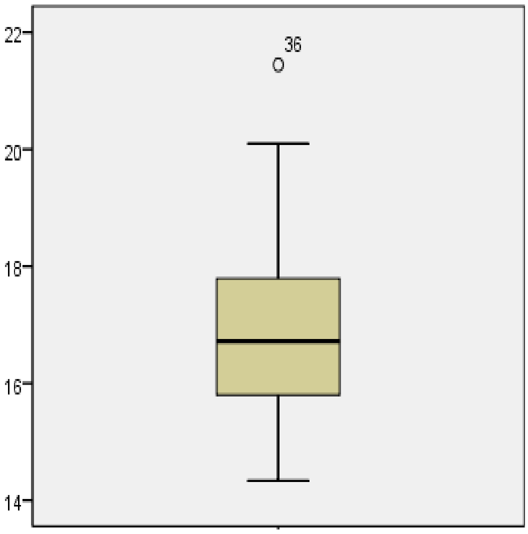


Figure 2.6. Box and whisker plot for Mg

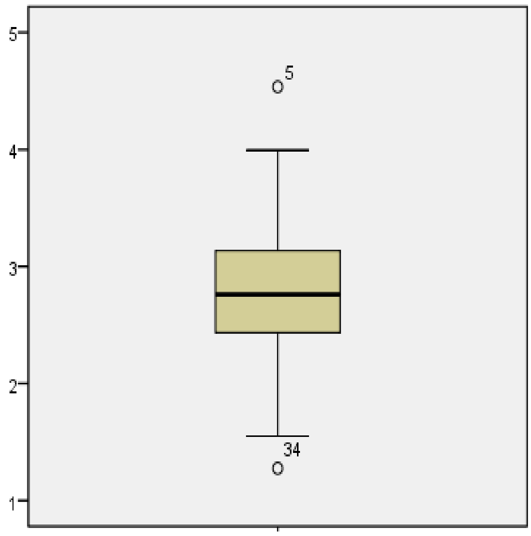


Figure 2.7. Box and whisker plot for NO_3

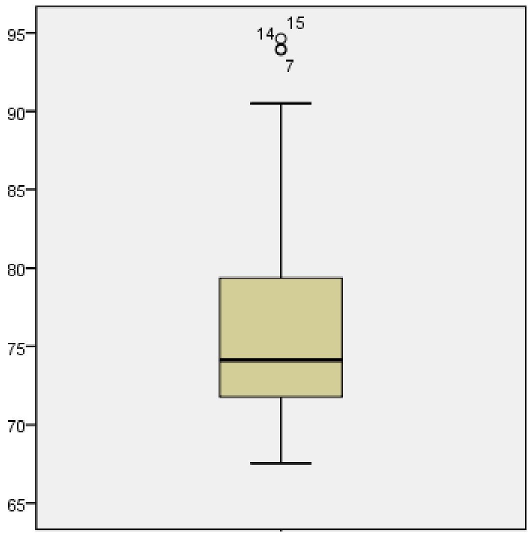


Figure 2.8. Box and whisker plot for Na

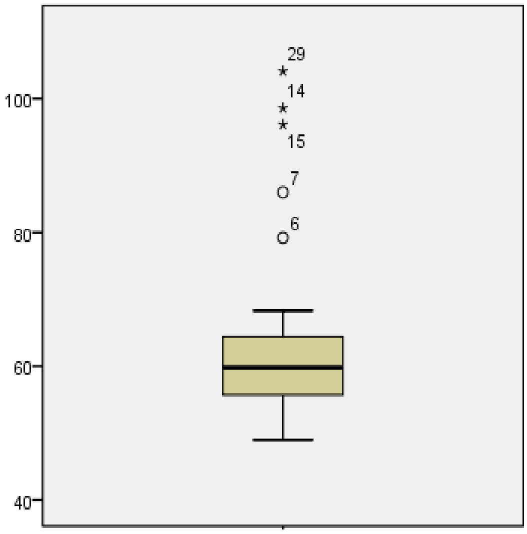


Figure 2.9. Box and whisker plot SO_4

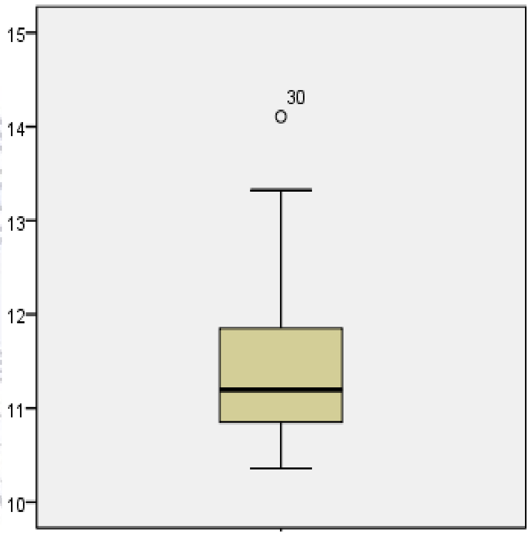


Figure 2.10. Box and whisker plot for Si

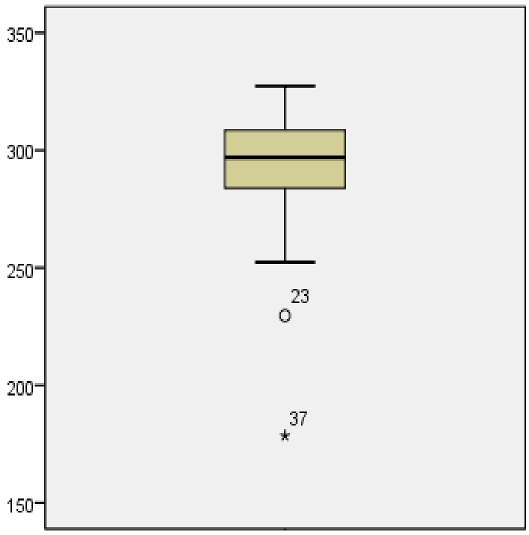


Figure 2.11. Box and whisker plot for TAL

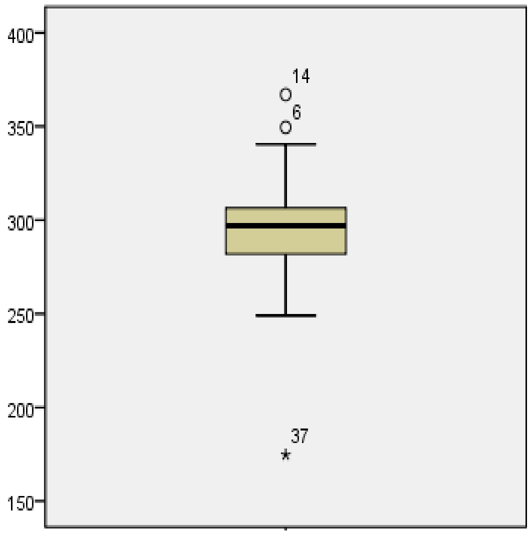


Figure 2.12. Box and whisker plot for Hardness

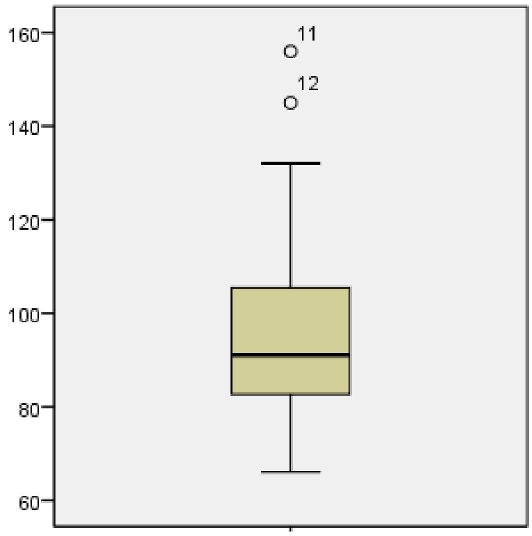


Figure 2.13. Box and whisker plot for EC

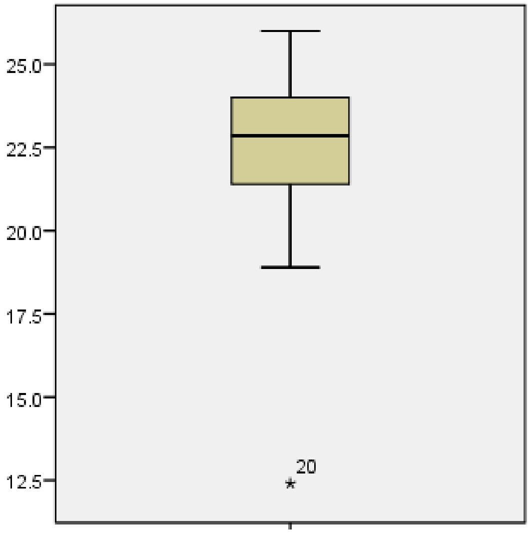


Figure 2.14. Box and whisker plot for Temp

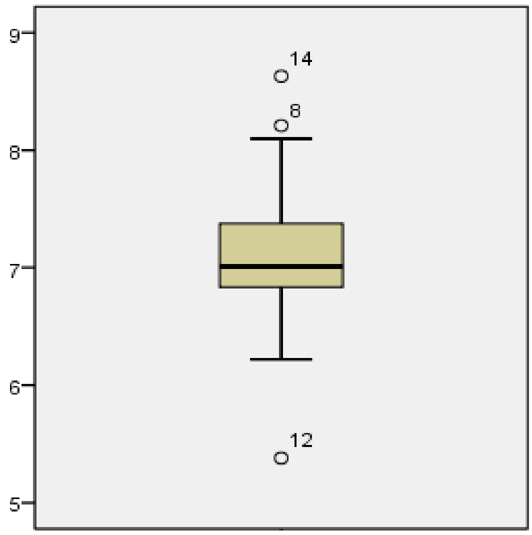


Figure 2.15. Box and whisker plot for pH

1.2.4. Time-series graphs for ZQMWVL1

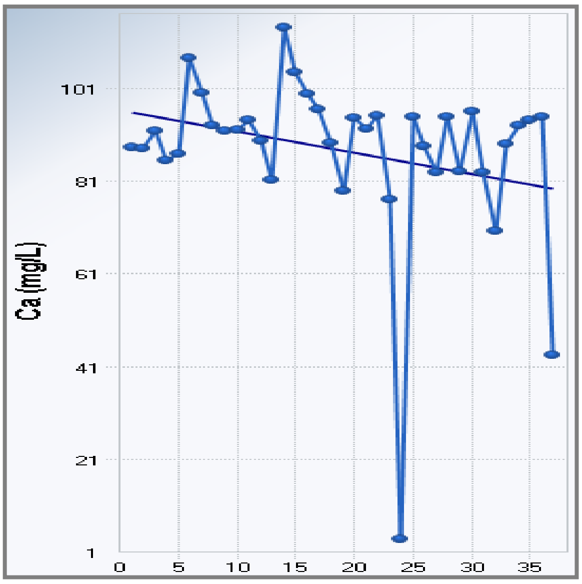


Figure 3.1. Time series graph for Ca concentrations

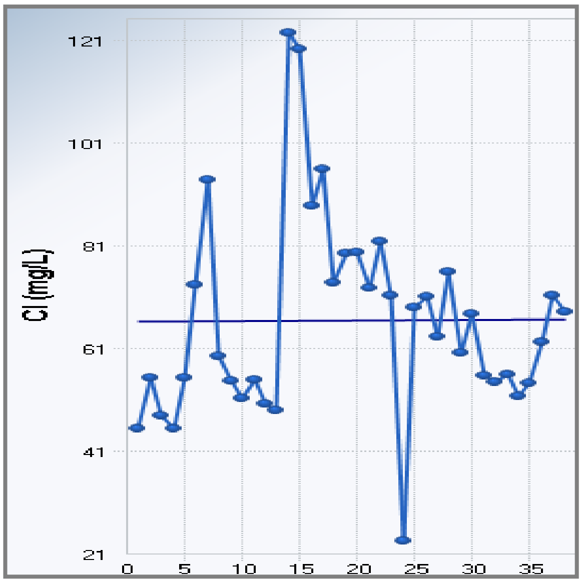


Figure 3.2. Time series graph for Cl concentrations

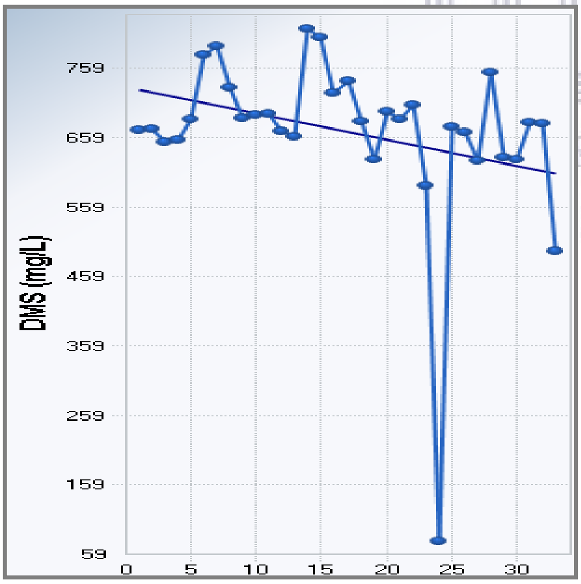


Figure 3.3. Time series graph for DMS concentrations

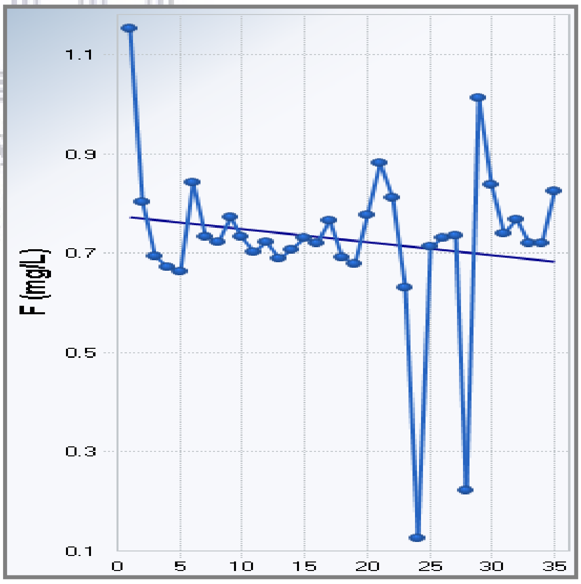


Figure 3.4. Time series graph for F concentrations

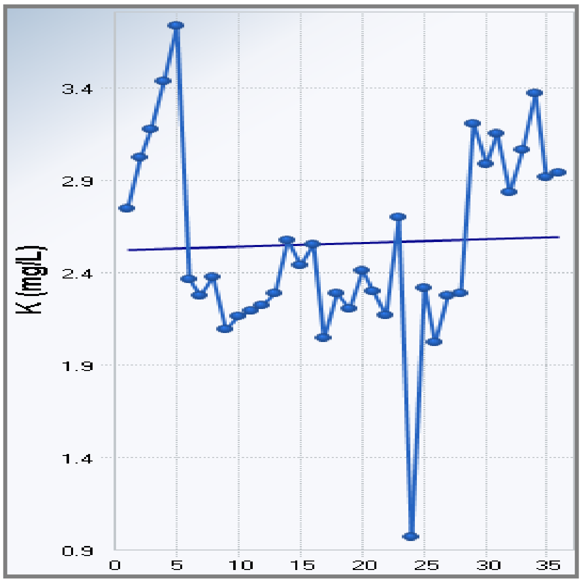


Figure 3.5. Time series graph for K concentrations

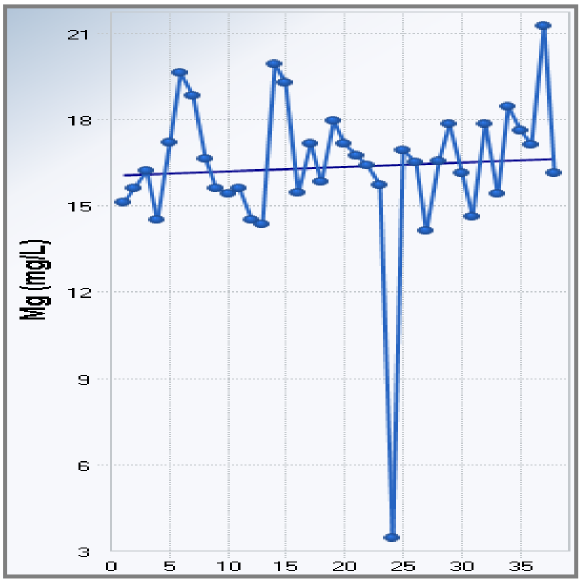


Figure 3.6. Time series graph for Mg concentrations

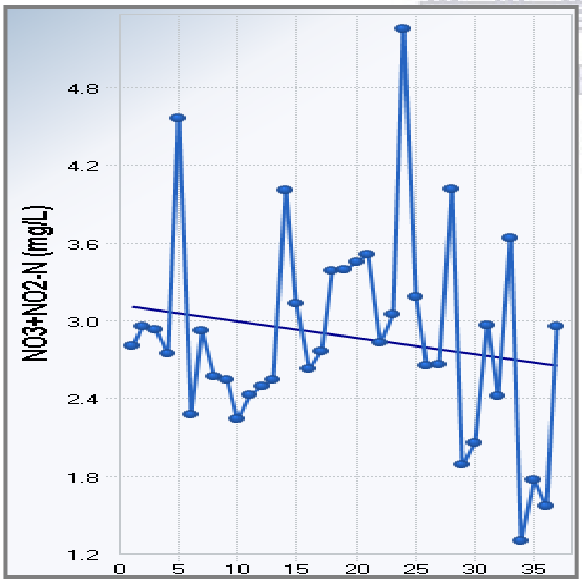


Figure 3.7. Time series graph for NO_3+NO_2 concentrations

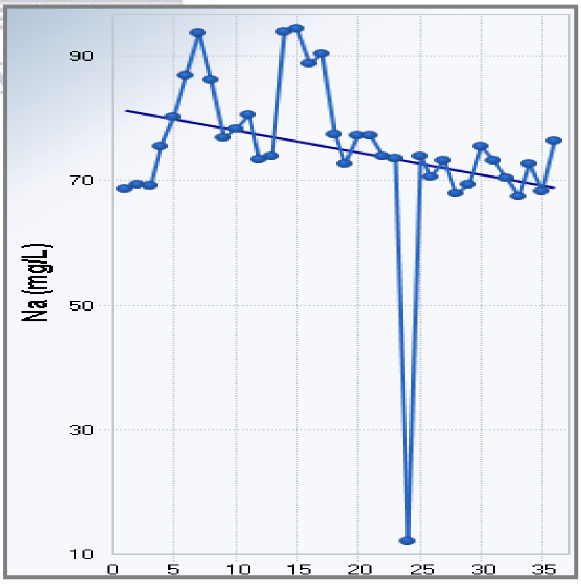


Figure 3.8. Time series graph for Na concentrations

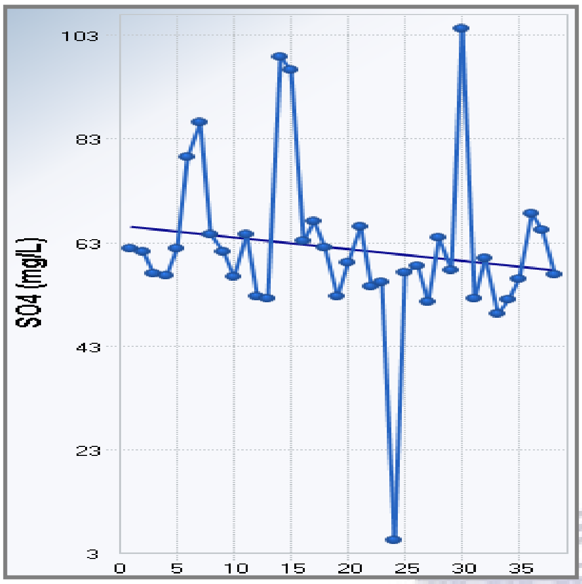


Figure 3.9. Time series graph for SO_4 concentrations

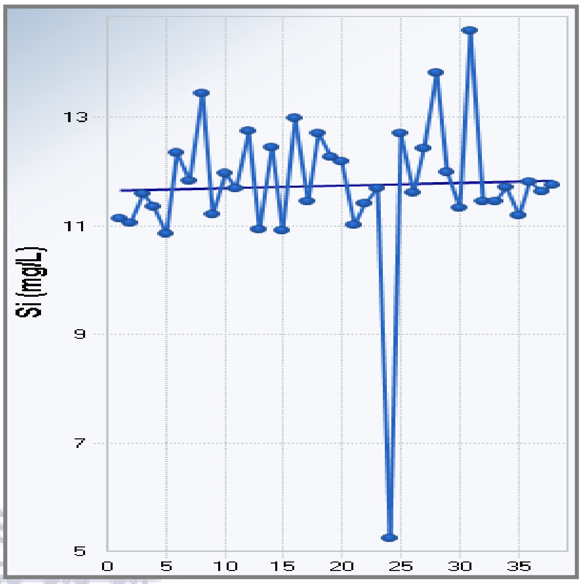


Figure 3.10. Time series graph for Si concentrations

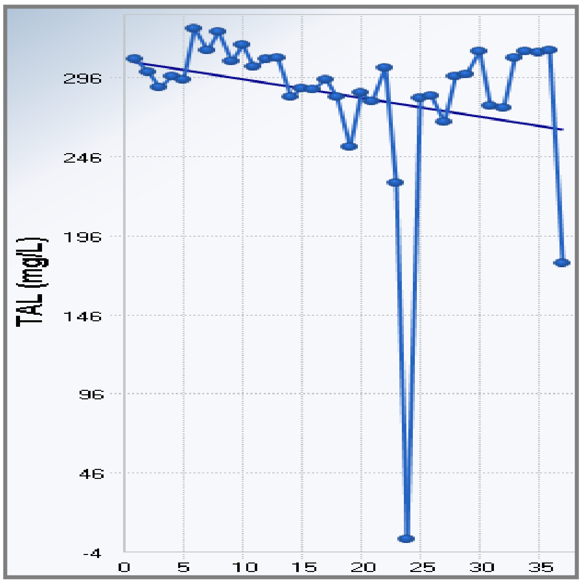


Figure 3.11. Time series graph for TAL concentrations

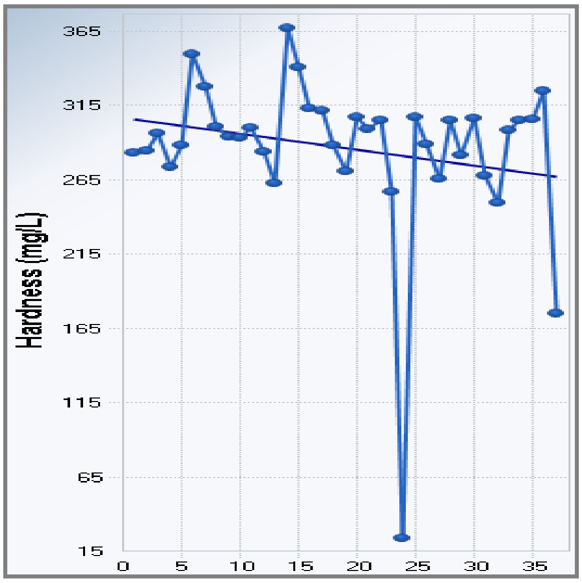


Figure 3.12. Time series graph for Hardness concentrations

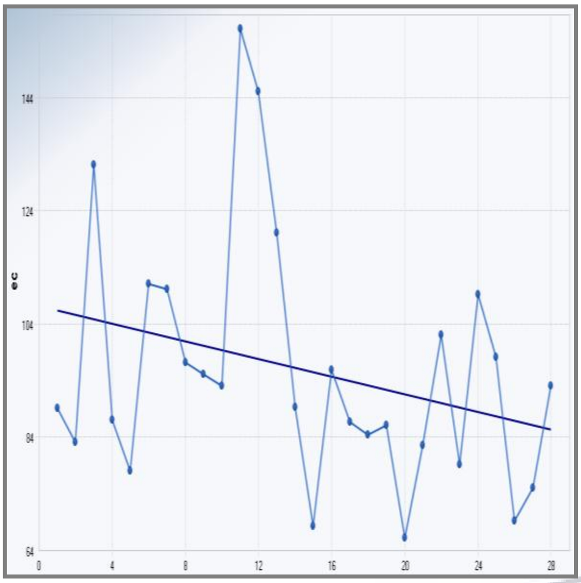


Figure 3.13 Time series graph for EC observations

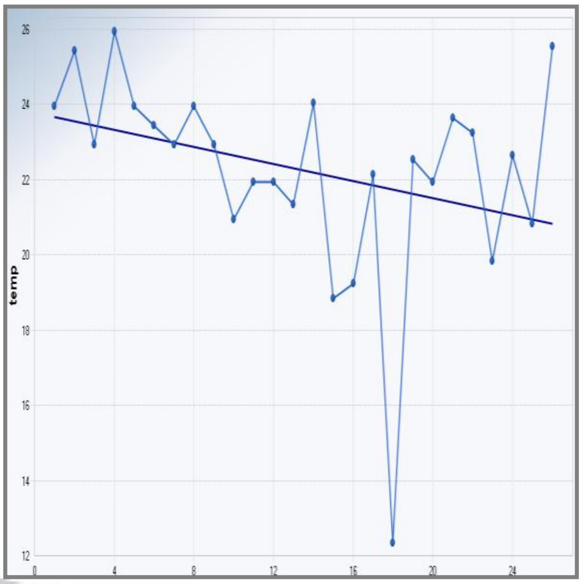


Figure 3.14. Time series graph for Temp observations

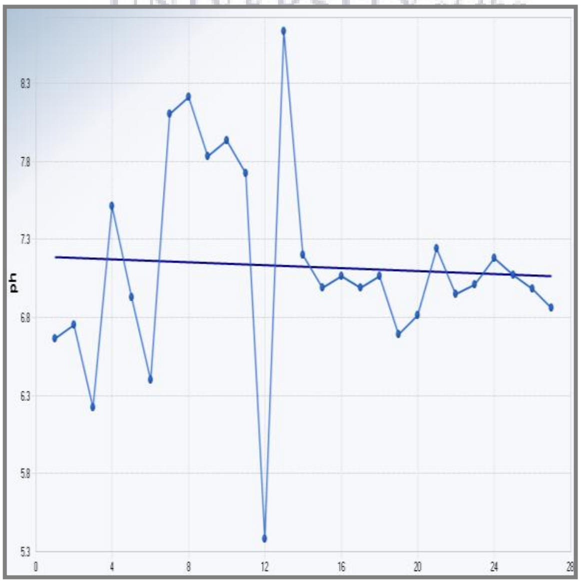


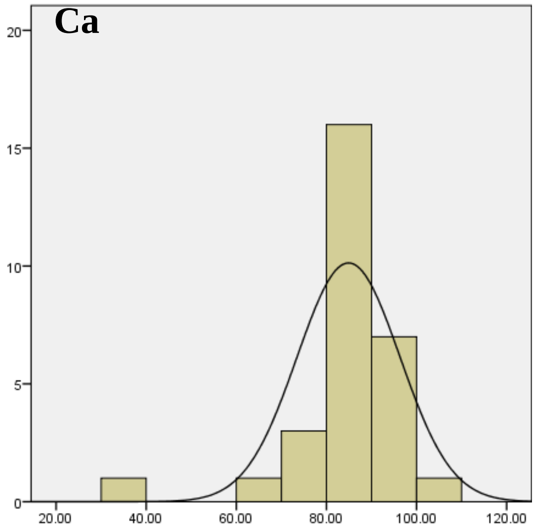
Figure 3.15. Time series graph for pH observations

