Statistical summary of Pilbara-wide hydraulic conductivity data derived from aquifer pumping tests

By

Nashel Kalaka

Thesis
Submitted to Flinders University
for the degree of

Master of Science (Groundwater Hydrology)

College of Science and Engineering
06th December 2021

TABLE OF CONTENTS

ABSTRACT

Rio Tinto hydraulic parameters database spans multiple deposits, and data acquisition is vast compared to other global mining houses. Data acquisition is at the centre of the Water Resource Department. Understanding hydraulic parameters and the formulation of statistics and probability density functions of these parameters is crucial to inform the analytical and numerical modelling of deposits and water resources. Historically hydraulic aquifer testing has been undertaken for several reasons including, dewatering, water supply and environmental impact assessments. Hydraulic aquifer properties such as hydraulic conductivity (K), transmissivity (T), storativity (S) and specific yield (S_y) are fundamental aquifer input parameters required for groundwater numerical and analytical modelling. There is no formal process within the Water Resource Evaluation (WRE) group about how aquifer parameters are assigned to analytical or numerical groundwater models. The primary objectives of this research aim to establish statically how hydraulic conductivity varies across deposit, site, and aquifer types.

The datasets used for this research were sourced from Rio Tinto's internal databases. After data collation, cleaning of datasets was undertaken to remove data missing observations or values and column names. Data analysis was undertaken using the R programming code.

The research concluded that the lognormality of the hydraulic conductivity distribution (K) generally holds and is supported by conceptual knowledge; although, there are some data deviations in the tails. The deviations in the distributed tails are likely to be attributed to fractured rock (i.e., right-skewed) and low permeability shales (i.e., left-skewed).

It is observed that the hydraulic conductivities of aquifer types are within approximately one order of magnitude. Individual interquartile ranges of hydraulic conductivity (K) for each aquifer type generally cover less than an order of magnitude; however, CID and MM have noticeably larger interquartile ranges. The CID, DG_M, LowP and Witt_K have very low permeability outliers. The mean distribution of hydraulic conductivity does not differ between mineralised and unmineralised units. It was also observed from the hydraulic parameters that there are more extensive ranges of hydraulic conductivity for Yandicoogina, Marandoo and Hope Downs 4. There are noticeably higher mean values for Marandoo than the rest of the sites. It can be attributed to drilling into the orebody aquifer type at Marandoo. The Yandi channel iron deposit has a noticeable right-shifted distribution (i.e., on average, higher values) of K by plotting hydraulic conductivity by the site.

The research recommends further reconciling the early stage pumping data and validating the yields against geology to determine whether geology is a factor. In addition, it is recommended that further study is required to correlate airlift yields with permeability, transmissivity, and other relevant hydraulic parameters. It would validate an approach that combines statistical learning with well-

completed data. It is also recommended to transfer or automate the extraction of relevant PDF files into a centralised database to extract additional data and enhance parameter knowledge.

DECLARATION

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university and that to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed: Nashel Kalaka

Date: 06th December 2021

ACKNOWLEDGEMENTS

It is with gratitude and great pleasure to acknowledge and thank the people who significantly contributed to the completion of this research. Firstly, I would like to thank supervisors Professor Peter Cook (NCGRT) and Dr Ilka Wallis for their patience, motivation, continuous support, and immense knowledge in all aspects of this study

I also take this opportunity to appreciate the support received from Keith Brown, the Principal Hydrogeologist, Stan Blank and Kris Longmore for their significant contributions and discussions regarding my research**.**

Finally, my appreciation also goes to my family and friends for their continuous support and ongoing encouragement throughout my years of study.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Background

The Rio Tinto database of the hydraulic parameters captures data across a range of deposits in the Pilbara region of Western Australia. The data acquisition is vast compared to other mining houses across the globe and is at the centre of the Water Resource Evaluation Department (WRE). Historically in Rio Tinto, Iron Ore aquifer pumping tests have been undertaken for several reasons including, dewatering, water supply and environmental impact assessments. Aquifer parameters generated from pumping tests such as hydraulic conductivity (K), transmissivity (T), storativity (S) and specific yield (S_y) are fundamental aquifer input parameters required for numerical and analytical modelling of groundwater. Currently, no formal process exists within the WRE department regarding how aquifer parameters are assigned to an analytical or numerical groundwater model.

Understanding hydraulic parameters and the formulation of statistics and probability density functions is crucial to inform analytical and numerical models. It is challenging to inform and make decisions supported by aquifer parameters without drawing insight from the vast in-house dataset. Therefore, this study aimed to formalise this process so that parameters used in our groundwater models are more transparent in analytical or numerical groundwater models.

1.2 Problem Statement

Rio Tinto Iron Ore (RTIO) has many iron ore deposits across the Pilbara, and more importantly, completed more aquifer tests than anyone else over the past five decades of mining. However, this data is currently not drawn together to inform groundwater modellers and practitioners about the variability of aquifer parameters within the same geological units across different locations. The knowledge is crucial for predicting groundwater drawdowns, recovery predictions, mine closure, and environmental impact assessments during dewatering.

1.3. Research Objectives

The primary objective of this research is:

- to determine how aquifer hydraulic conductivity is distributed across Iron ore deposits and by aquifer type.

1.4 Thesis structure

Chapter 1 provided a research background, objectives and methodologies undertaken for this study. Chapter 2 discusses the literature research relevant to this study, regional hydro-stratigraphically sequence, and then hydrogeological settings of the Pilbara region are detailed in Chapter 2. Chapter

3 details the research methods and describes how the data was aggregated from different internal Rio Tinto sources, data cleaning and data transformation into a format before importing into R-programming source code. Chapter 4 details the data analysis and statistical distributions generated. Chapter 5 summarises the research findings, conclusions, and further research suggestions.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

A vast and growing body of literature was reviewed detailing hydraulic properties in the fractured, porous, and spatial variation of aquifer properties in heterogeneous media (Clauser, 1992; Brace, 1980 & 1984; Freeze, 1975; Bear, 1972; Neumann, 1990). This chapter begins with the literature review on hydraulic conductivity and its importance in understanding how groundwater flows in the porous medium. Following this, the literature on spatial variability of hydraulic properties. The review of statistical methods is discussed in this chapter as well. The last section discusses the Pilbara hydrogeological characterisation of the Pilbara, climate, and regional geology.

2.1.1 Hydraulic Conductivity (K) and Transmissivity (T)

An understating of groundwater hydraulics is based upon the fundamental properties of water or fluid and the media through which water moves. The property that describes how the water moves through the media is hydraulic conductivity (Equation 2). Another critical parameter is transmissivity that is the product of hydraulic conductivity multiplied by formation thickness. These hydraulic parameters or properties vary spatially or in space due to the formation's complex geological processes or deposition history (Kolterman and Gorelick, 1996).

Darcy (1856) described hydraulic conductivity as a coefficient of proportionality that describes the rate