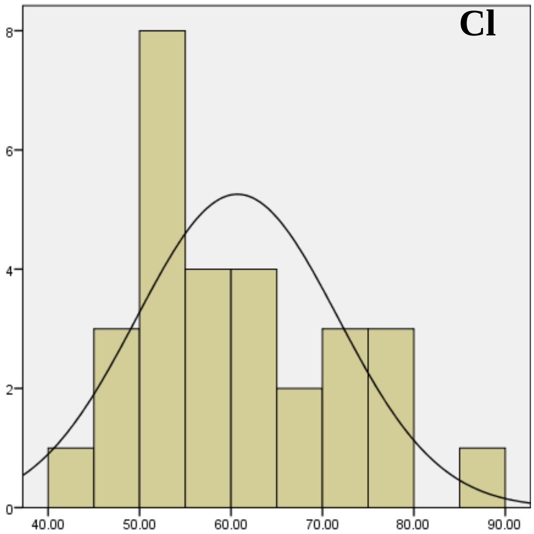
1.3.1. Exploratory statistical analysis for ZQMWVL2

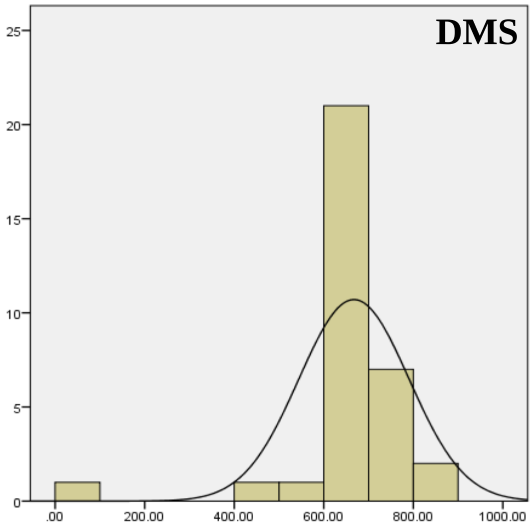
Table 1.2. Descriptive statistics for ZQMWVL2

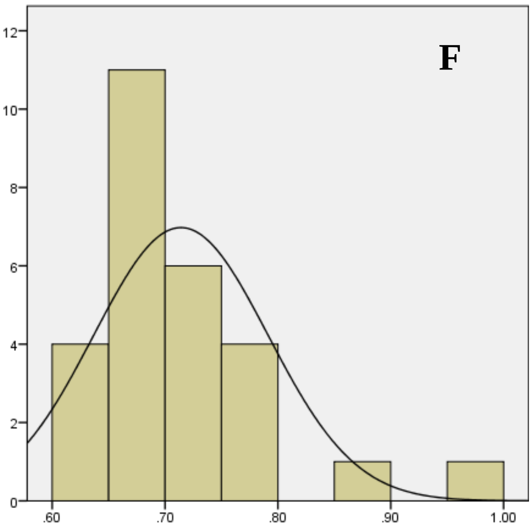
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Ca	29	62.30	38.67	100.98	84.90	11.42	130.36	-2.59	9.37
Cl	29	44.16	44.80	88.96	60.69	11.00	121.08	0.62	-0.17
DMS	25	266.36	482.57	748.93	660.31	53.09	2819.04	-1.50	4.48
F	27	0.36	0.63	0.99	0.71	0.08	0.01	1.93	5.37
K	27	2.76	1.94	4.70	2.84	0.60	0.37	1.22	1.97
Mg	29	8.02	13.26	21.28	17.19	1.59	2.54	0.16	0.89
NO3NO2	28	2.90	2.14	5.03	3.87	0.73	0.53	-0.69	0.12
Na	27	19.72	62.60	82.32	71.46	4.32	18.68	0.46	0.28
SO4	29	52.02	46.50	98.52	58.92	11.25	126.58	1.77	4.31
Si	29	4.56	9.09	13.64	10.74	0.84	0.70	1.34	4.36
TAL	28	126.78	191.47	318.25	284.99	25.38	644.31	-2.18	6.30
EC	20	122.00	26.00	148.00	89.41	25.01	625.33	0.03	2.22
Temp	20	9.40	15.60	25.00	22.65	2.13	4.54	-1.96	5.48
pH	20	1.50	6.88	8.38	7.37	0.47	0.22	0.67	-0.84

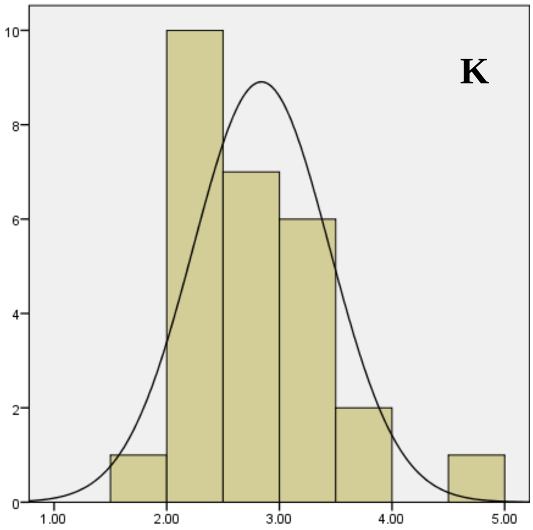
1.3.2. Histograms for ZQMWVL2

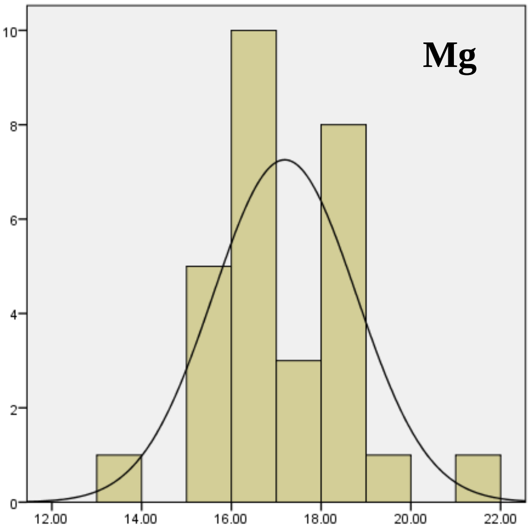


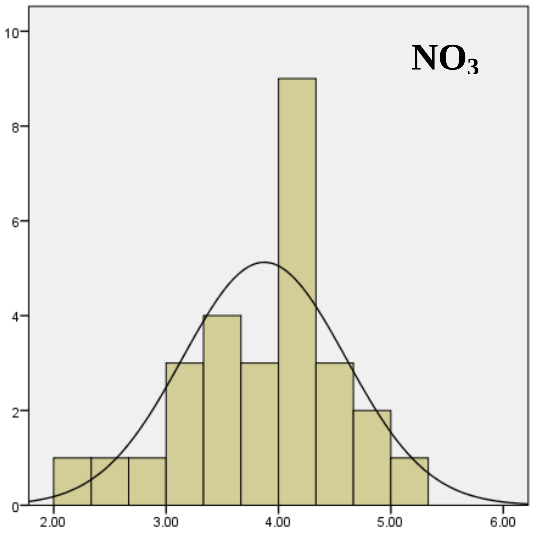


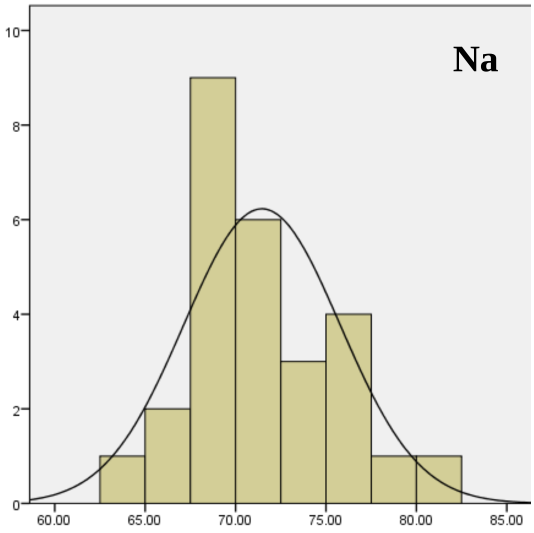


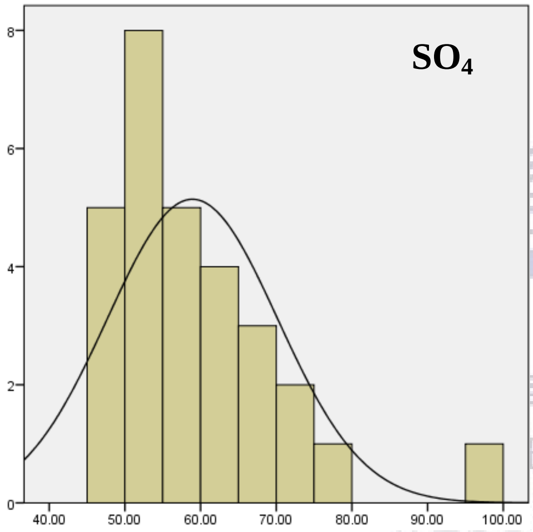


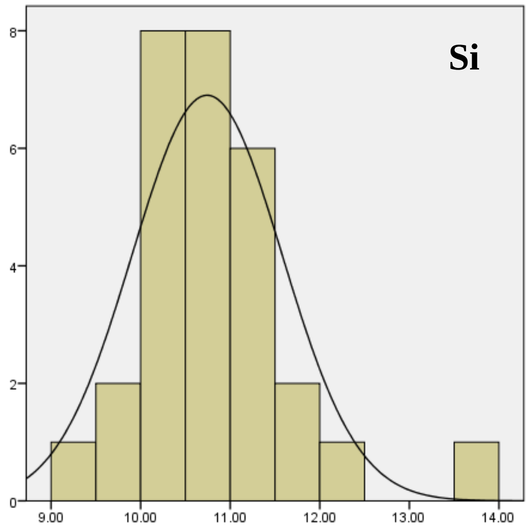


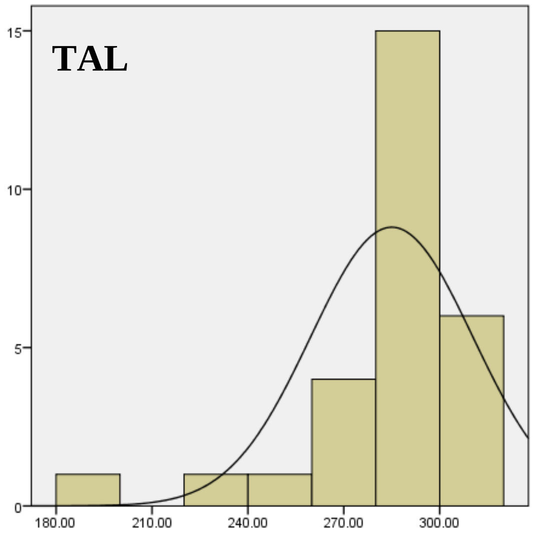


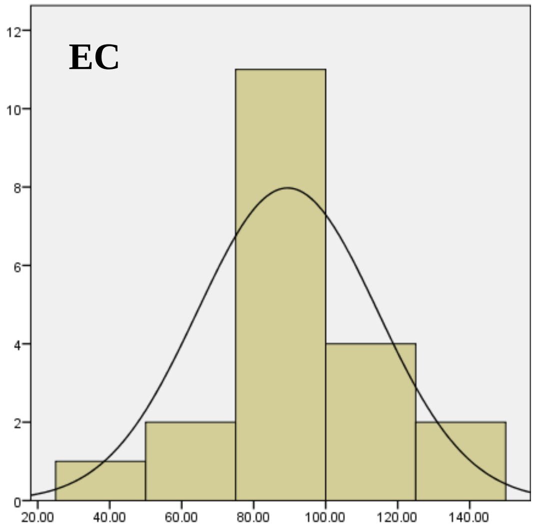

NO3
0
2
4
6
8
10
2.00
3.00
4.00
5.00
6.00

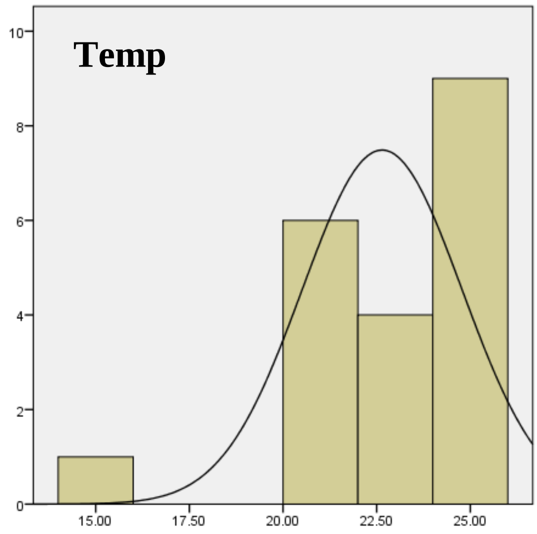

Na
0
2
4
6
8
10
60.00
65.00
70.00
75.00
80.00
85.00

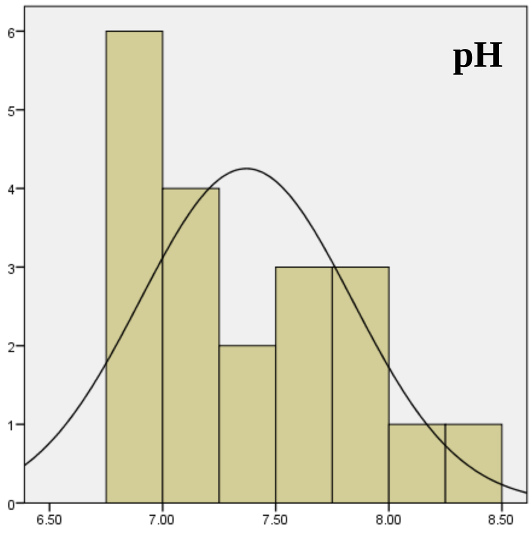

SO4
0
2
4
6
8
40.00
50.00
60.00
70.00
80.00
90.00
100.00


Si
0
2
4
6
8
9.00
10.00
11.00
12.00
13.00
14.00

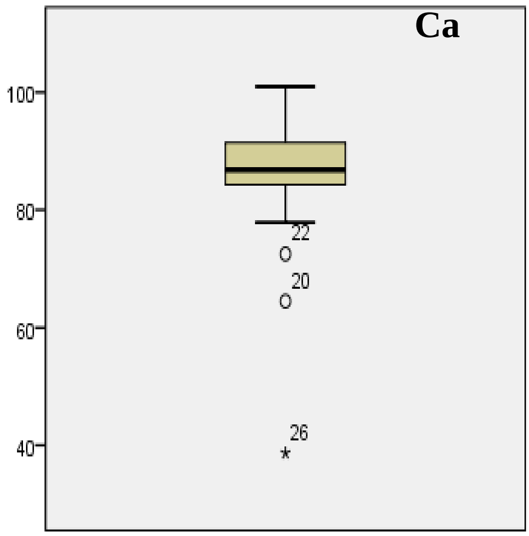

TAL
0
5
10
15
180.00
210.00
240.00
270.00
300.00

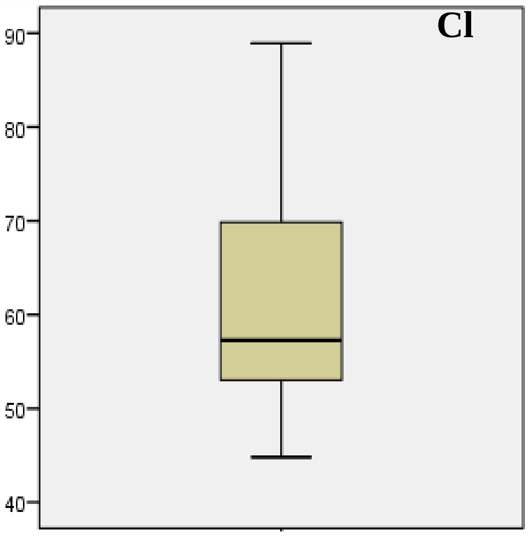

EC
0
2
4
6
8
10
12
20.00
40.00
60.00
80.00
100.00
120.00
140.00

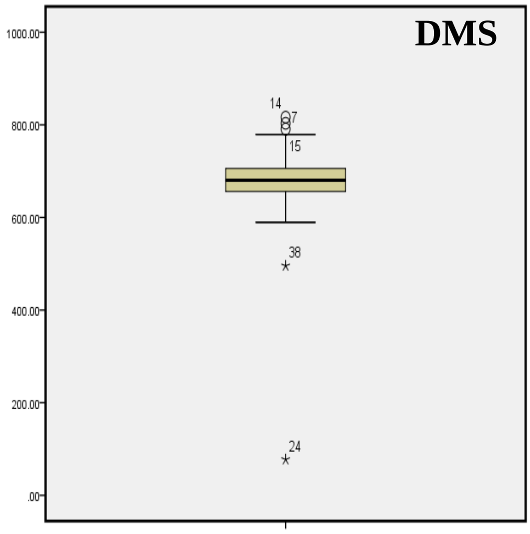


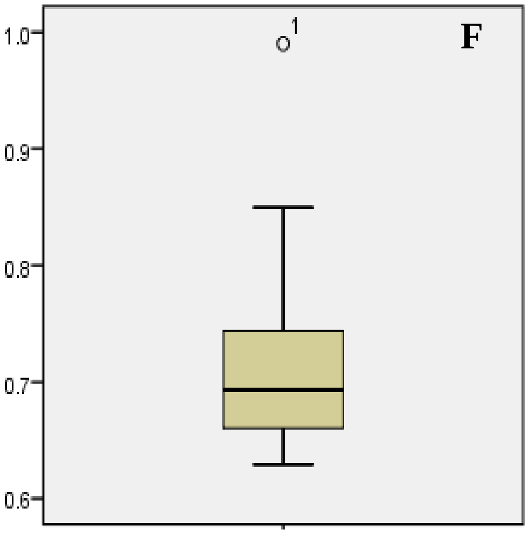


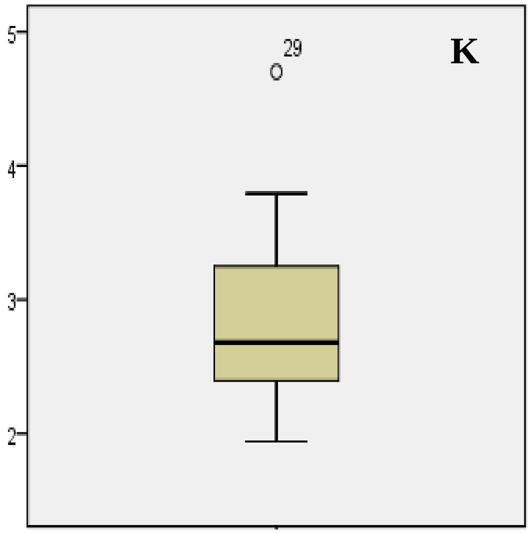
1.3.3. Box & Whisker plots for ZQMWVL2

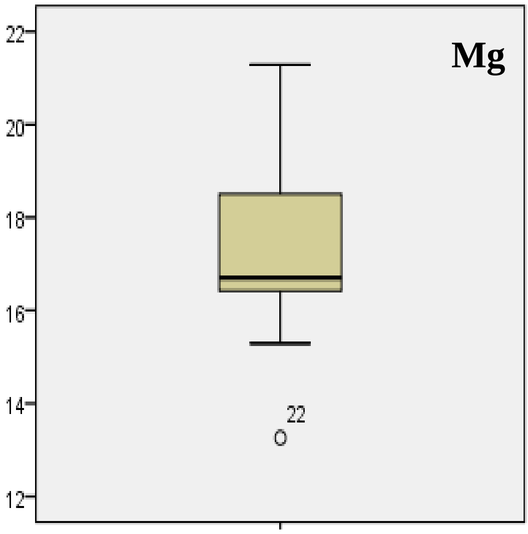


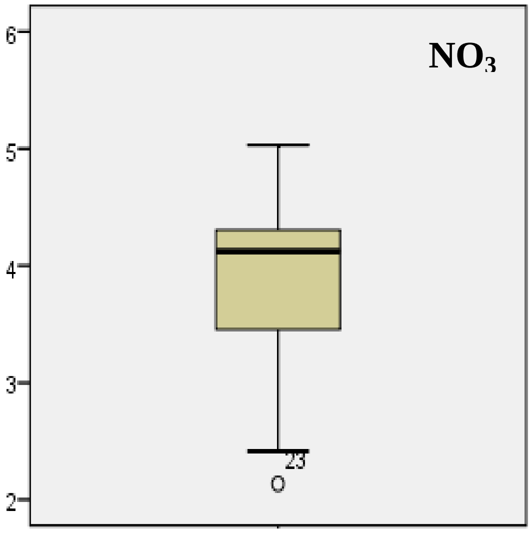


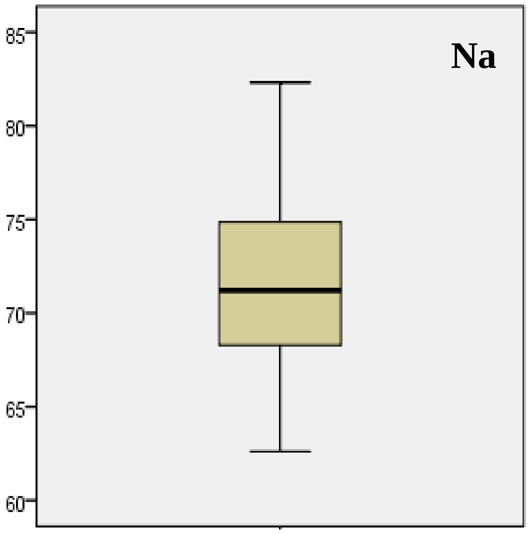


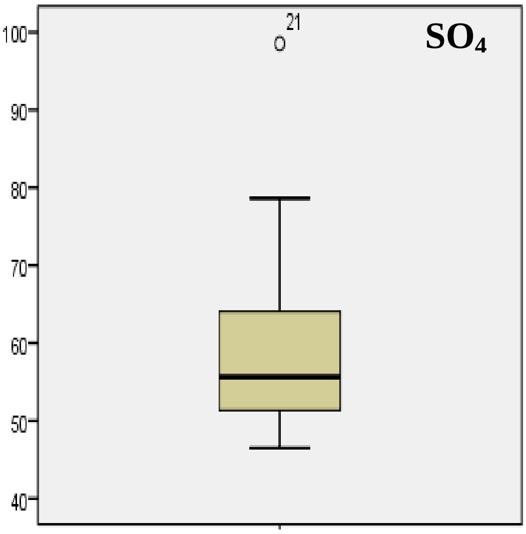


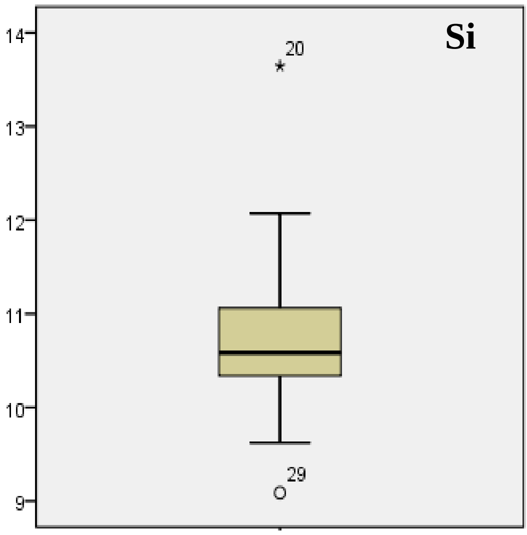

K
5
4
3
2
29

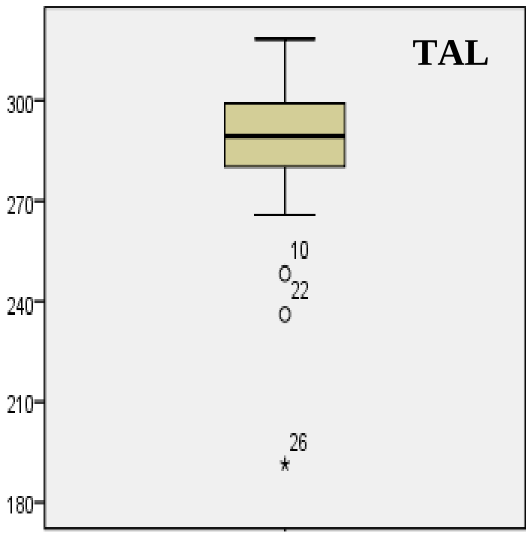

Mg
22
20
18
16
14
12
22

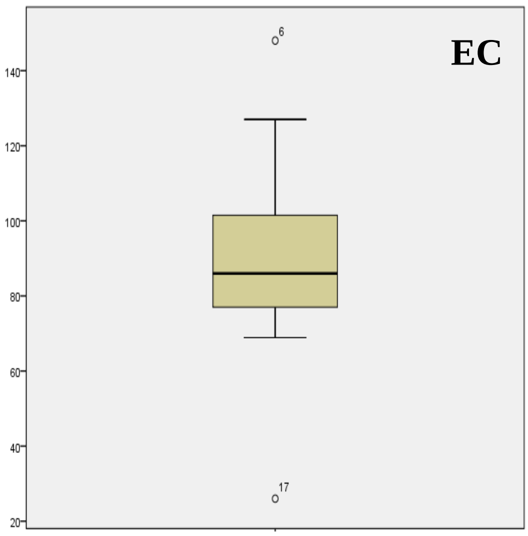

NO₃
6
5
4
3
2
23

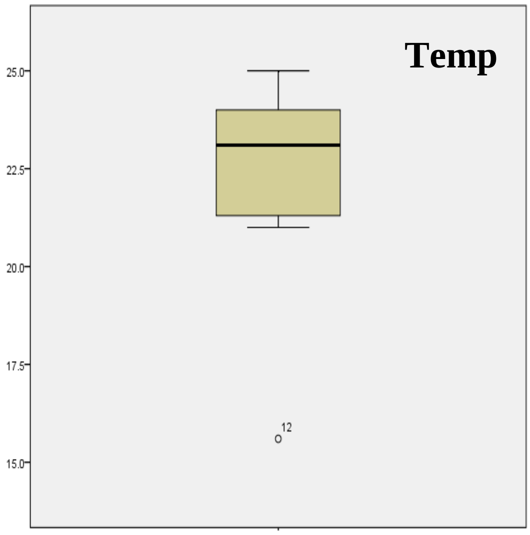

Na
85
80
75
70
65
60

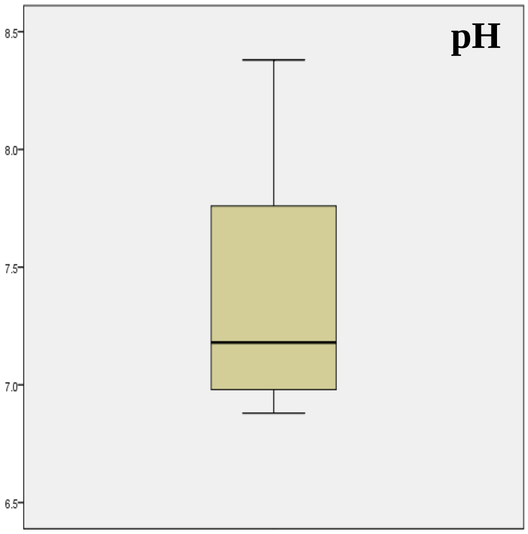

SO₄
100
90
80
70
60
50
40
21


Si
14
13
12
11
10
9
20
29

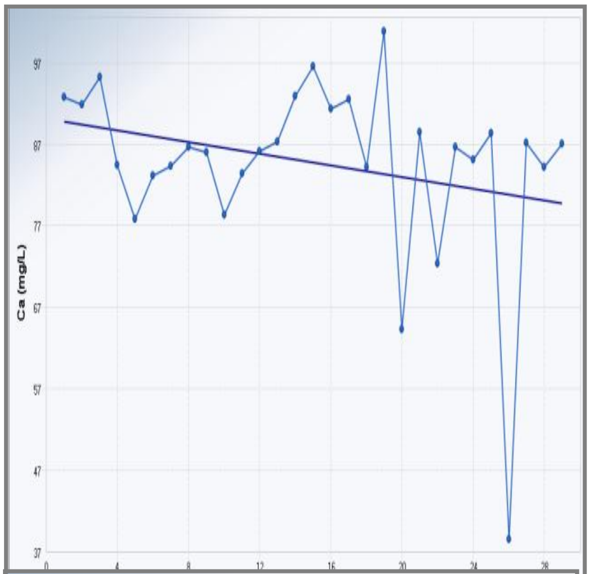

TAL
300
270
240
210
180
10
22
26

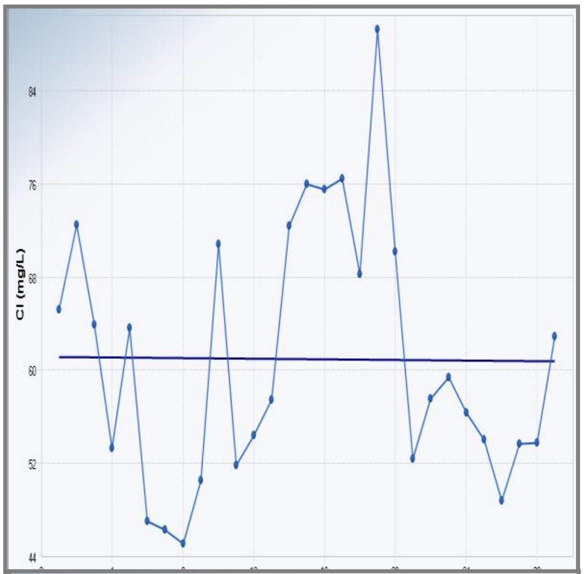

EC
140
120
100
80
60
40
20
6
17

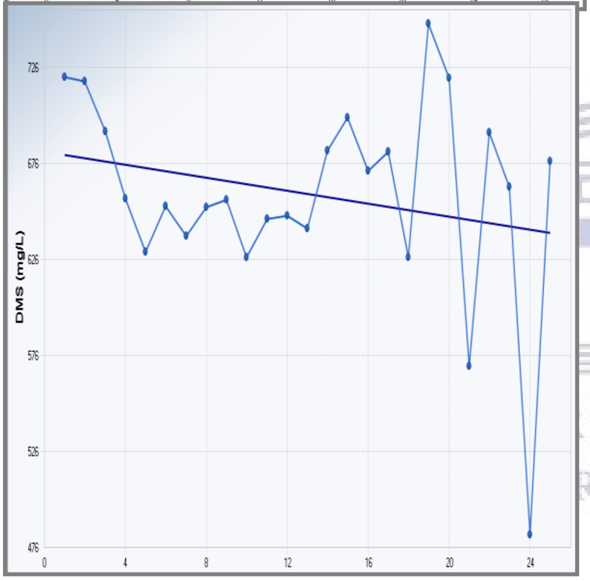

Temp
25.0
22.5
20.0
17.5
15.0
12

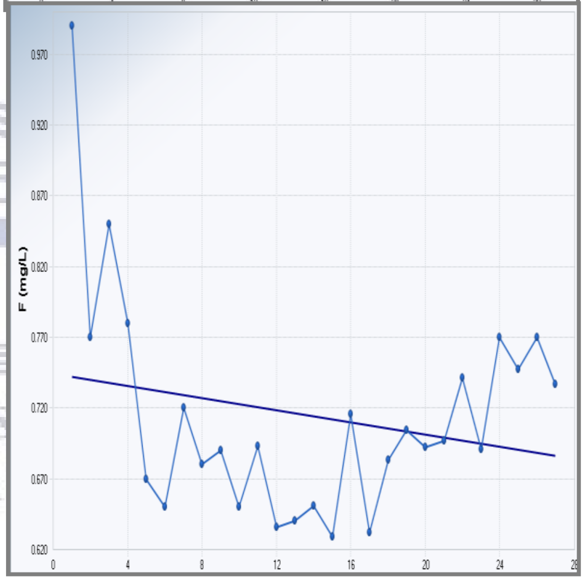

pH
8.5
8.0
7.5
7.0
6.5

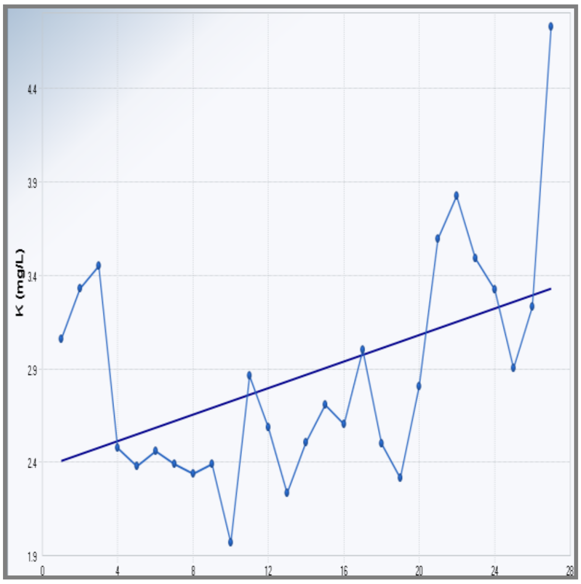
1.3.4. Time-series graphs for ZQMWVL2

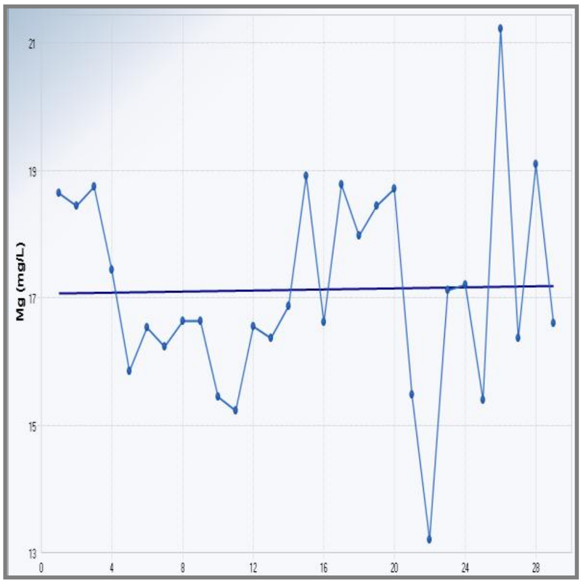


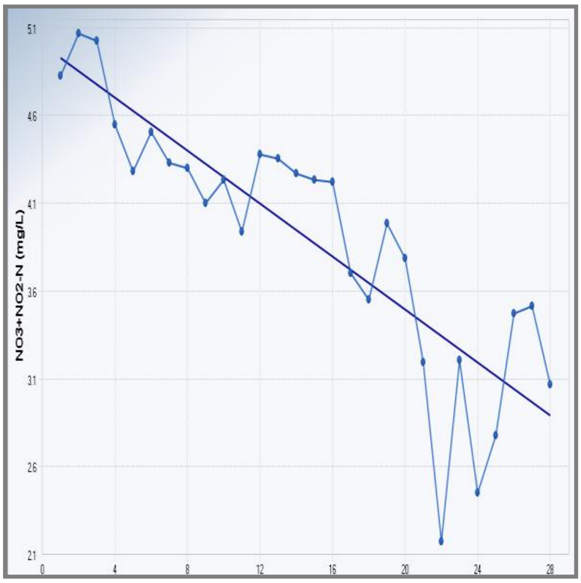


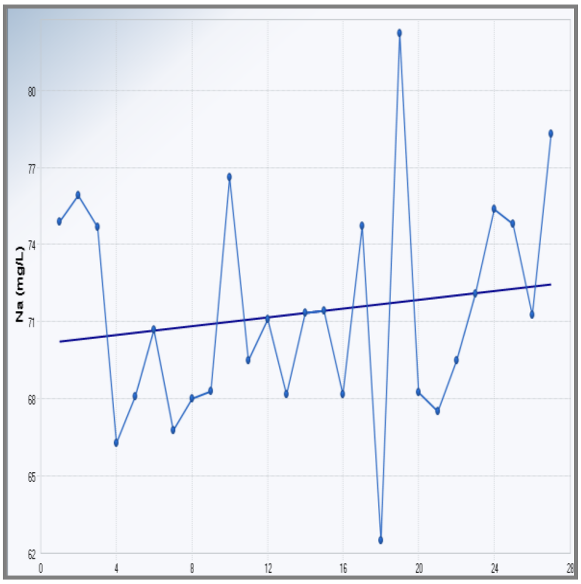


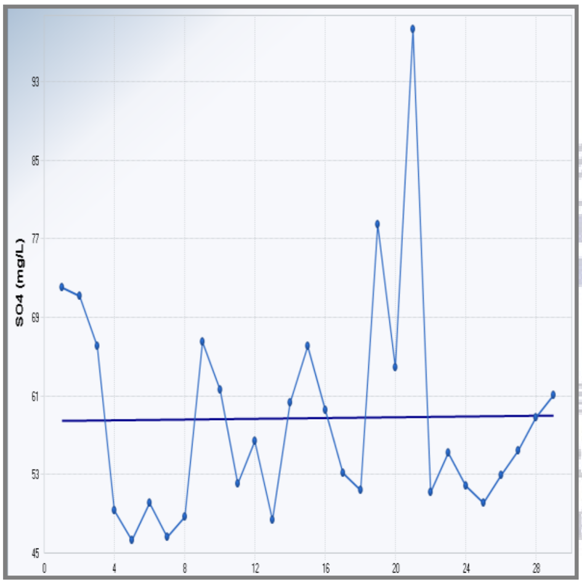


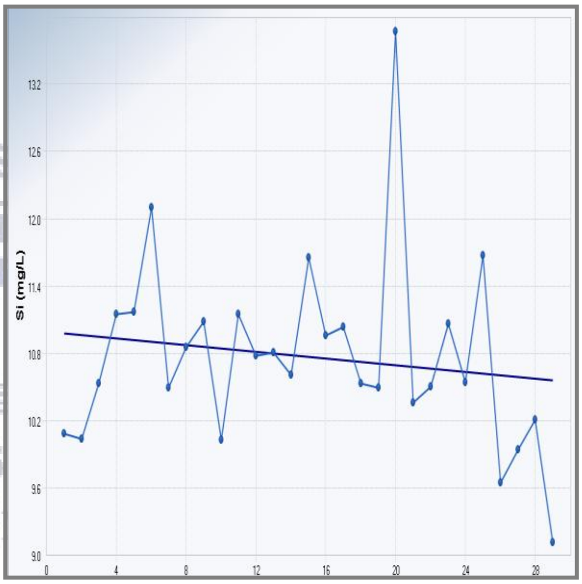


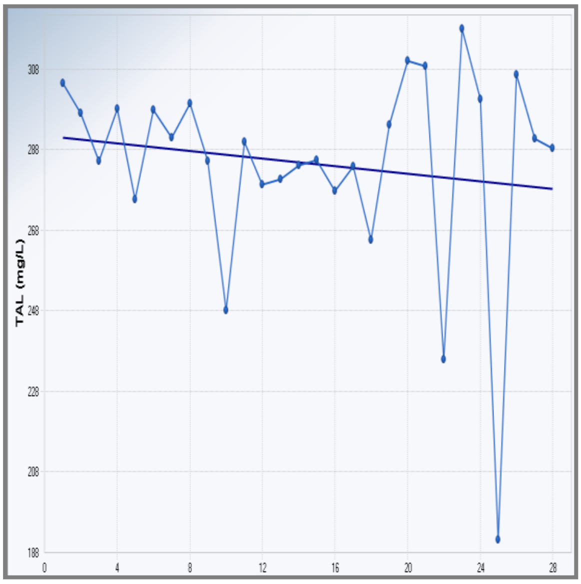


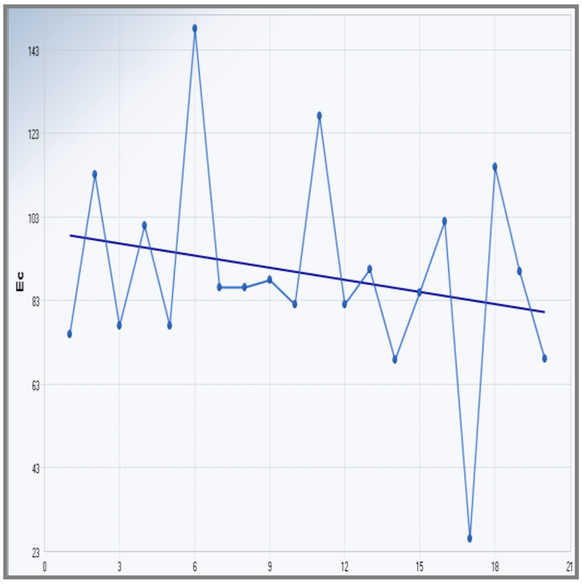

NO3+NO2-N (mg/L)
5.1
4.6
4.1
3.6
3.1
2.6
2.1
0
4
8
12
16
20
24
28

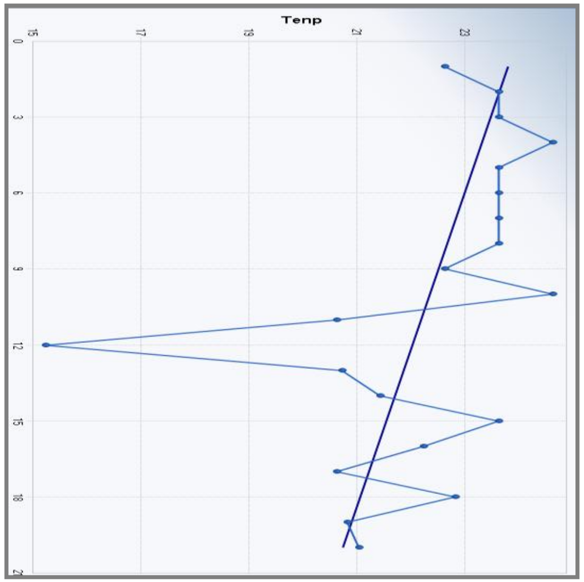

Na (mg/L)
80
77
74
71
68
65
62
0
4
8
12
16
20
24
28

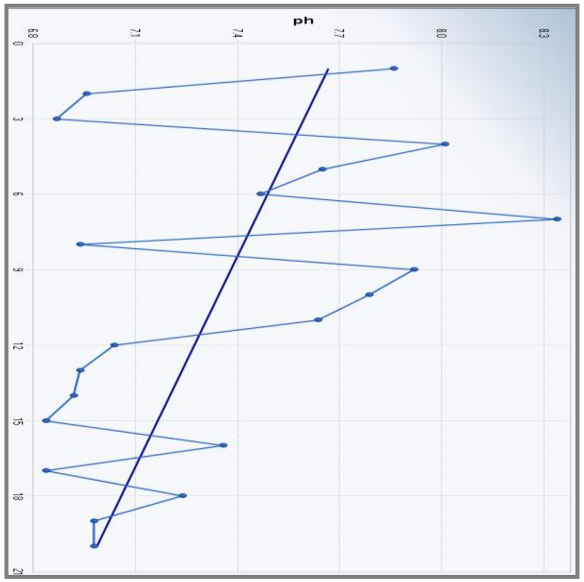

SO4 (mg/L)
93
85
77
69
61
53
45
0
4
8
12
16
20
24
28


Si (mg/L)
13.2
12.6
12.0
11.4
10.8
10.2
9.6
9.0
0
4
8
12
16
20
24
28


TAL (mg/L)
308
288
268
248
228
208
188
0
4
8
12
16
20
24
28


Ec
143
123
103
83
63
43
23
0
3
6
9
12
15
18
21



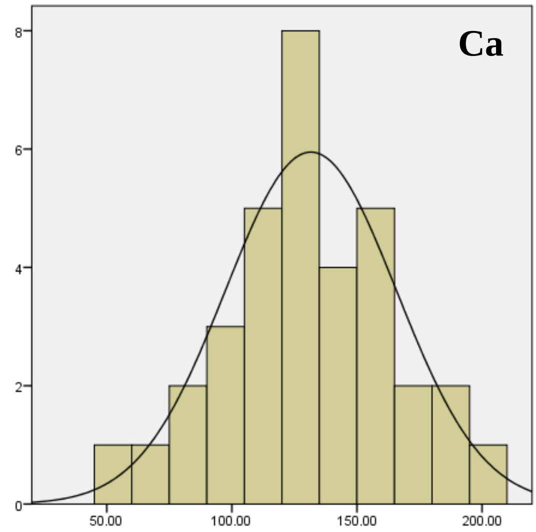


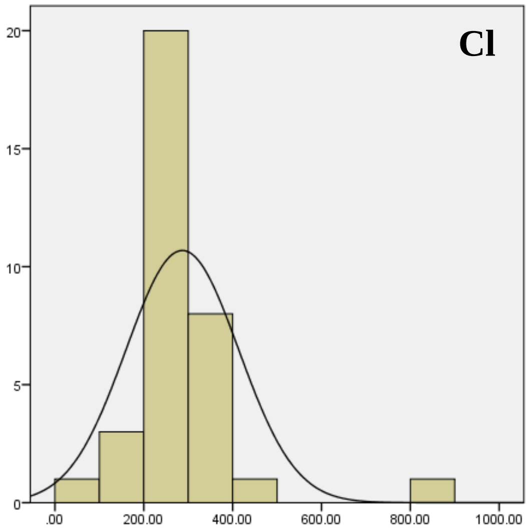
1.4.1. Exploratory statistical analysis for ZQMLUE1

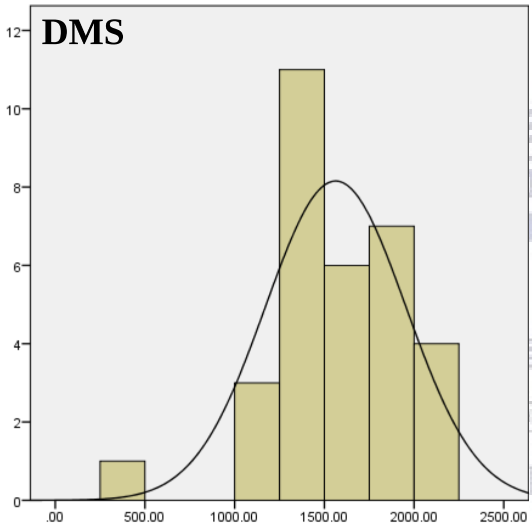
Table 1.3. Descriptive statistics for ZQMLUE1

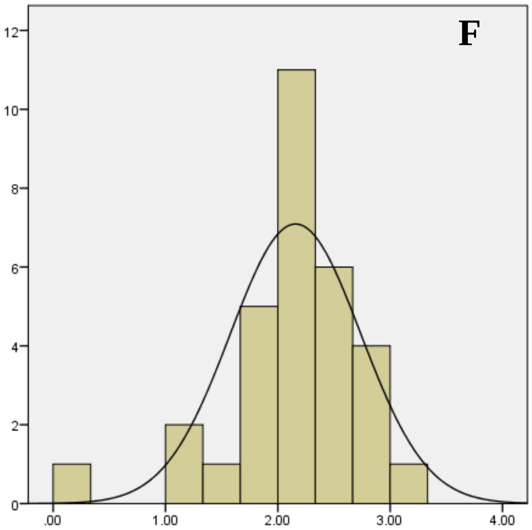
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Ca	34	159.560	49.740	209.300	131.587	34.197	1169.418	0.033	0.229
Cl	34	823.980	24.730	848.720	286.715	126.898	16103.134	2.453	11.644
DMS	32	1934.850	250.240	2185.100	1562.812	391.313	153126.022	-1.042	2.774
F	31	2.960	0.120	3.080	2.158	0.581	0.338	-1.541	4.071
K	33	12.680	0.480	13.160	7.607	3.291	10.828	-0.151	-0.913
Mg	34	41.120	5.380	46.500	18.203	10.459	109.395	1.233	1.309
NH4	32	0.190	0.020	0.200	0.040	0.040	0.002	2.693	8.007
NO3NO2	33	4.080	0.010	4.090	0.388	0.931	0.867	2.868	8.166
Na	33	503.120	11.810	514.930	334.382	92.585	8571.982	-1.127	3.835
PO4	32	0.180	0.010	0.180	0.026	0.032	0.001	3.939	17.805
SO4	34	995.840	4.490	1000.330	450.919	187.127	35016.680	-0.228	1.985
Si	33	9.210	5.730	14.940	9.285	1.757	3.088	1.642	4.562
TAL	33	653.310	110.990	764.300	295.349	130.162	16942.227	1.243	3.813
EC	24	266.200	163.800	430.000	272.908	70.231	4932.323	1.069	0.449
Temp	24	11.600	14.400	26.000	21.863	3.200	10.242	-0.986	0.005
pH	22	3.310	5.100	8.410	7.233	0.692	0.480	-1.005	3.328

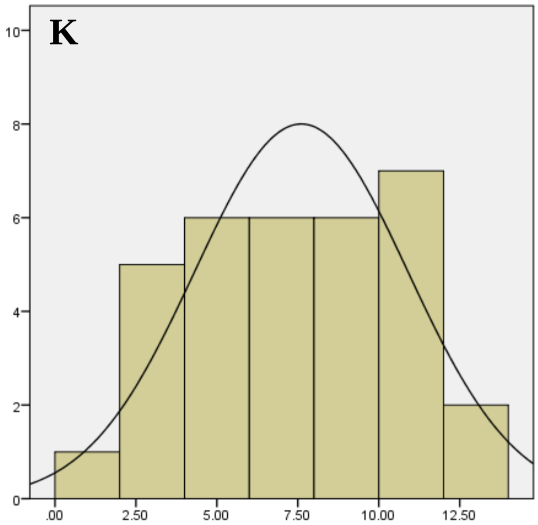
1.4.2. Histograms for ZQMLUE1

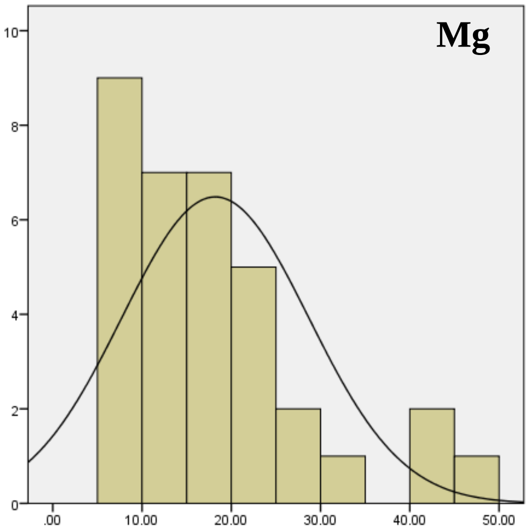


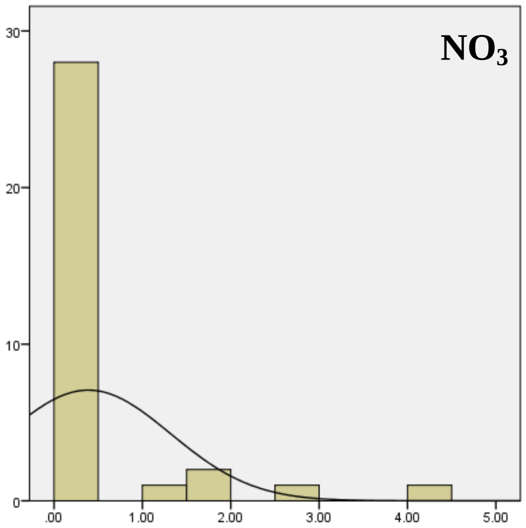


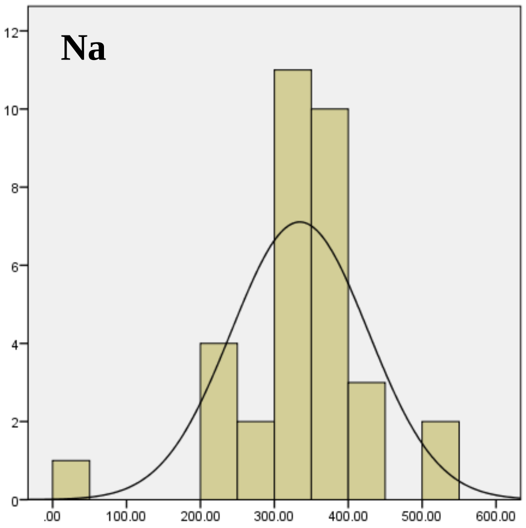


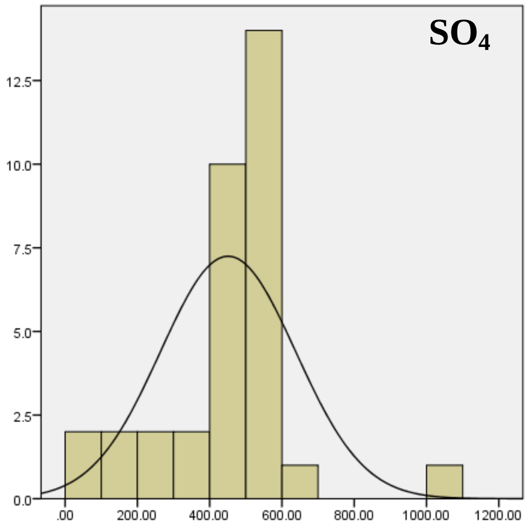


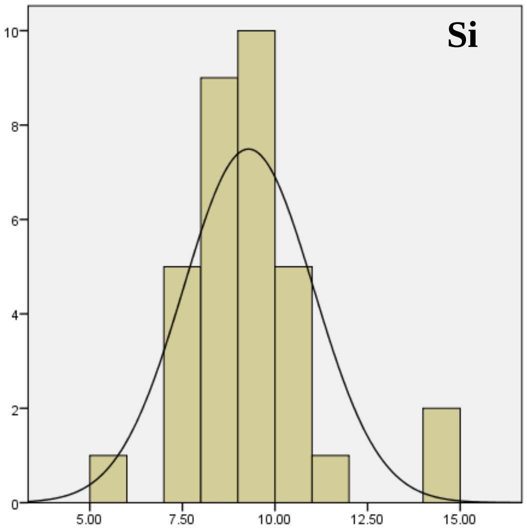


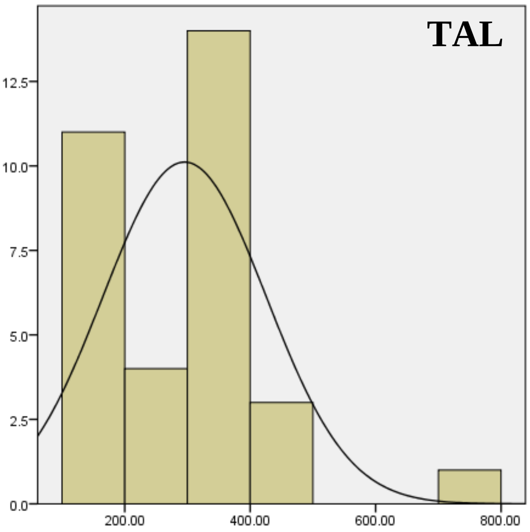


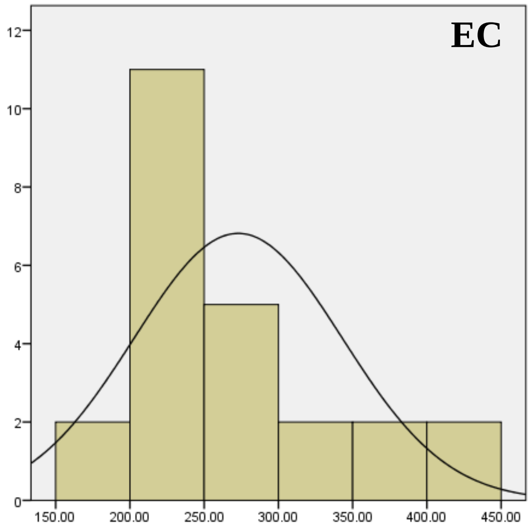

NO$_3$
30
20
10
0
.00
1.00
2.00
3.00
4.00
5.00

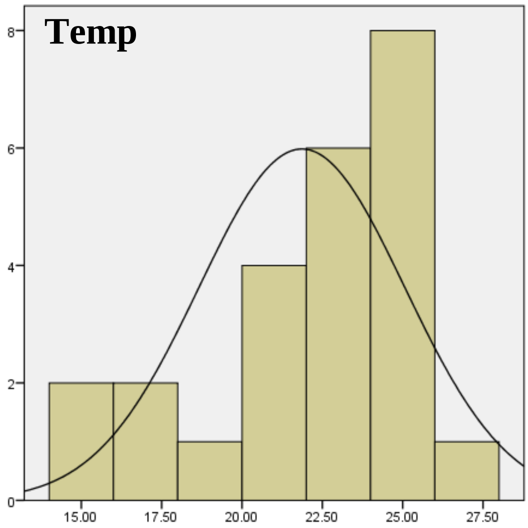

Na
12
10
8
6
4
2
0
.00
100.00
200.00
300.00
400.00
500.00
600.00

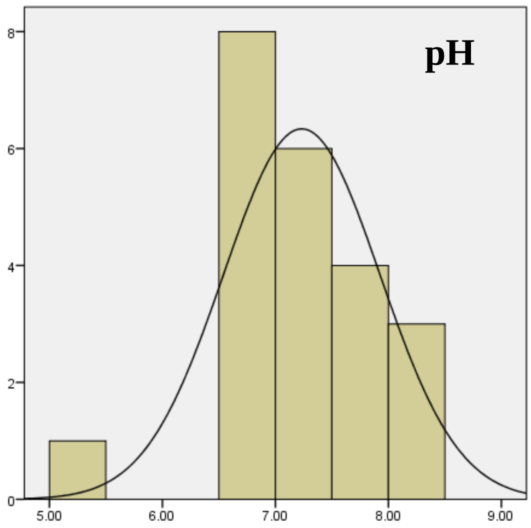

SO$_4$
12.5
10.0
7.5
5.0
2.5
0.0
.00
200.00
400.00
600.00
800.00
1000.00
1200.00


Si
10
8
6
4
2
0
5.00
7.50
10.00
12.50
15.00

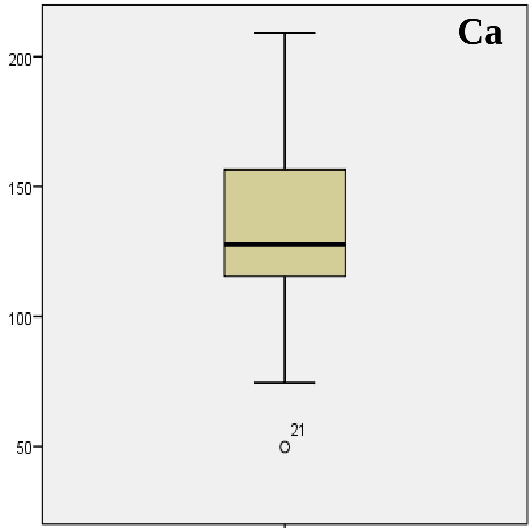

TAL
12.5
10.0
7.5
5.0
2.5
0.0
200.00
400.00
600.00
800.00

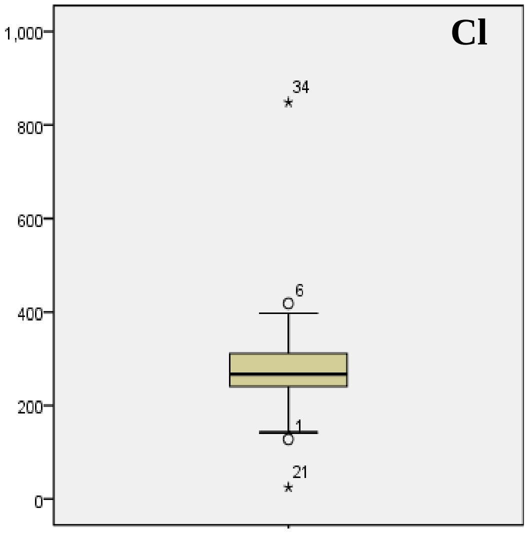

EC
12
10
8
6
4
2
0
150.00
200.00
250.00
300.00
350.00
400.00
450.00

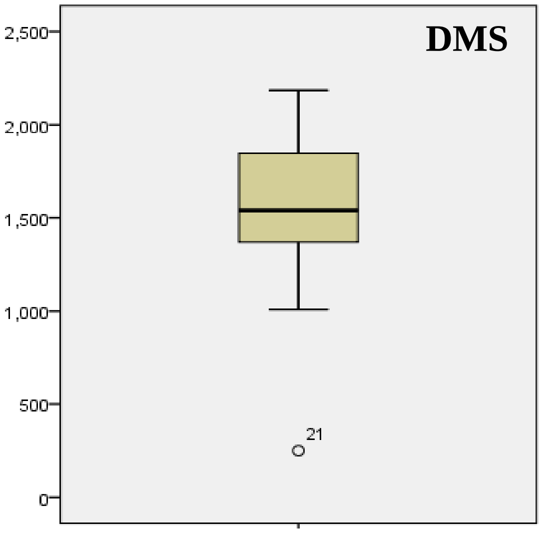


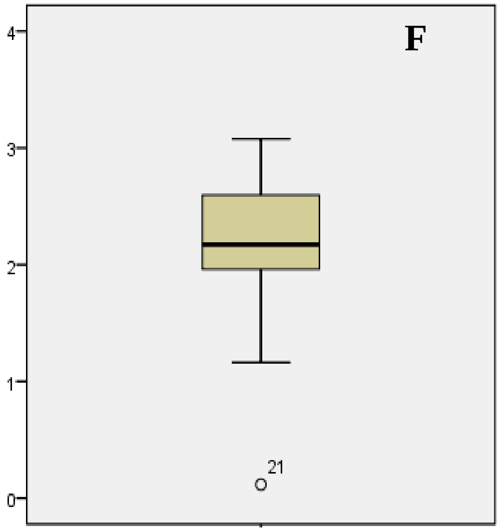


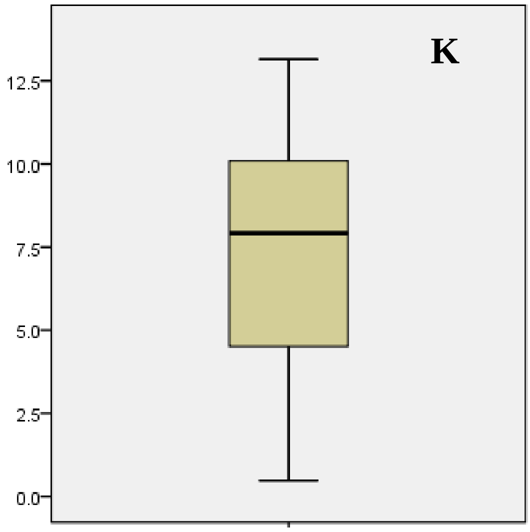
1.4.3. Box and Whisker plots for ZQMLUE1

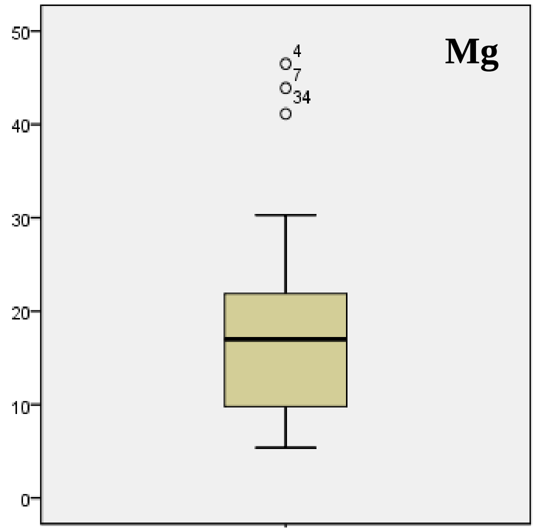


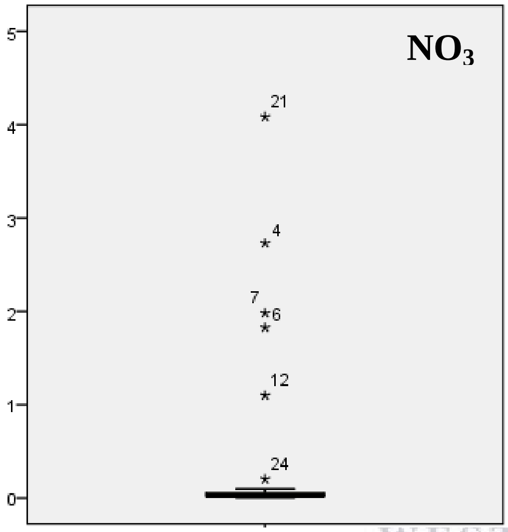


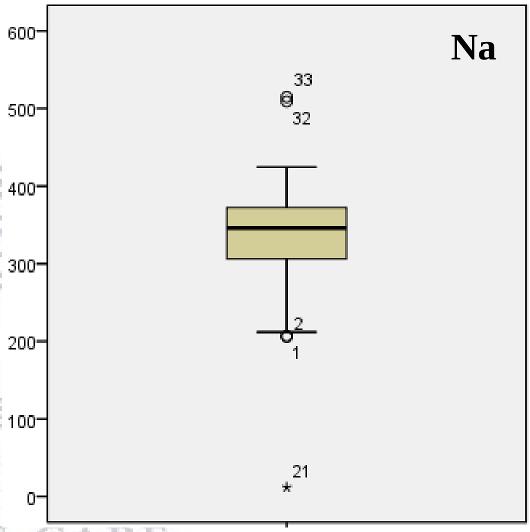


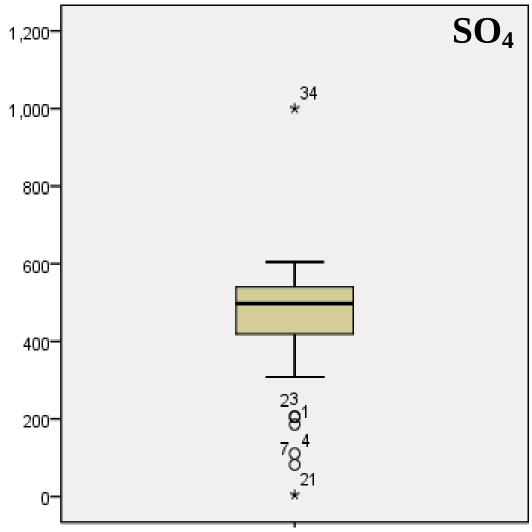


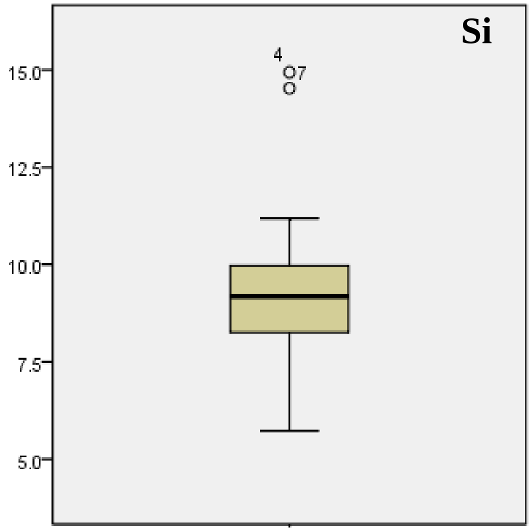

K
12.5
10.0
7.5
5.0
2.5
0.0

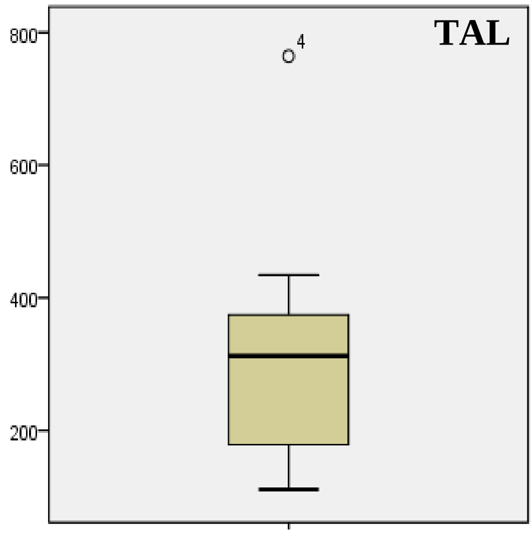

Mg
50
40
30
20
10
0
4
7
34

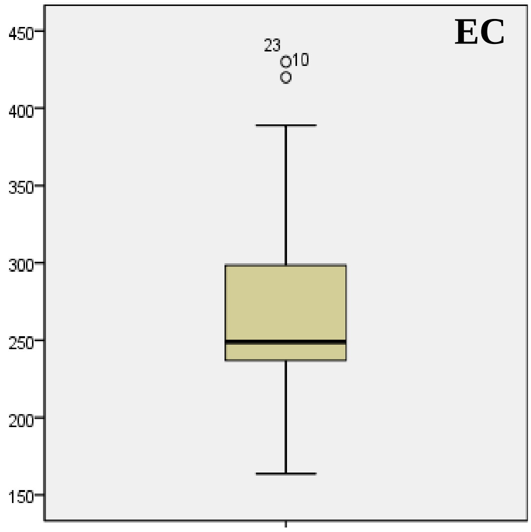

NO3
5
4
3
2
1
0
21
4
7
6
12
24

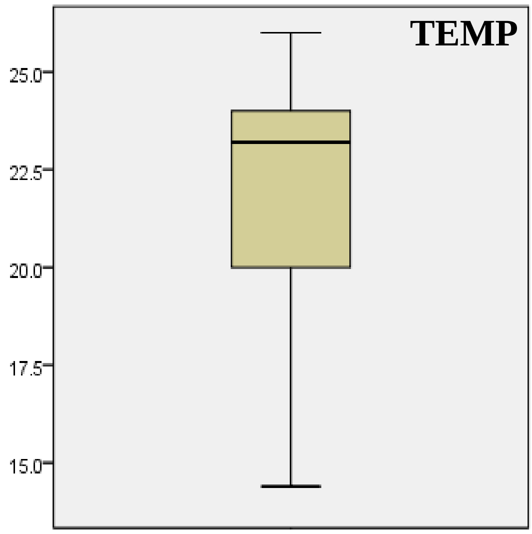

Na
600
500
400
300
200
100
0
33
32
2
1
21

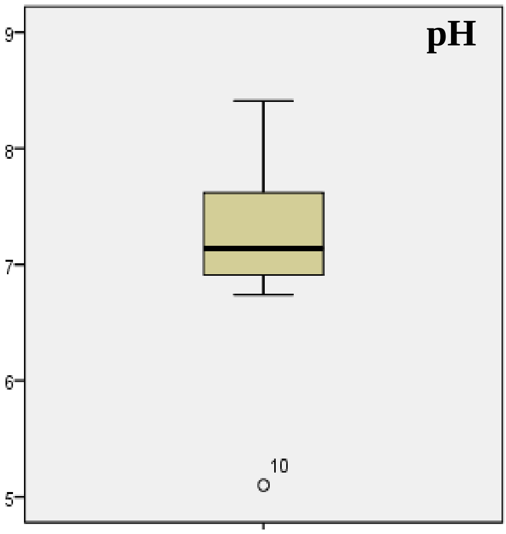

SO4
1,200
1,000
800
600
400
200
0
34
23
1
7
4
21


Si
15.0
12.5
10.0
7.5
5.0
4
7

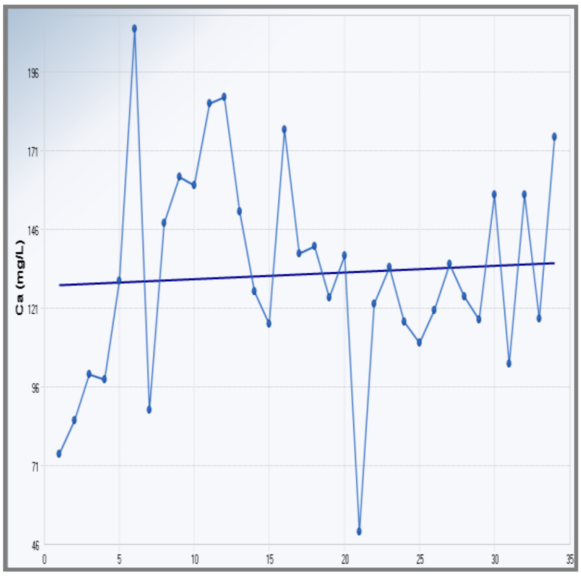

TAL
800
600
400
200
4

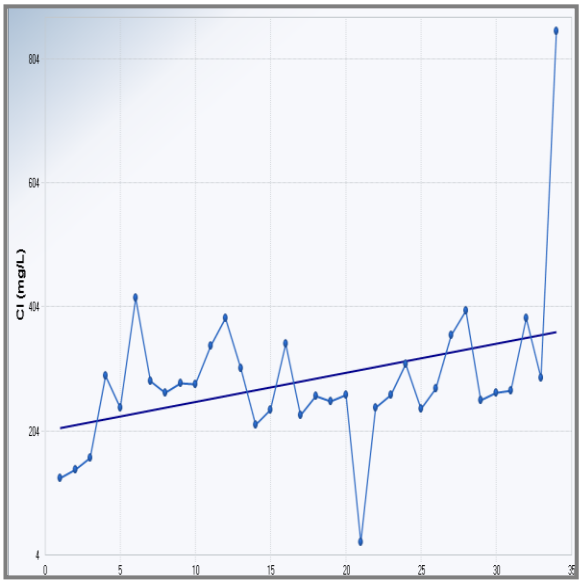

EC
450
400
350
300
250
200
150
23
10

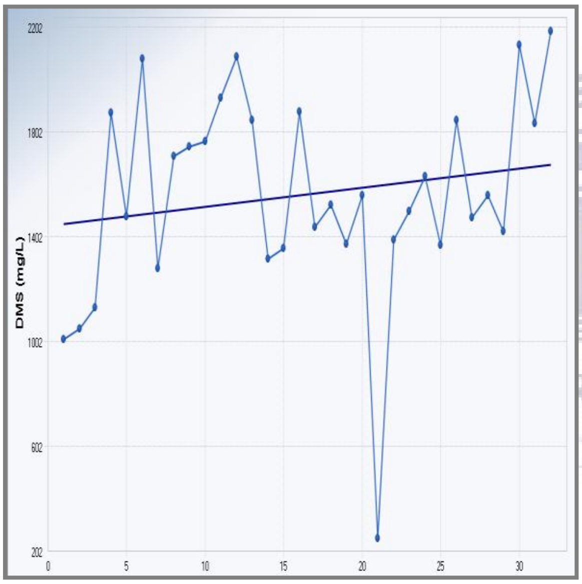

TEMP
25.0
22.5
20.0
17.5
15.0

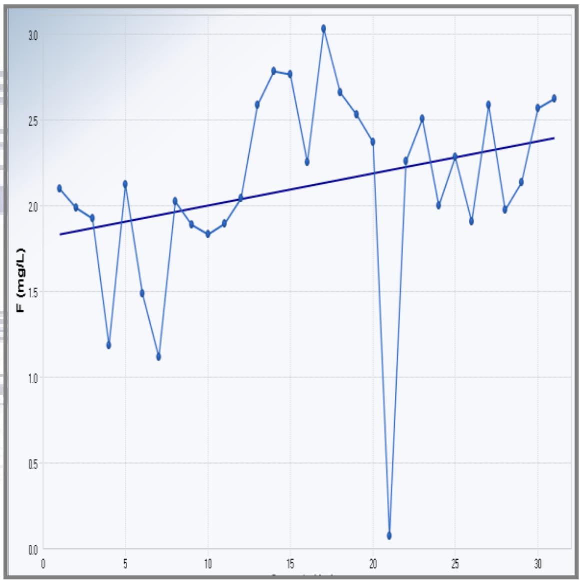

pH
9
8
7
6
5
10

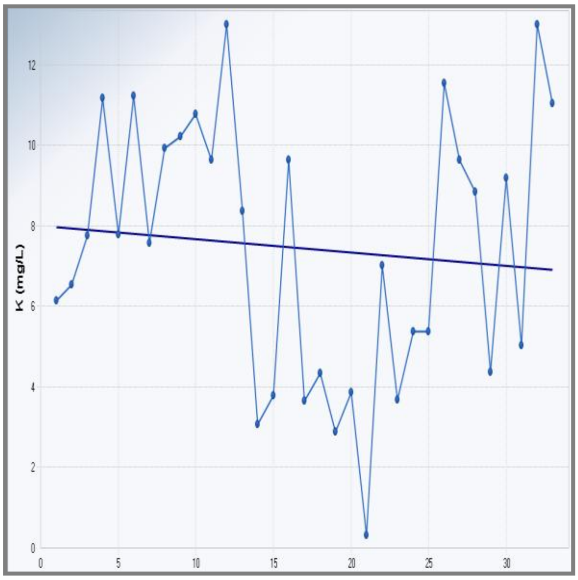
1.4.4. Time-series graphs of ZQMLUE1

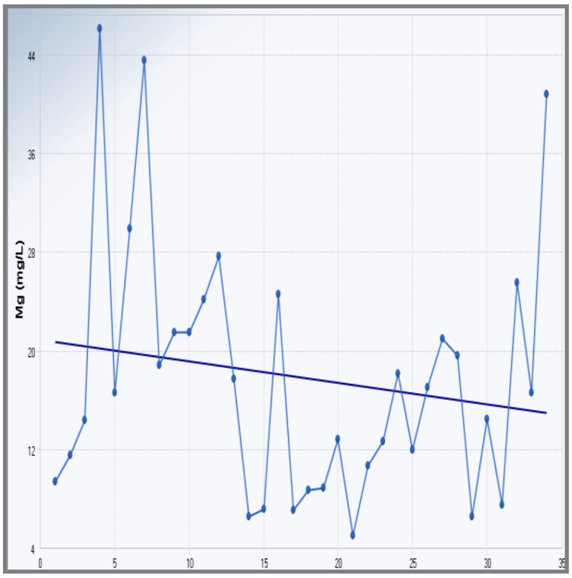


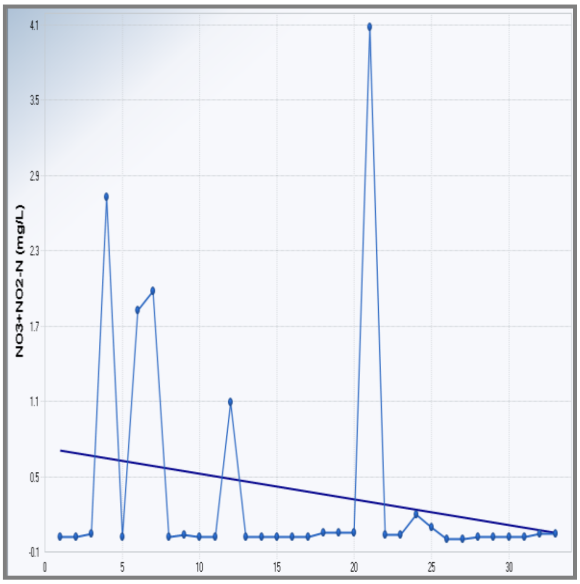


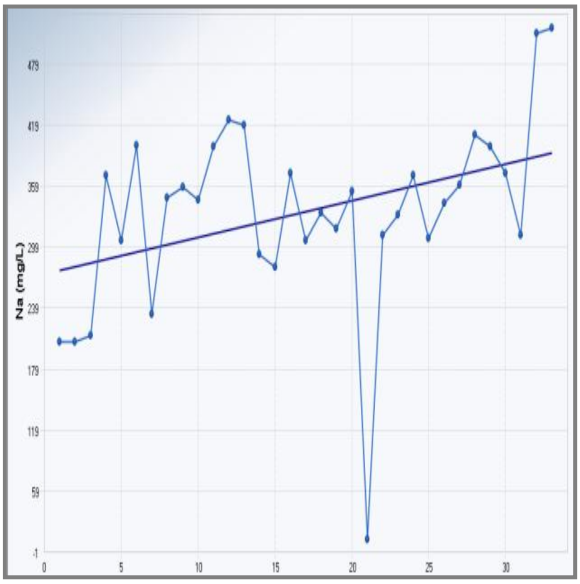


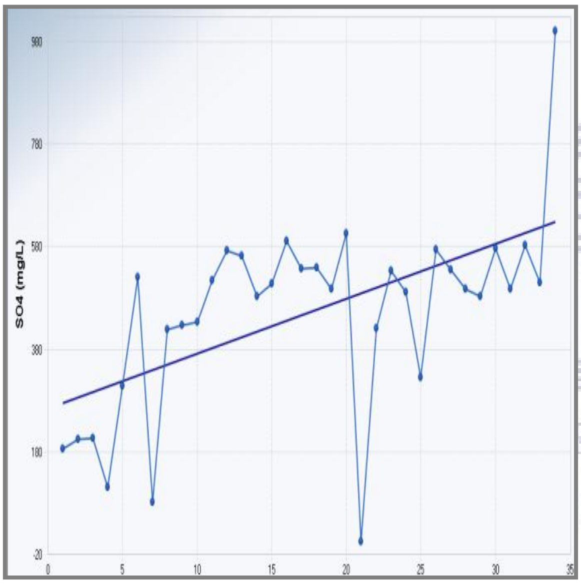


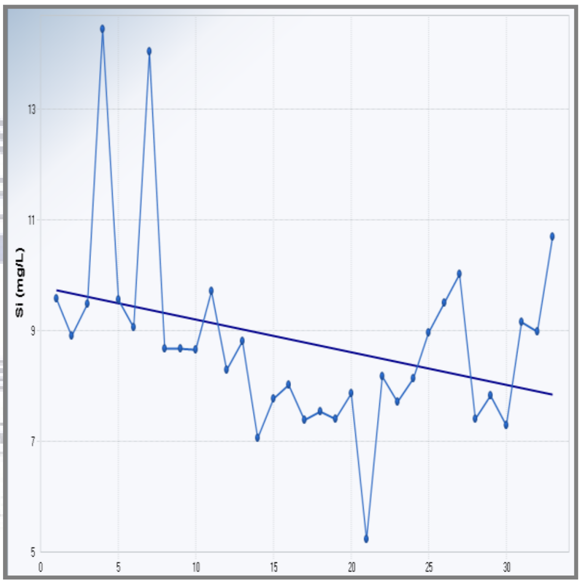


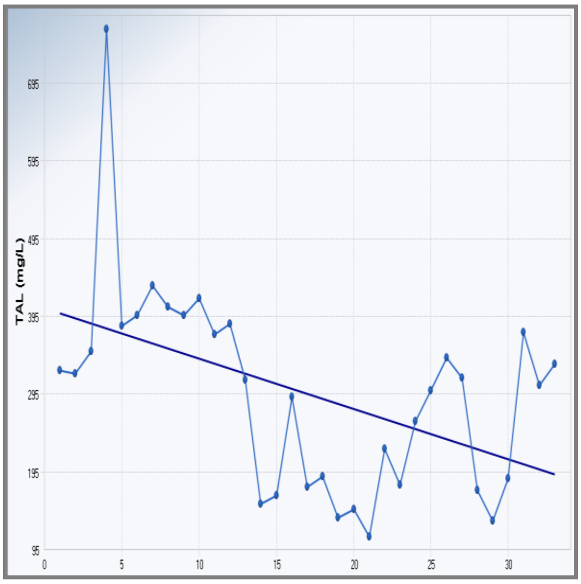


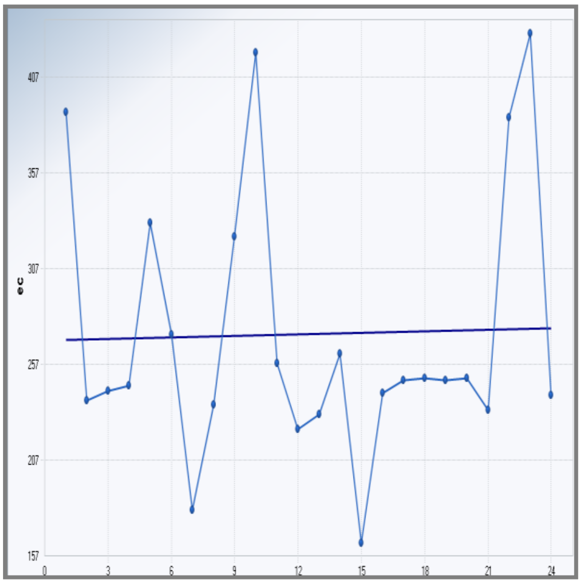

NO3+NO2-N (mg/L)
4.1
3.5
2.9
2.3
1.7
1.1
0.5
-0.1
0
5
10
15
20
25
30

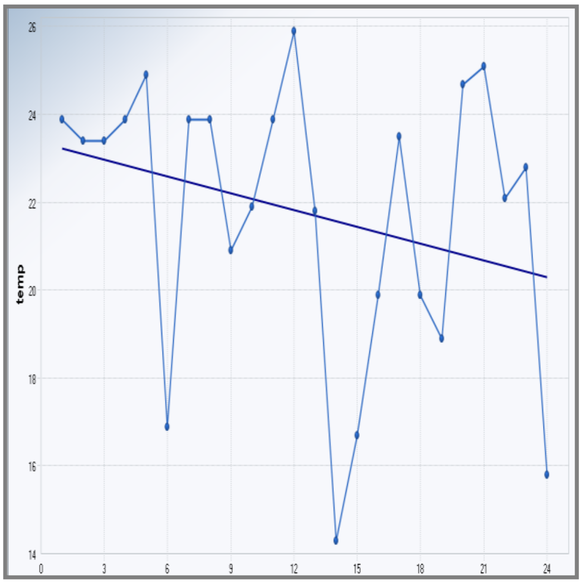

Na (mg/L)
479
419
359
299
239
179
119
59
-1
0
5
10
15
20
25
30

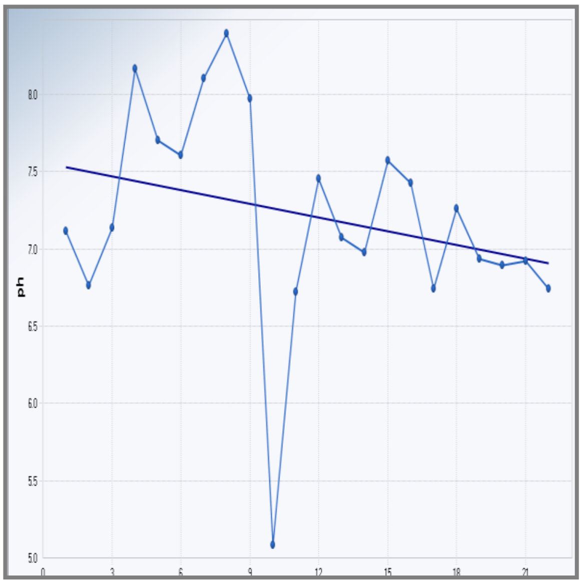

SO4 (mg/L)
980
780
580
380
180
-20
0
5
10
15
20
25
30
35


Si (mg/L)
13
11
9
7
5
0
5
10
15
20
25
30


TAL (mg/L)
695
595
495
395
295
195
95
0
5
10
15
20
25
30


ec
407
357
307
257
207
157
0
3
6
9
12
15
18
21
24



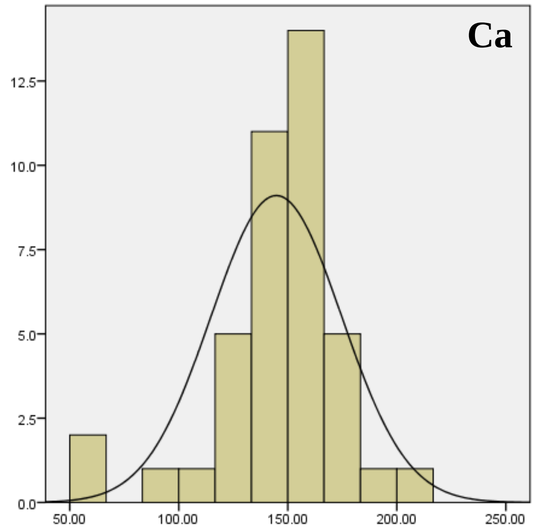


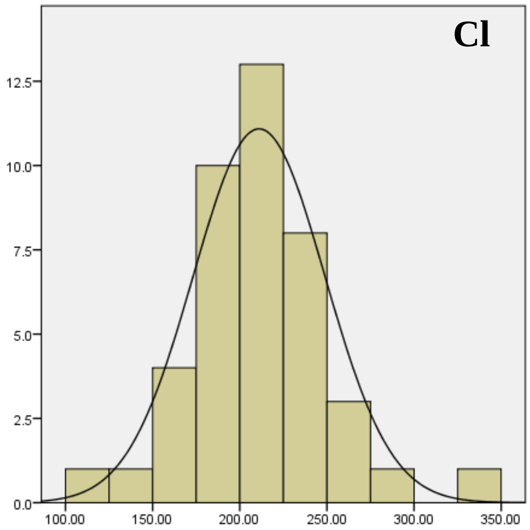
1.5.1. Exploratory statistical analysis for ZQMRSK1

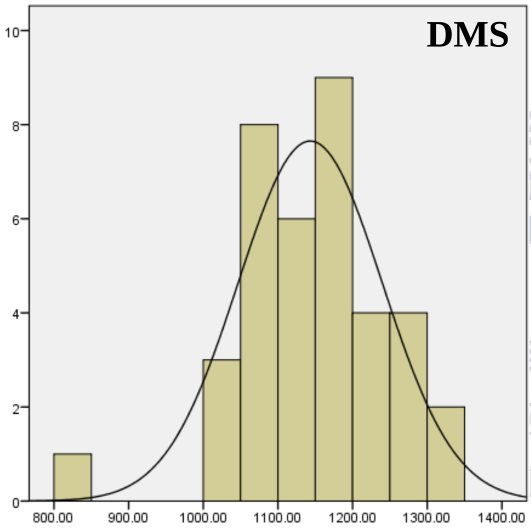
Table 1.4. Descriptive statistics for ZQMRSK1

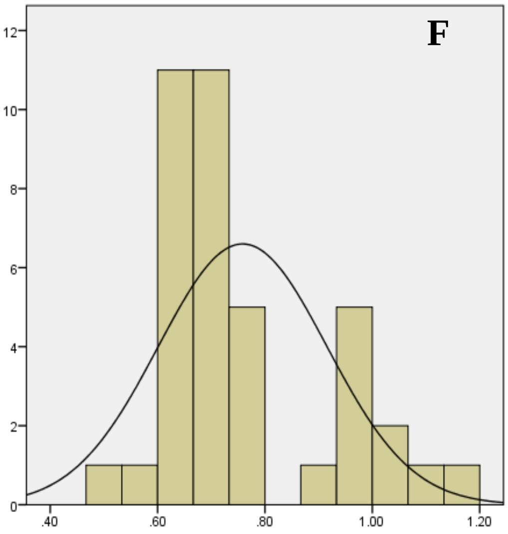
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Ca	41	150.02	51.57	201.59	144.74	29.94	896.58	-1.40	3.15
Cl	42	209.49	116.39	325.88	211.05	37.77	1426.95	0.32	1.79
DMS	37	459.71	849.29	1309.00	1143.25	96.44	9301.61	-0.54	0.98
F	39	0.65	0.49	1.14	0.76	0.16	0.03	0.99	0.05
K	40	4.45	2.44	6.89	3.35	0.99	0.99	2.12	5.13
Mg	42	44.80	14.96	59.76	38.36	7.23	52.33	-0.40	3.81
NO3NO2	41	5.75	0.03	5.78	2.09	1.70	2.88	0.60	-0.76
Na	40	226.39	40.46	266.84	148.59	31.13	969.33	0.38	7.70
SO4	42	740.26	3.00	743.26	285.66	101.29	10259.12	1.73	10.50
Si	42	4.51	8.28	12.79	10.76	0.97	0.94	0.02	0.49
TAL	41	200.28	160.58	360.86	253.30	39.40	1552.72	0.23	1.73
Hardness	15	426.00	284.00	709.00	481.06	125.79	15822.06	0.22	-0.10
EC	42	93.10	121.90	215.00	156.55	19.67	386.79	0.67	1.09
Temp	42	5.40	18.50	23.90	20.91	0.94	0.88	0.50	2.26
pH	37	1.02	6.70	7.72	7.15	0.17	0.03	0.34	3.06

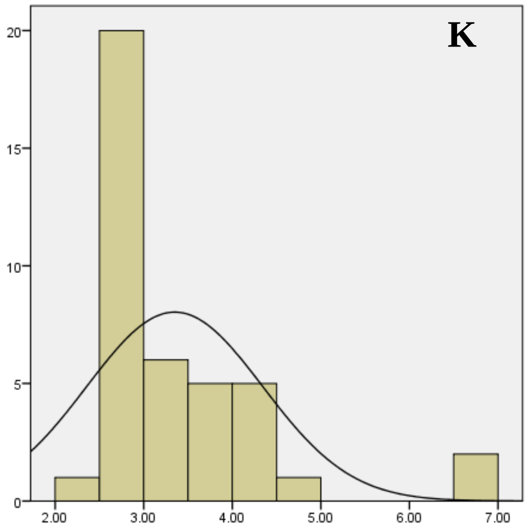
1.5.2. Histograms for ZQMRSK1

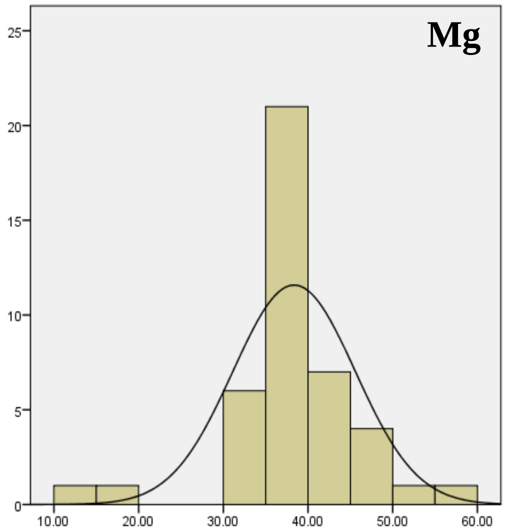


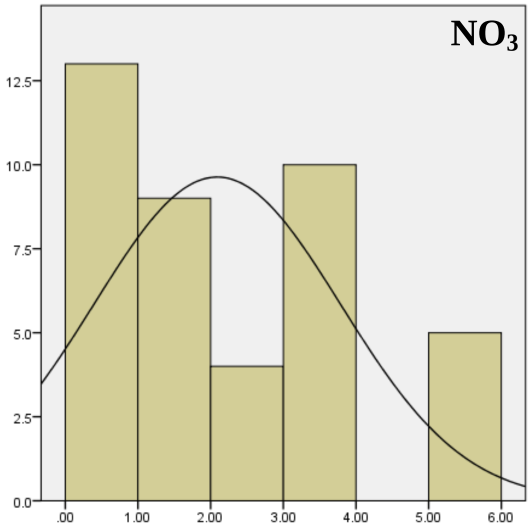


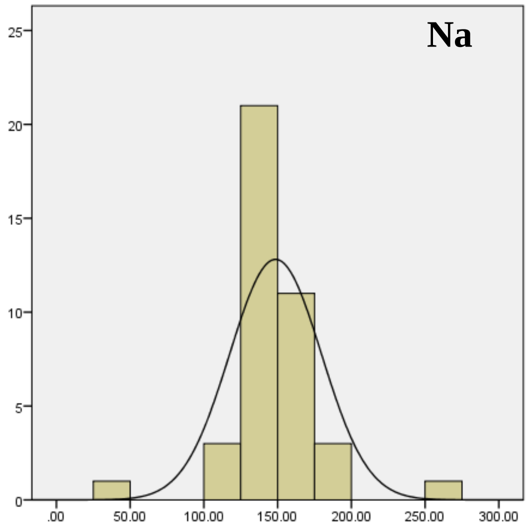


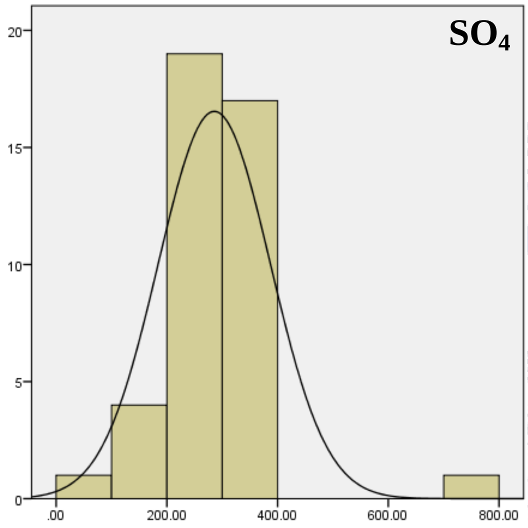


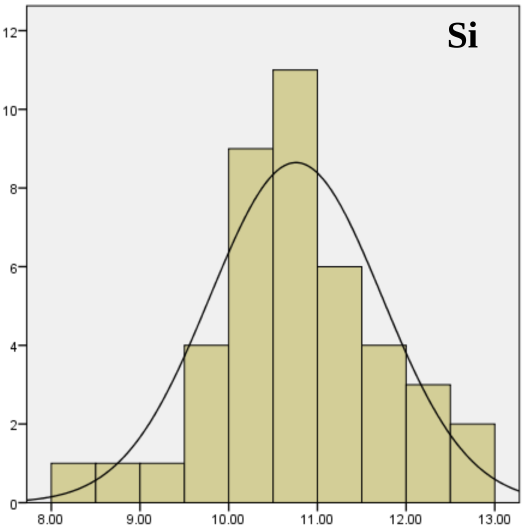


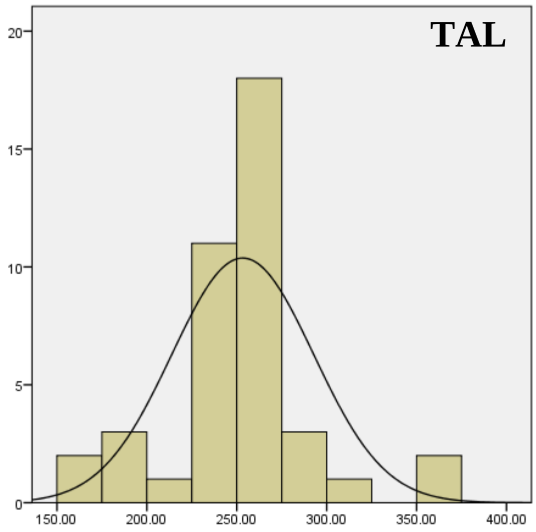


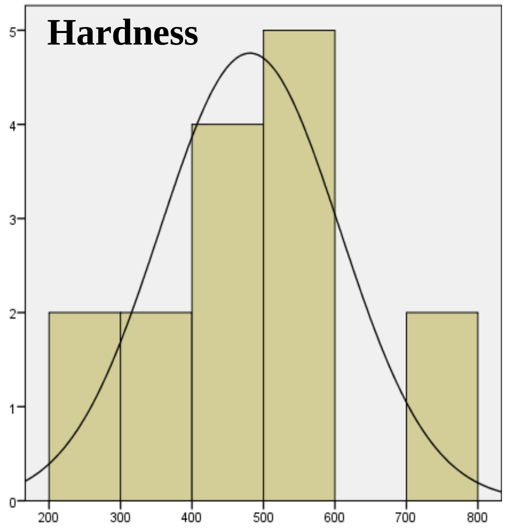

NO$_3$
12.5
10.0
7.5
5.0
2.5
0.0
.00
1.00
2.00
3.00
4.00
5.00
6.00

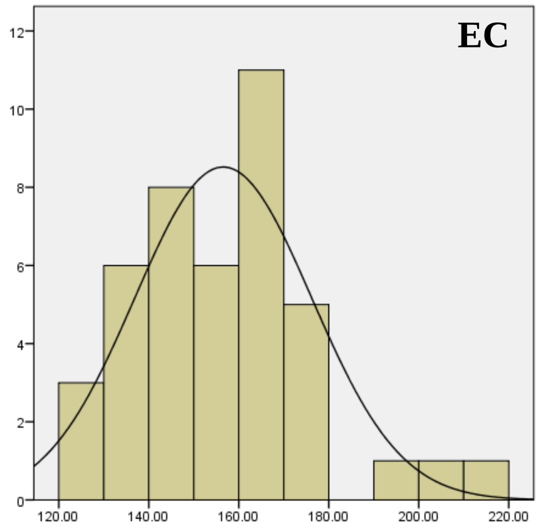

Na
25
20
15
10
5
0
.00
50.00
100.00
150.00
200.00
250.00
300.00

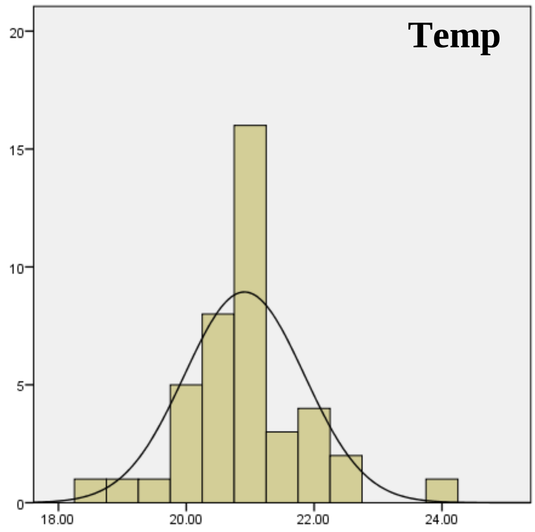

SO$_4$
20
15
10
5
0
.00
200.00
400.00
600.00
800.00

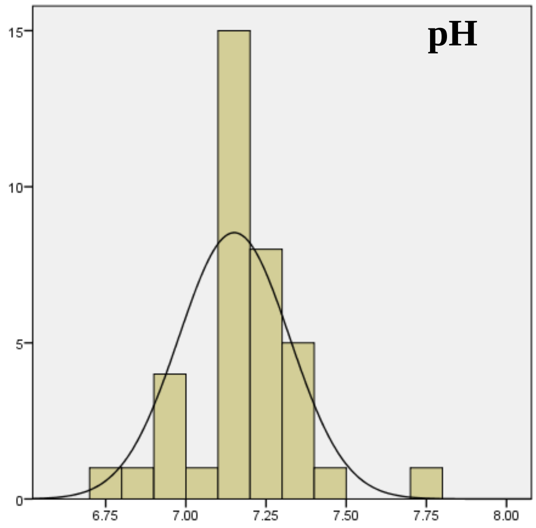

Si
12
10
8
6
4
2
0
8.00
9.00
10.00
11.00
12.00
13.00


TAL
20
15
10
5
0
150.00
200.00
250.00
300.00
350.00
400.00

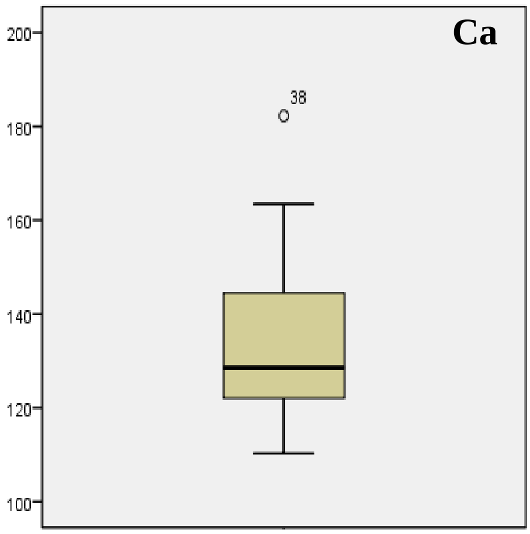

Hardness
5
4
3
2
1
0
200
300
400
500
600
700
800

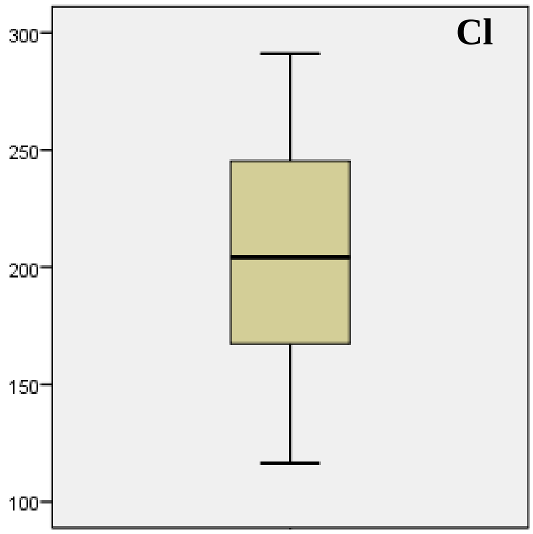

EC
0
2
4
6
8
10
12
120.00
140.00
160.00
180.00
200.00
220.00

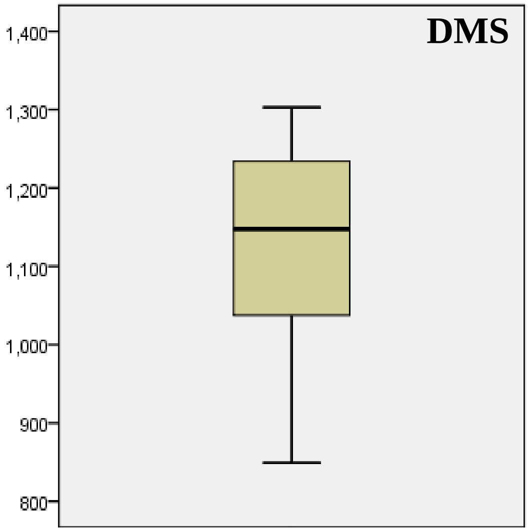

Temp
0
5
10
15
20
18.00
20.00
22.00
24.00

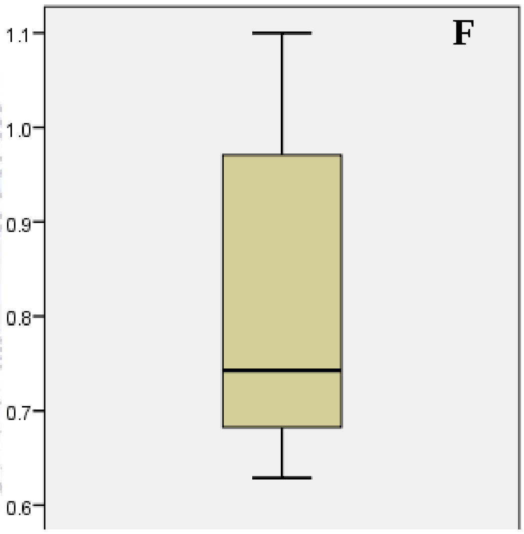

pH
0
5
10
15
6.75
7.00
7.25
7.50
7.75
8.00

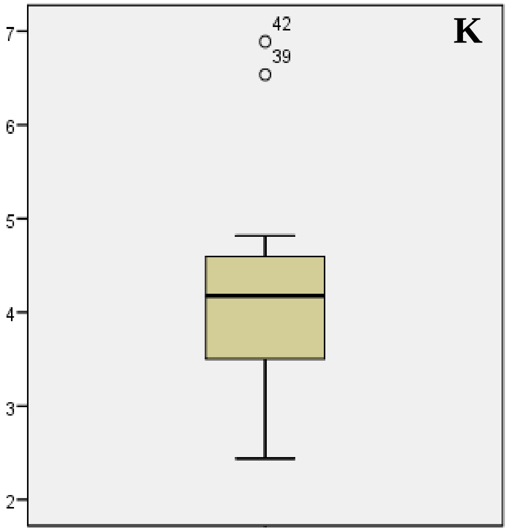
1.5.3. Box & Whisker plots for ZQMRSK1

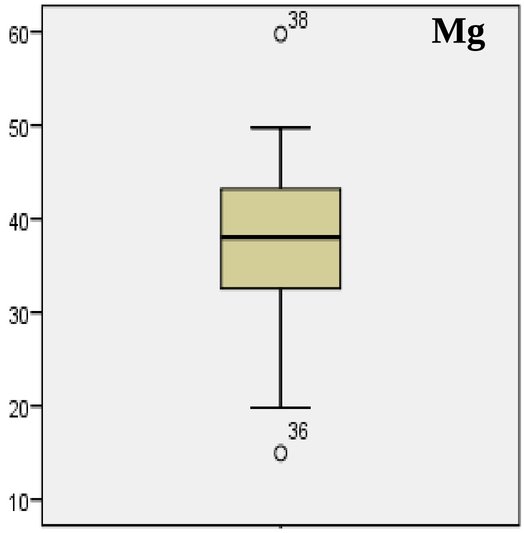


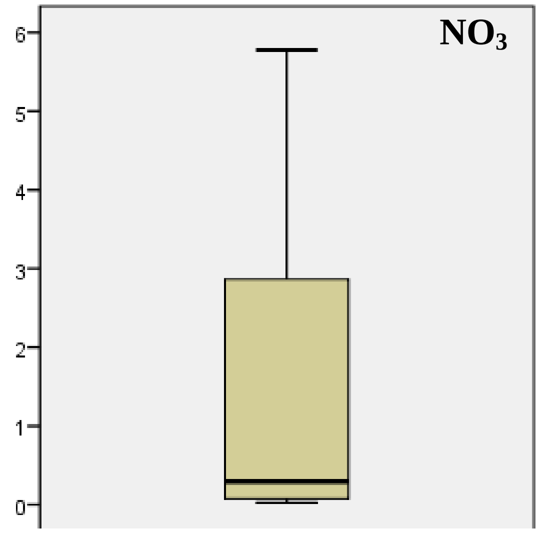


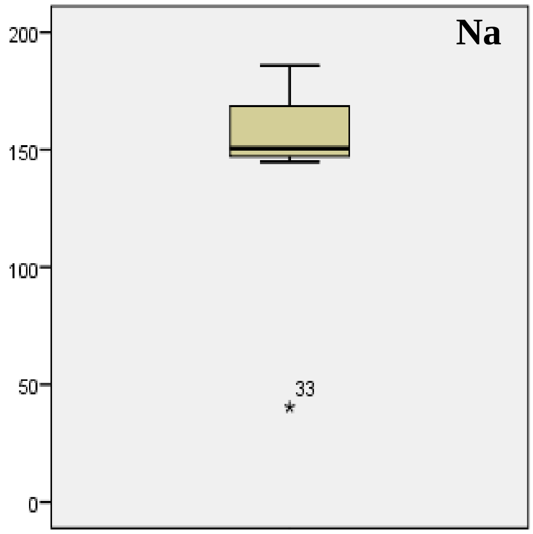


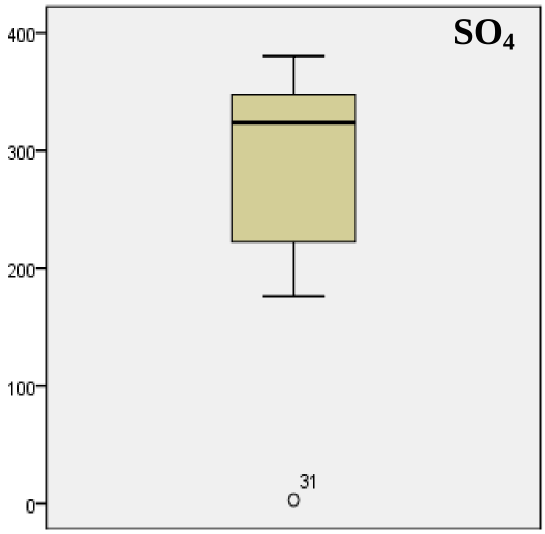


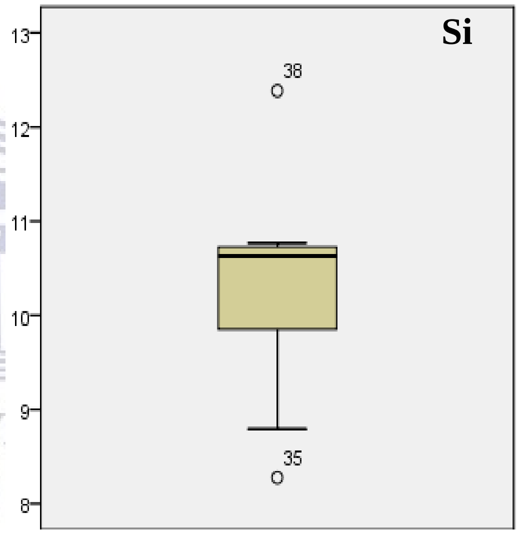


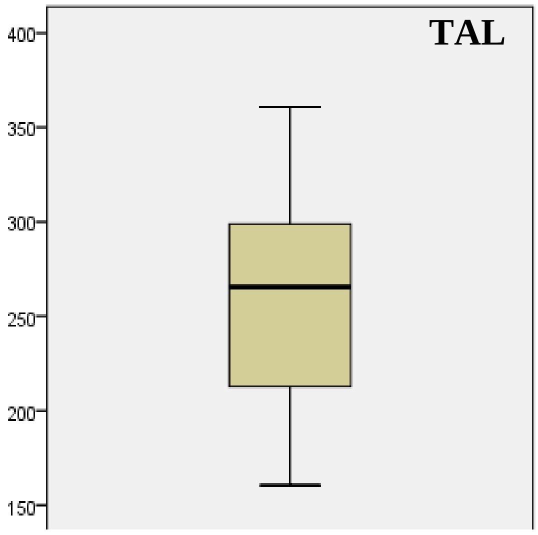


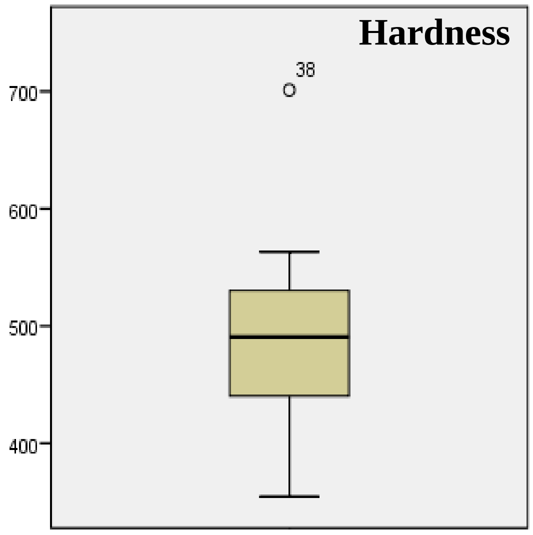

NO_3
6
5
4
3
2
1
0

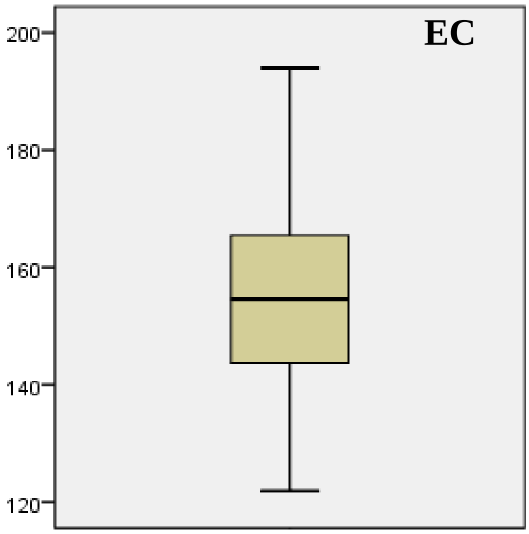

Na
200
150
100
50
0
33
*

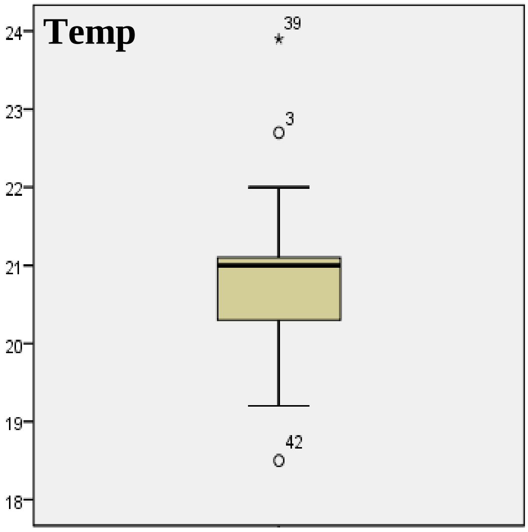

SO_4
400
300
200
100
0
31

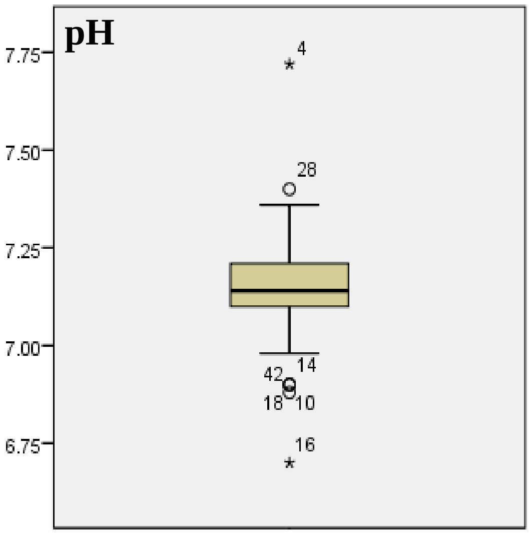

Si
13
12
11
10
9
8
38
35


TAL
400
350
300
250
200
150

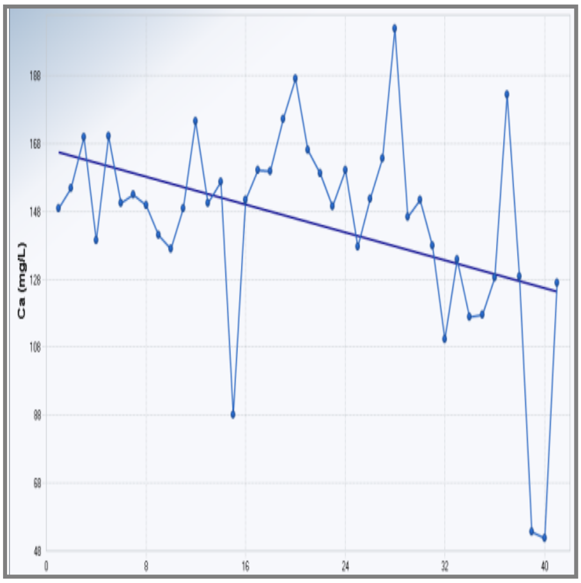

Hardness
700
600
500
400
38

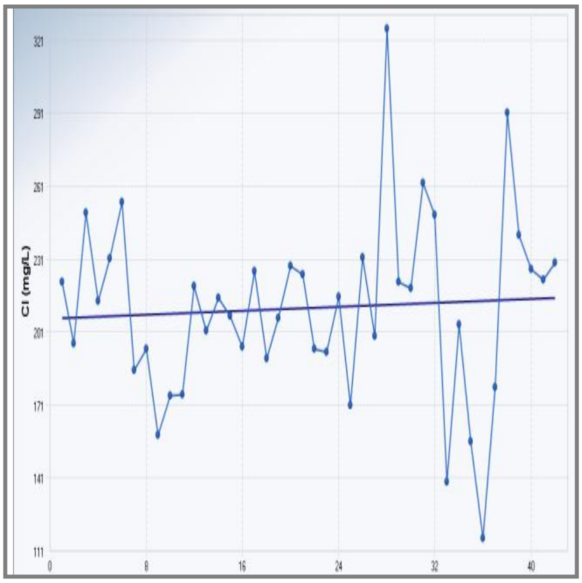

EC
200
180
160
140
120

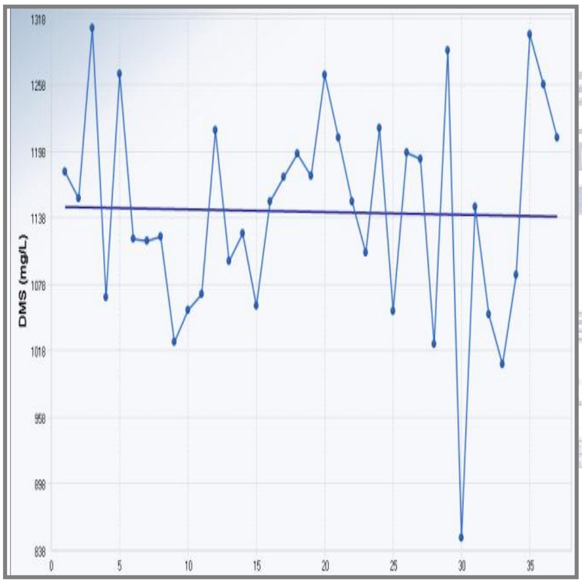

Temp
24
23
22
21
20
19
18
39
3
42

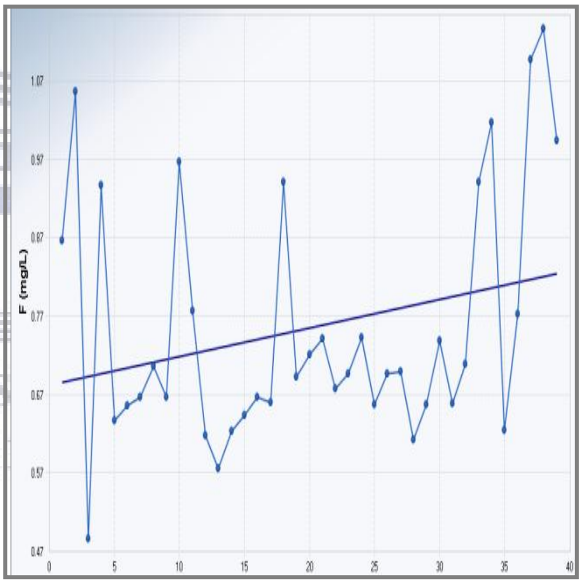

pH
7.75
7.50
7.25
7.00
6.75
4
28
14
42
18
10
16

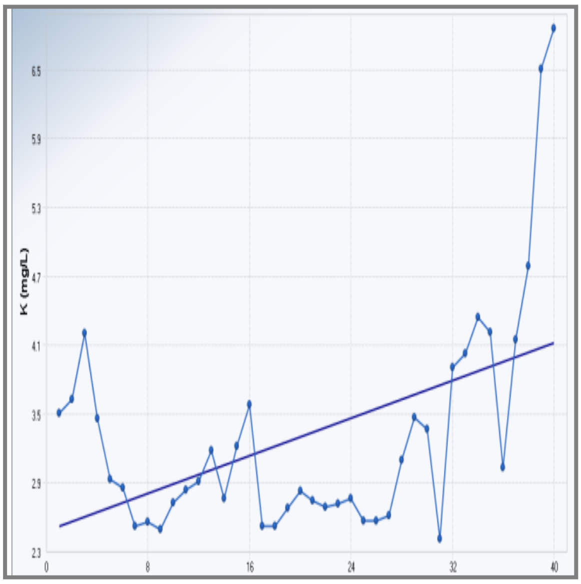
1.5.4. Time-series plots for ZQMRSK1

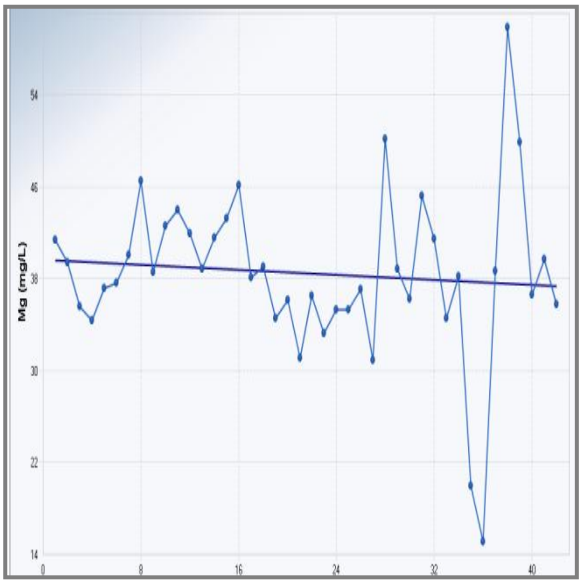


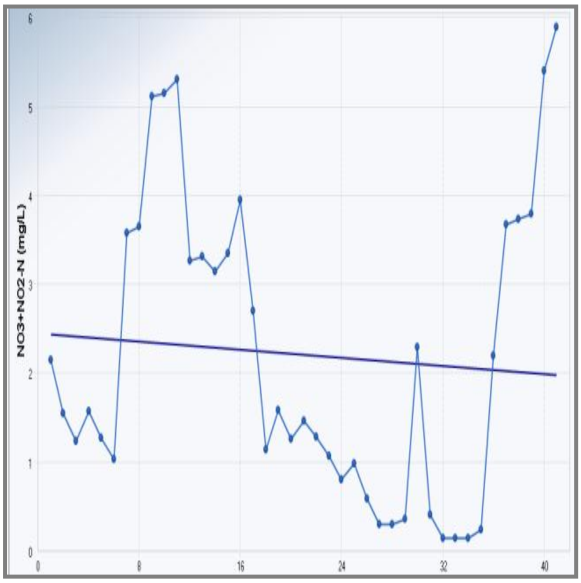


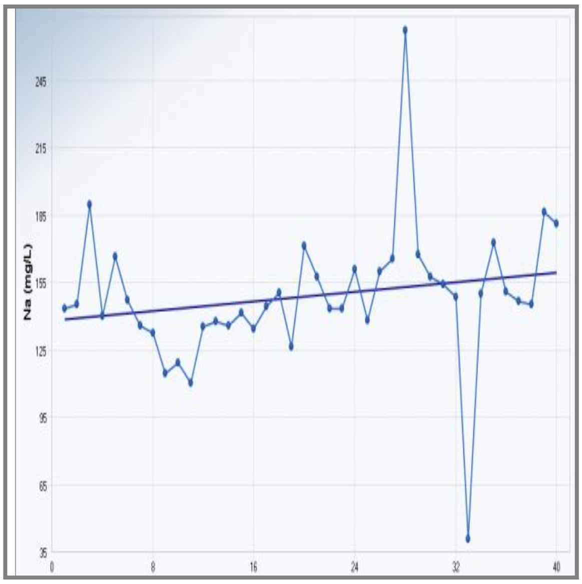


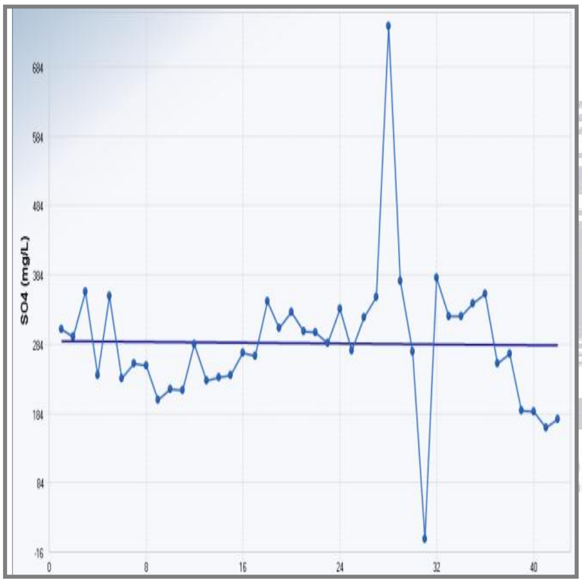


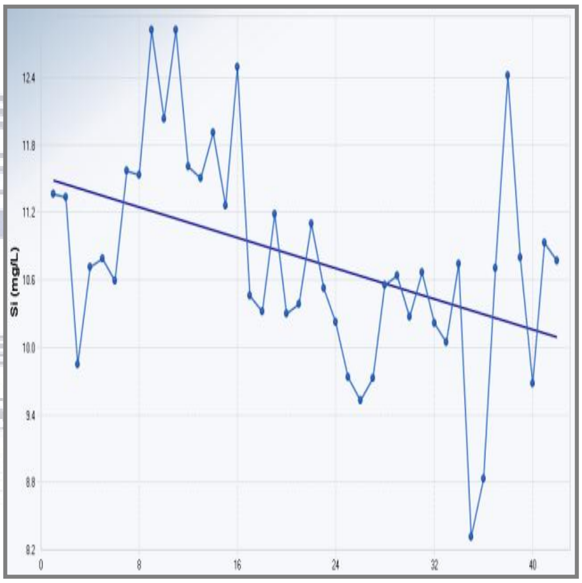


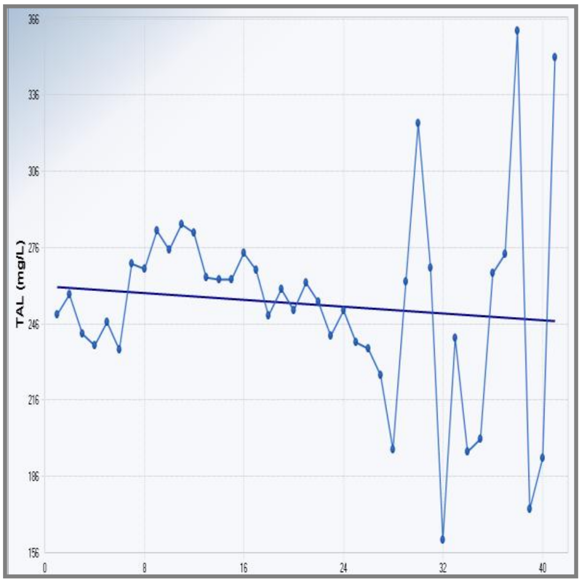


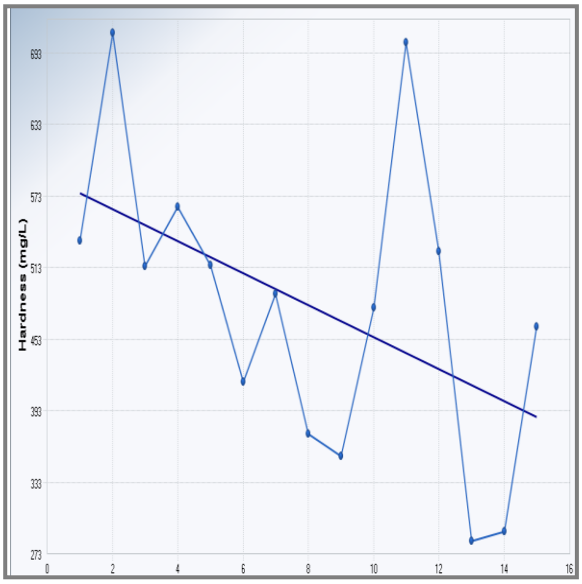

NO3+NO2-N (mg/L)
0
1
2
3
4
5
6
0
8
16
24
32
40

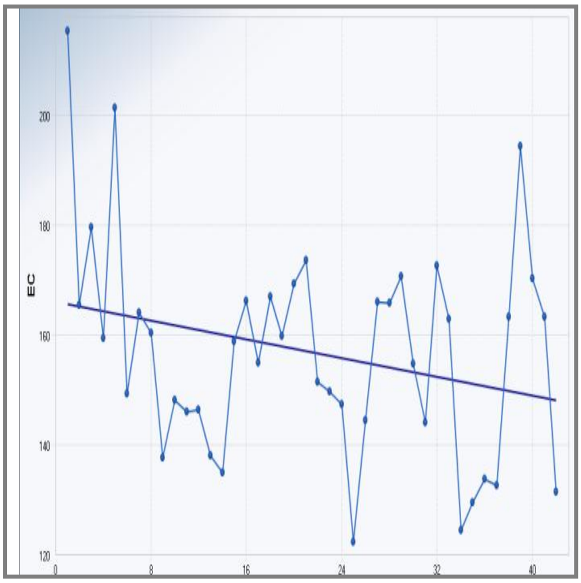

Na (mg/L)
35
65
95
125
155
185
215
245
0
8
16
24
32
40

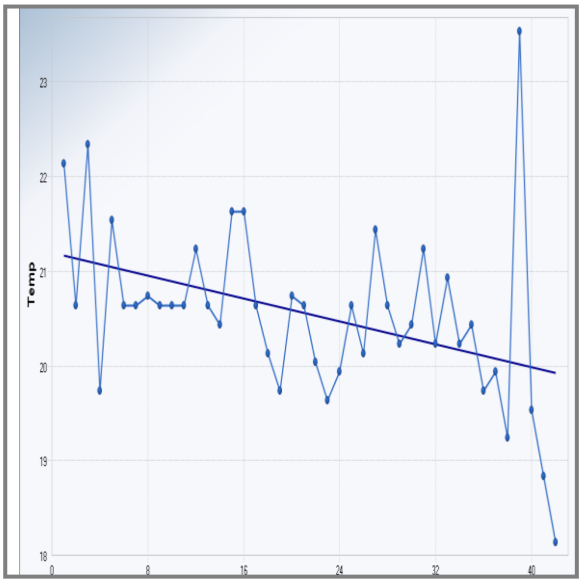

SO4 (mg/L)
-16
84
184
284
384
484
584
684
0
8
16
24
32
40

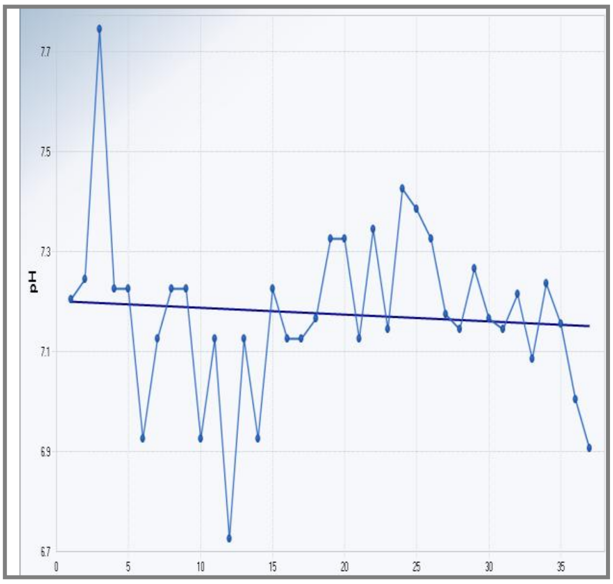

Si (mg/L)
8.2
8.8
9.4
10.0
10.6
11.2
11.8
12.4
0
8
16
24
32
40


TAL (mg/L)
156
186
216
246
276
306
336
366
0
8
16
24
32
40


Hardness (mg/L)
273
333
393
453
513
573
633
693
0
2
4
6
8
10
12
14
16

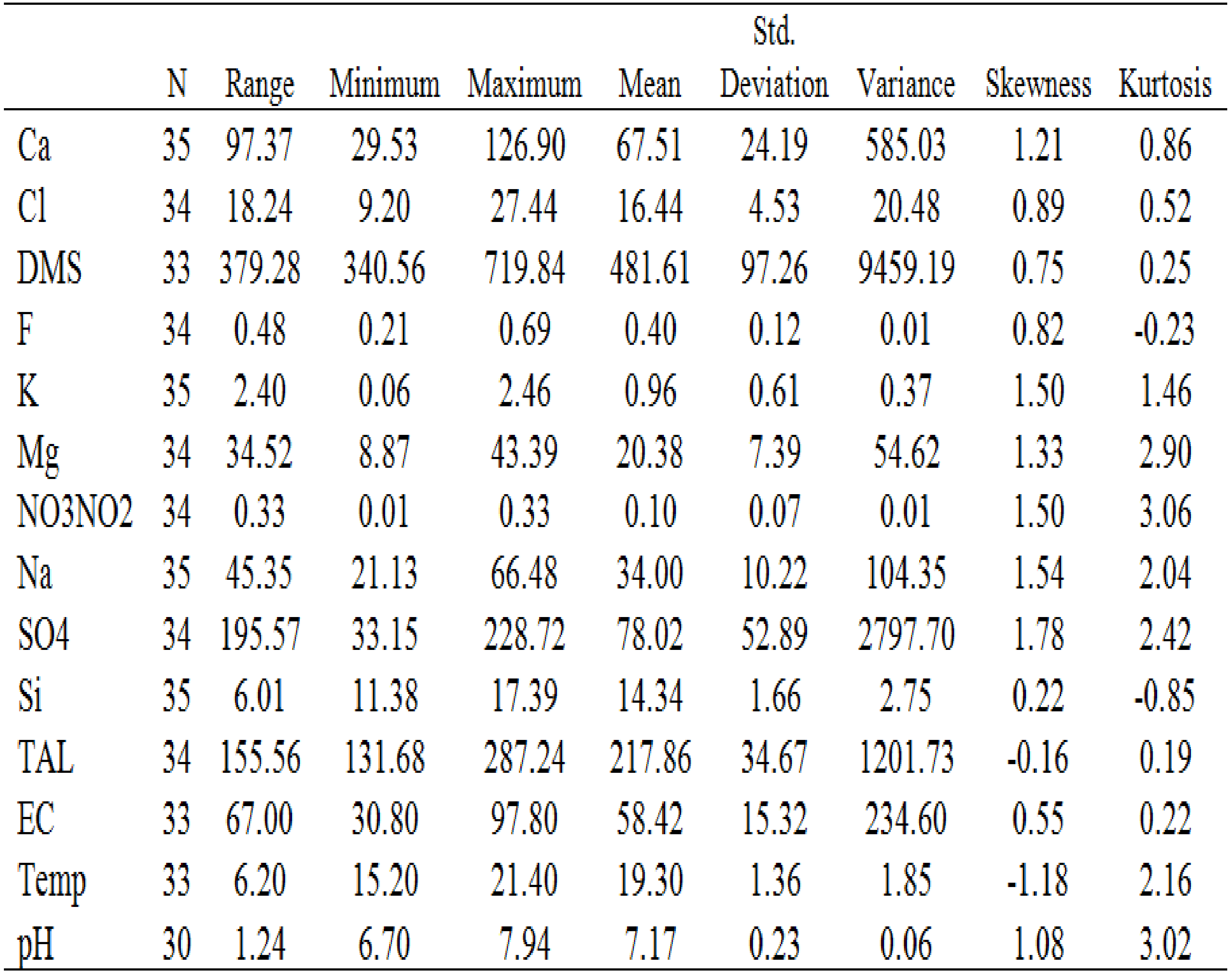

EC
200
180
160
140
120
0
8
16
24
32
40


Temp
23
22
21
20
19
18
0
8
16
24
32
40


pH
7.7
7.5
7.3
7.1
6.9
6.7
0
5
10
15
20
25
30
35

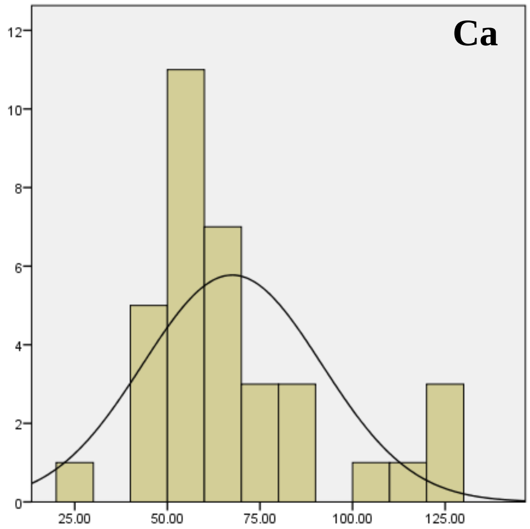
1.6.1. Exploratory statistical analysis for ZQMBWW1

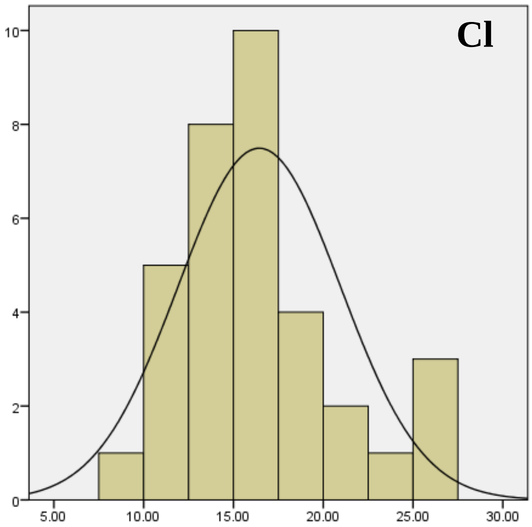
Table 1.5. Descriptive statistics for ZQMBWW1

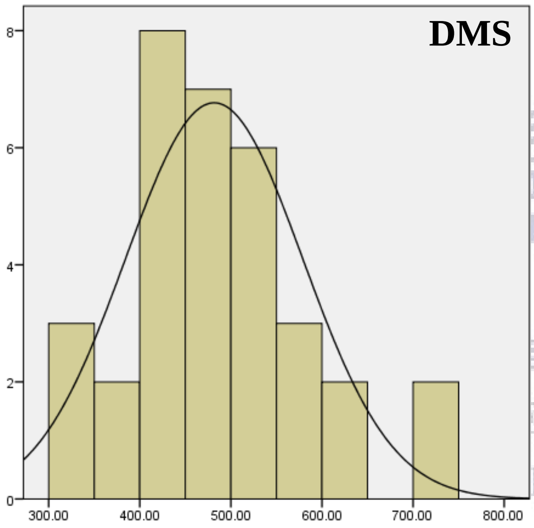


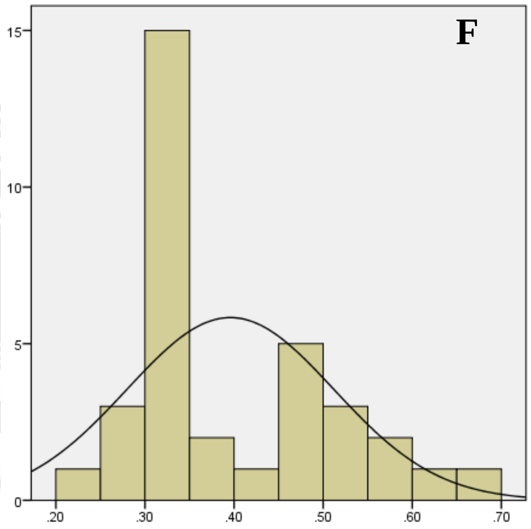
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Ca	35	97.37	29.53	126.90	67.51	24.19	585.03	1.21	0.86
Cl	34	18.24	9.20	27.44	16.44	4.53	20.48	0.89	0.52
DMS	33	379.28	340.56	719.84	481.61	97.26	9459.19	0.75	0.25
F	34	0.48	0.21	0.69	0.40	0.12	0.01	0.82	-0.23
K	35	2.40	0.06	2.46	0.96	0.61	0.37	1.50	1.46
Mg	34	34.52	8.87	43.39	20.38	7.39	54.62	1.33	2.90
NO3NO2	34	0.33	0.01	0.33	0.10	0.07	0.01	1.50	3.06
Na	35	45.35	21.13	66.48	34.00	10.22	104.35	1.54	2.04
SO4	34	195.57	33.15	228.72	78.02	52.89	2797.70	1.78	2.42
Si	35	6.01	11.38	17.39	14.34	1.66	2.75	0.22	-0.85
TAL	34	155.56	131.68	287.24	217.86	34.67	1201.73	-0.16	0.19
EC	33	67.00	30.80	97.80	58.42	15.32	234.60	0.55	0.22
Temp	33	6.20	15.20	21.40	19.30	1.36	1.85	-1.18	2.16
pH	30	1.24	6.70	7.94	7.17	0.23	0.06	1.08	3.02

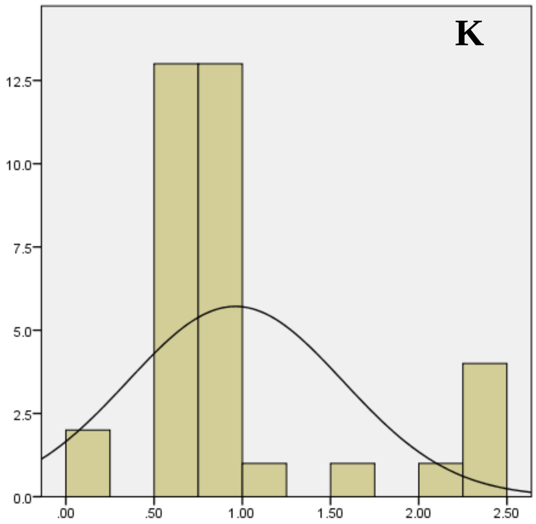
1.6.2. Histograms for ZQMBWW1

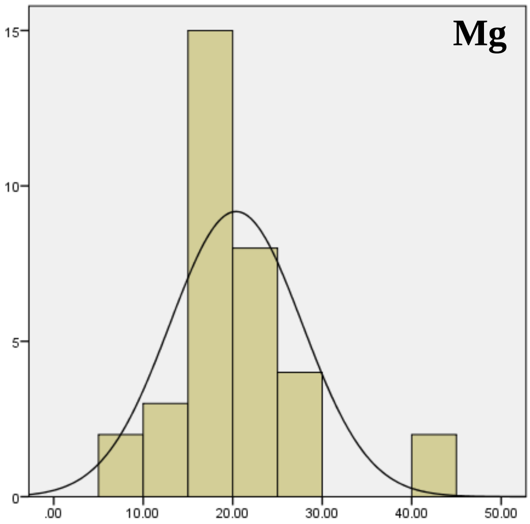


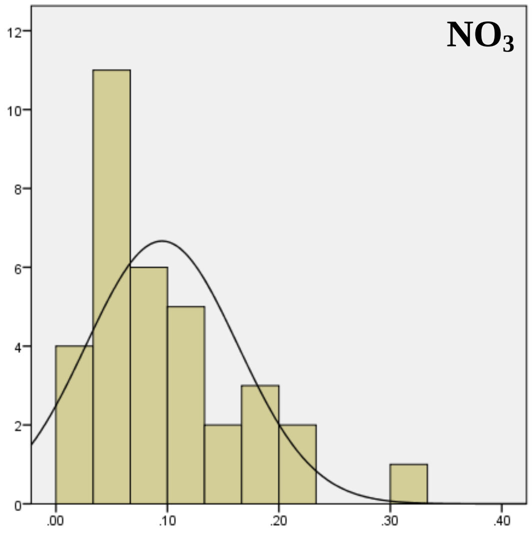


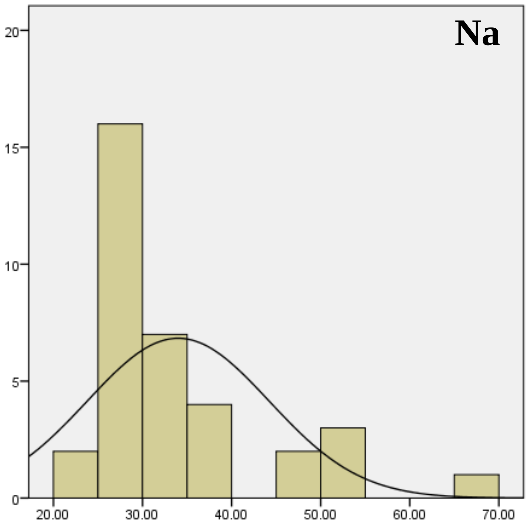


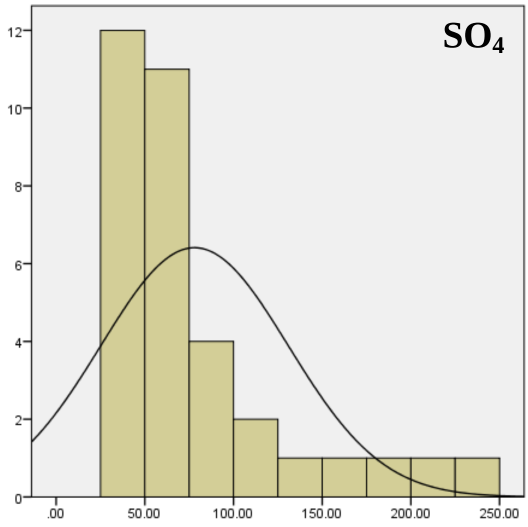


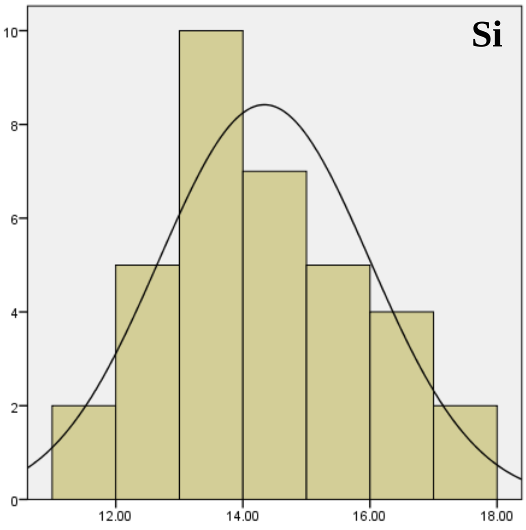


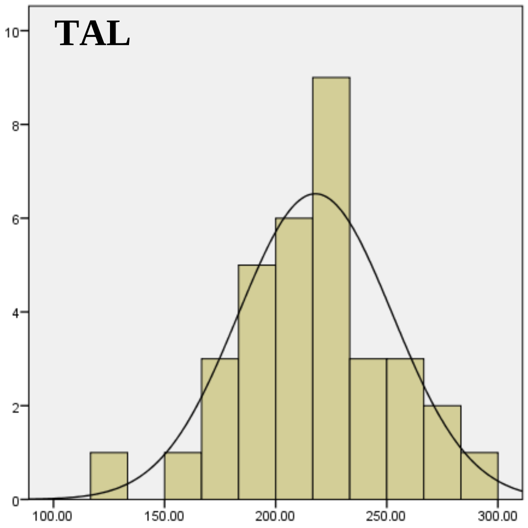


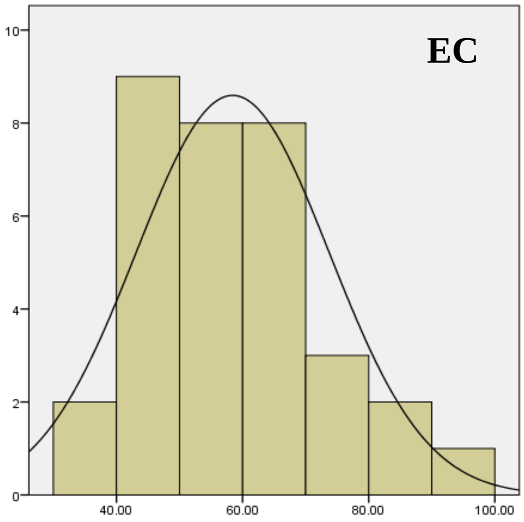

NO$_3$
0
2
4
6
8
10
12
.00
.10
.20
.30
.40

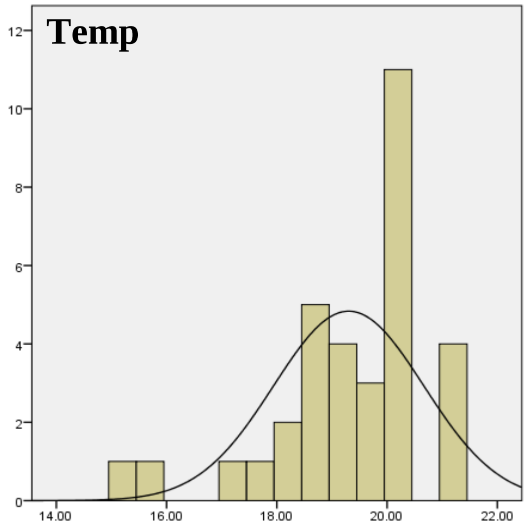

Na
0
5
10
15
20
20.00
30.00
40.00
50.00
60.00
70.00

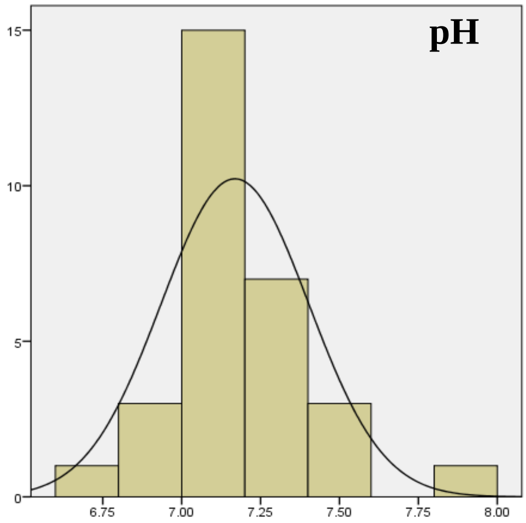

SO$_4$
0
2
4
6
8
10
12
.00
50.00
100.00
150.00
200.00
250.00


Si
0
2
4
6
8
10
12.00
14.00
16.00
18.00

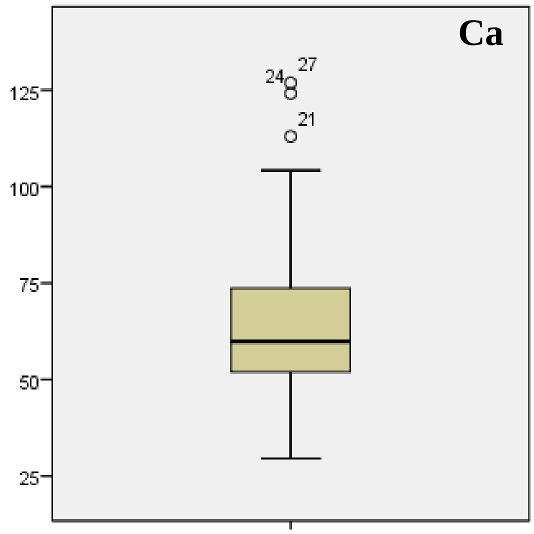

TAL
0
2
4
6
8
10
100.00
150.00
200.00
250.00
300.00

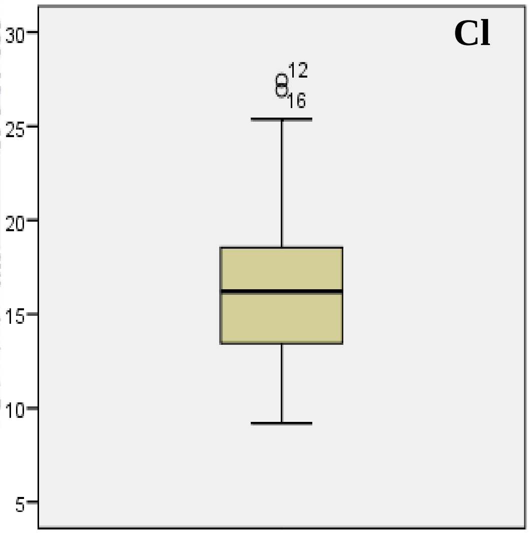

EC
0
2
4
6
8
10
40.00
60.00
80.00
100.00

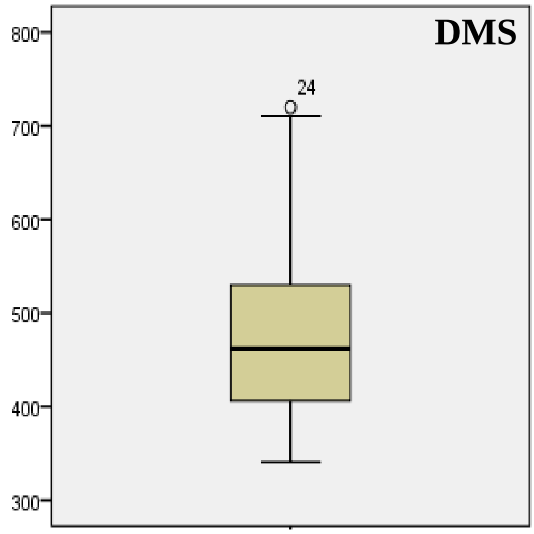


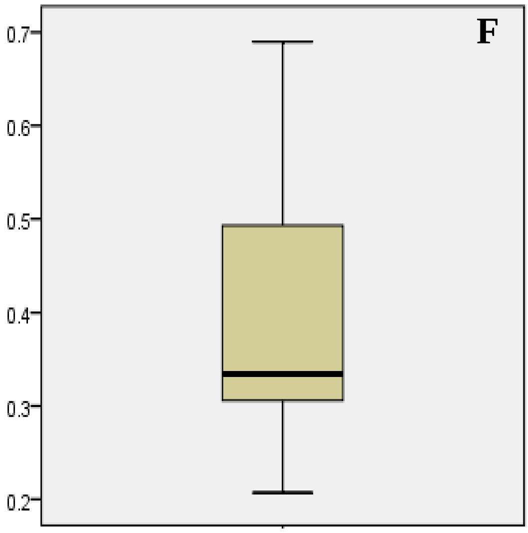


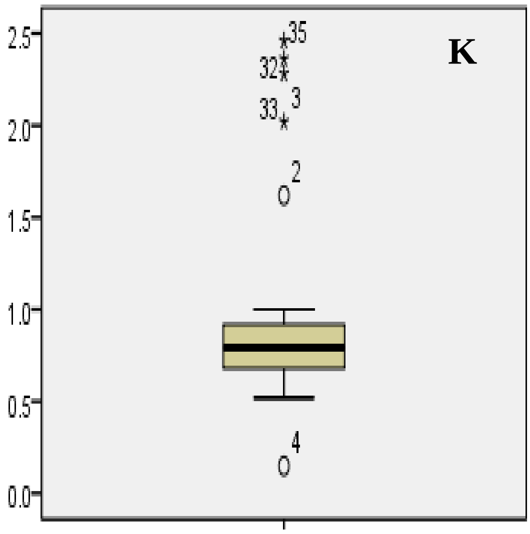
1.6.3. Box & Whisker plots for ZQMBWW1

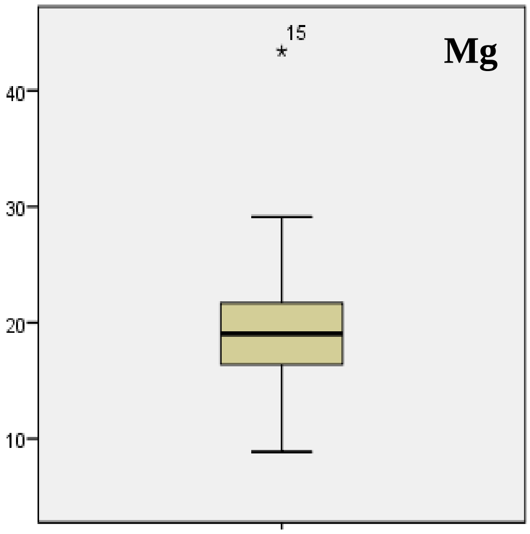


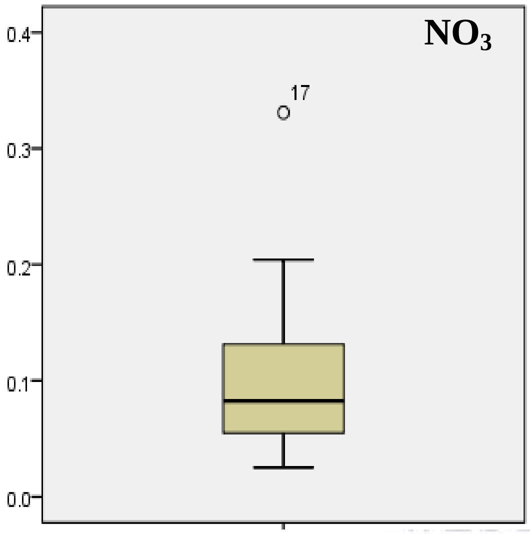


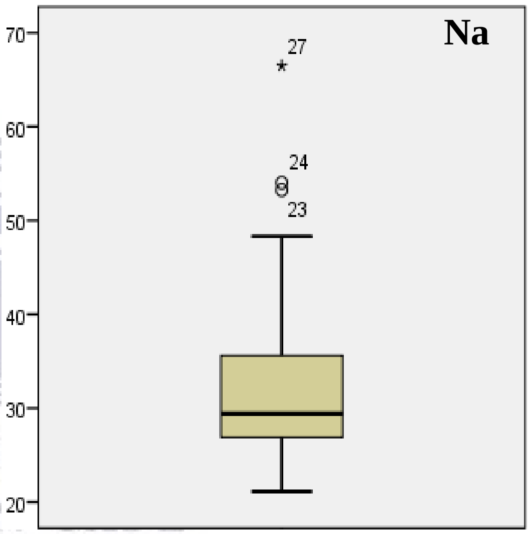


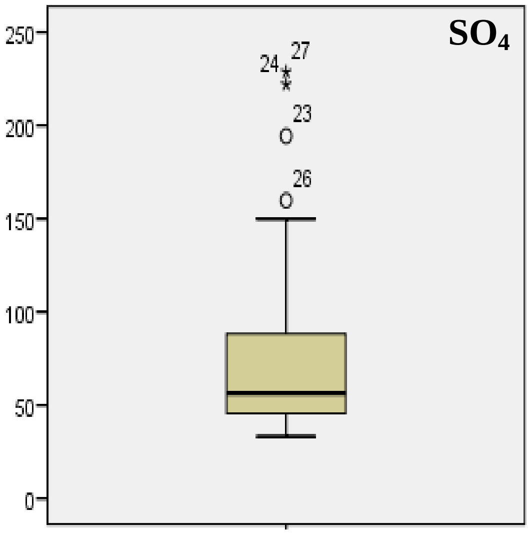


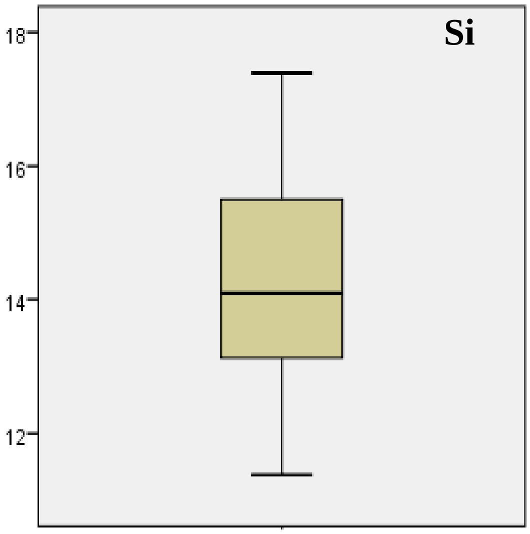

K
2.5
2.0
1.5
1.0
0.5
0.0
35
32
3
33
2
4

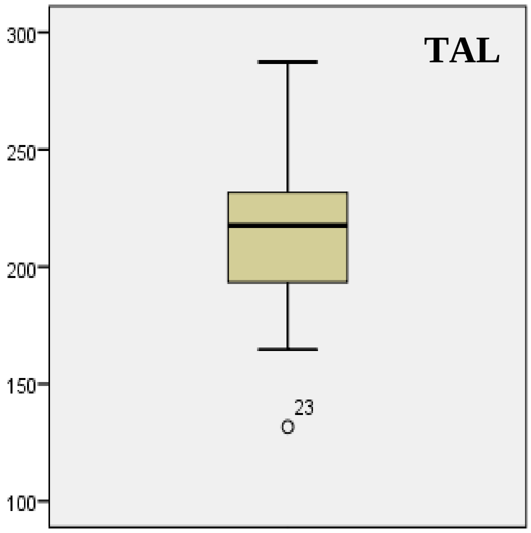

Mg
40
30
20
10
15

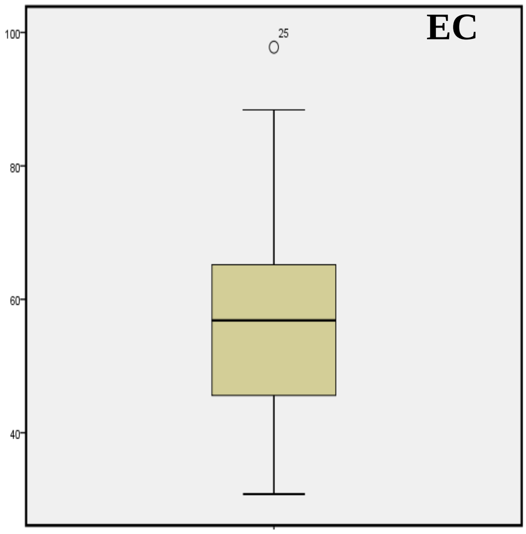

NO3
0.4
0.3
0.2
0.1
0.0
17

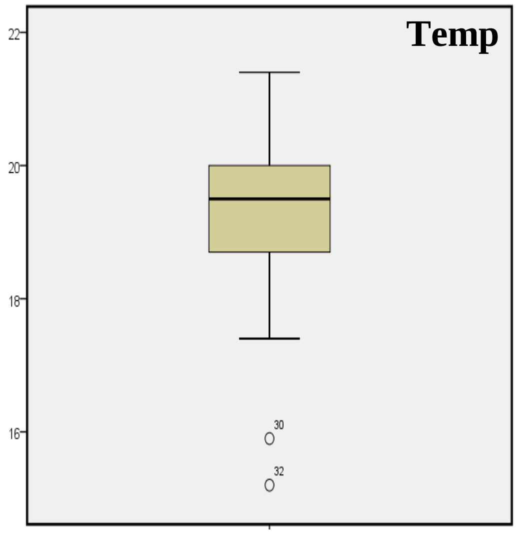

Na
70
60
50
40
30
20
27
24
23

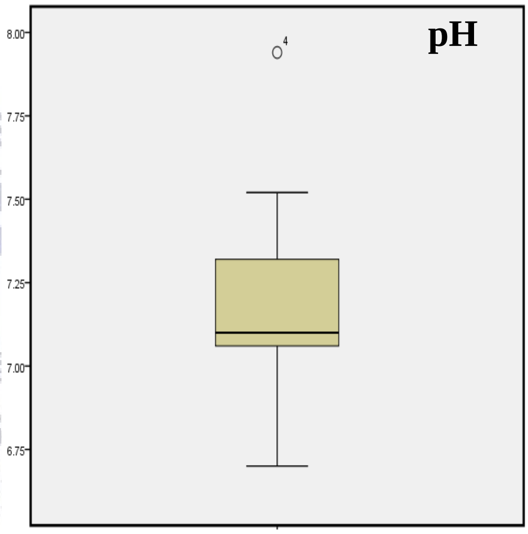

SO4
250
200
150
100
50
0
27
24
23
26


Si
18
16
14
12

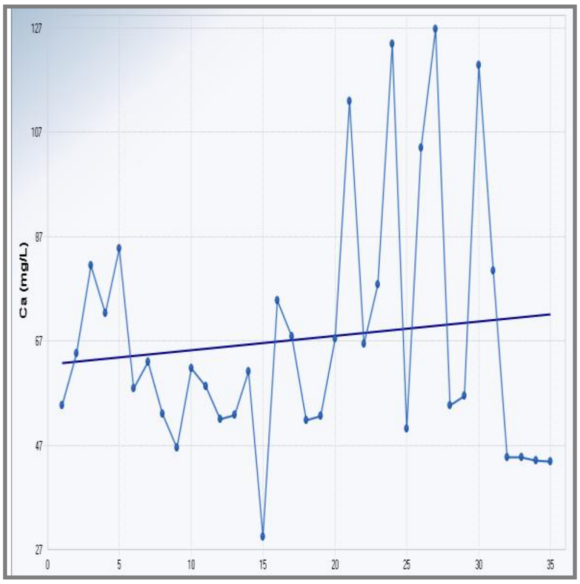

TAL
300
250
200
150
100
23

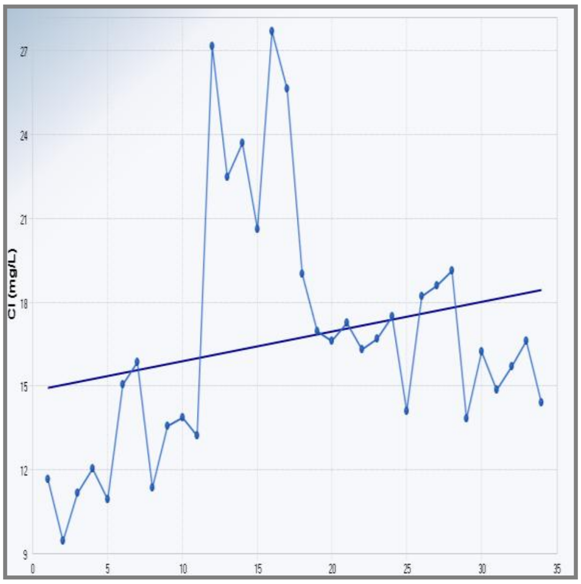

EC
100
80
60
40
25

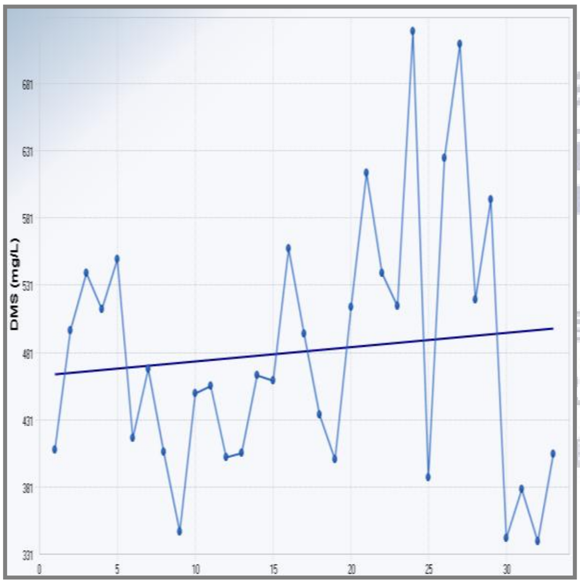

Temp
22
20
18
16
30
32

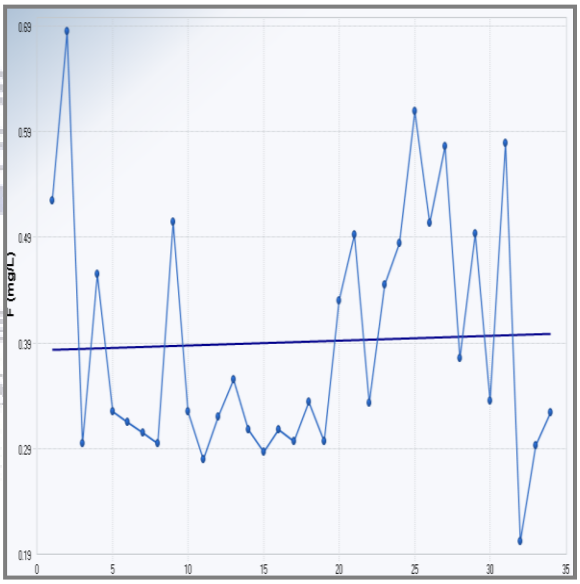

pH
8.00
7.75
7.50
7.25
7.00
6.75
4

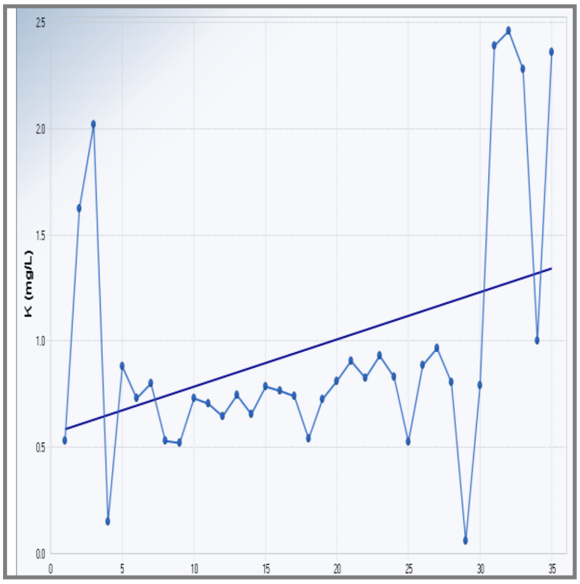
1.6.4. Time-series graphs for ZQMBWW1

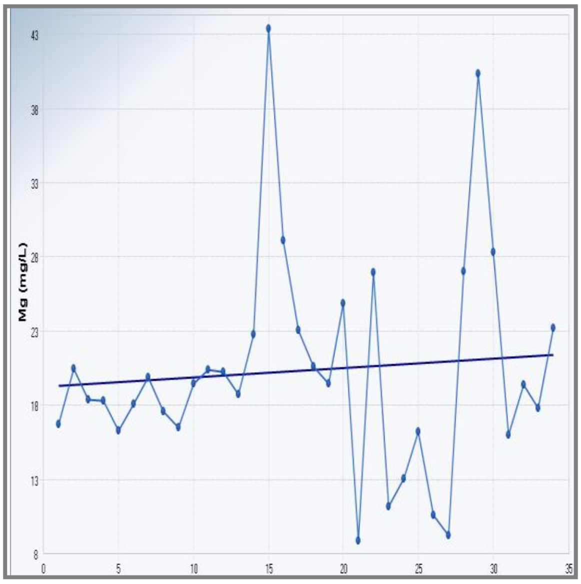


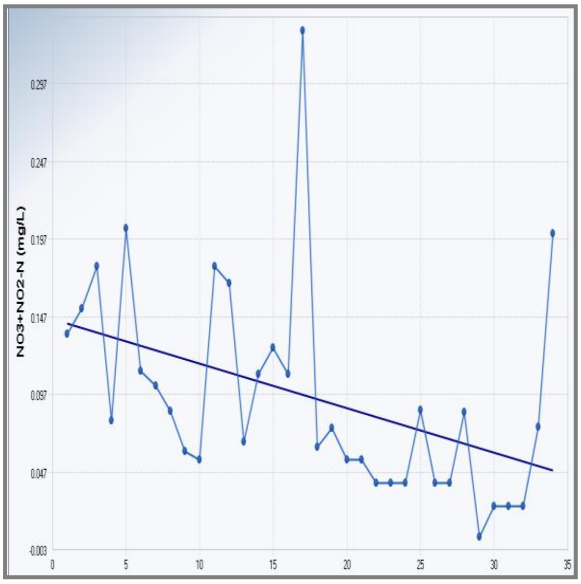


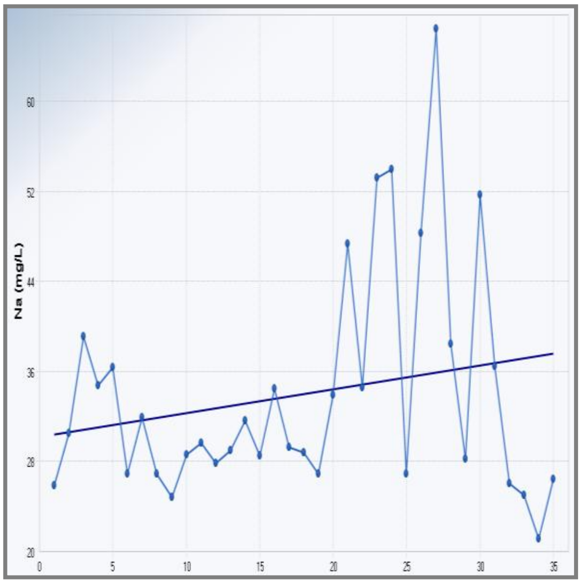


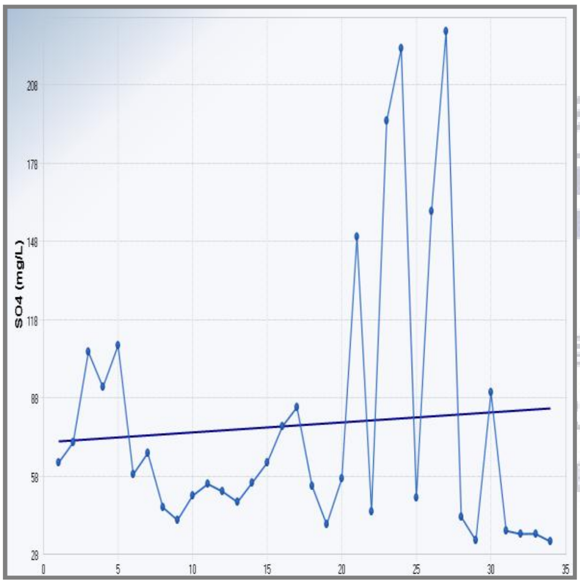


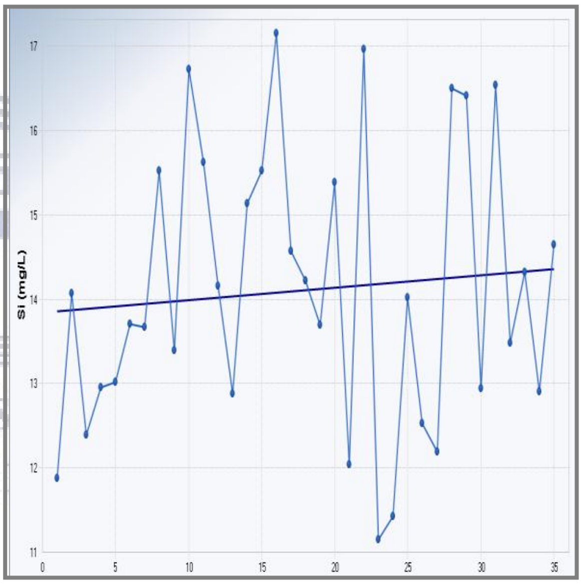


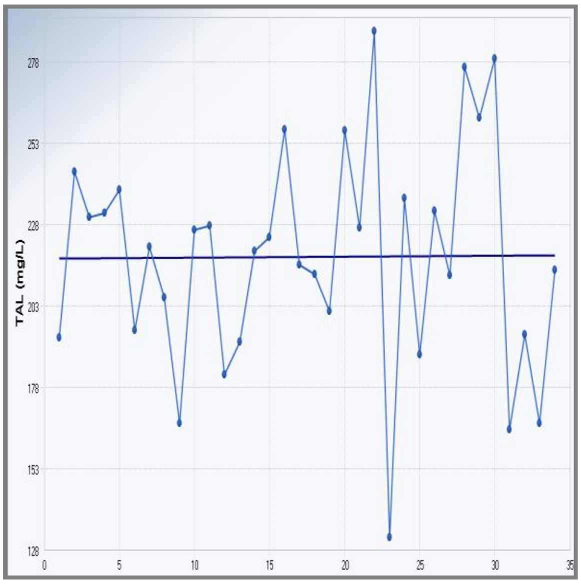


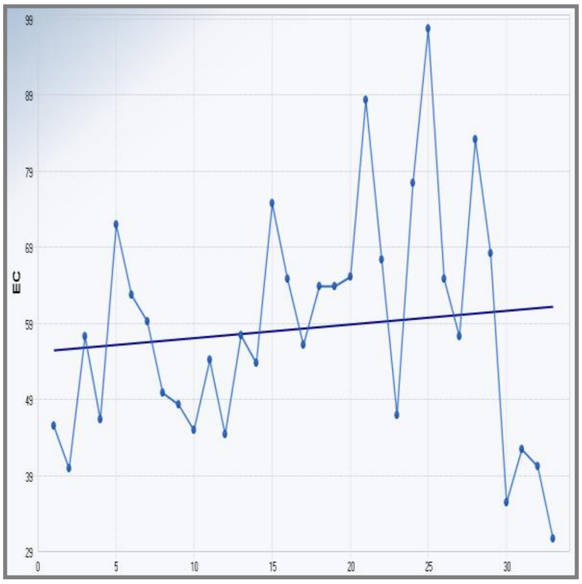

NO3+NO2-N (mg/L)
0.297
0.247
0.197
0.147
0.097
0.047
-0.003
0
5
10
15
20
25
30
35

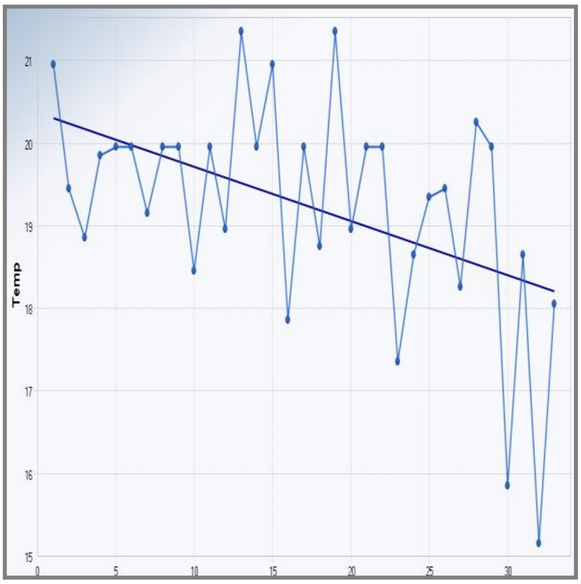

Na (mg/L)
60
52
44
36
28
20
0
5
10
15
20
25
30
35

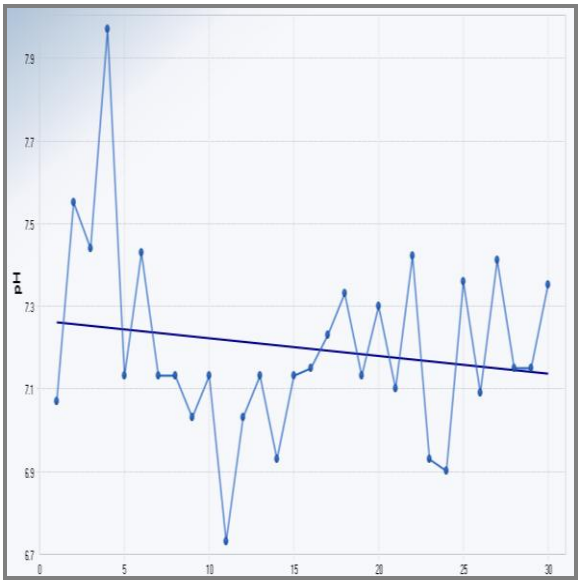

SO4 (mg/L)
208
178
148
118
88
58
28
0
5
10
15
20
25
30
35


Si (mg/L)
17
16
15
14
13
12
11
0
5
10
15
20
25
30
35


TAL (mg/L)
278
253
228
203
178
153
128
0
5
10
15
20
25
30
35


EC
99
89
79
69
59
49
39
29
0
5
10
15
20
25
30


Temp
21
20
19
18
17
16
15
0
5
10
15
20
25
30


pH
7.9
7.7
7.5
7.3
7.1
6.9
6.7
0
5
10
15
20
25
30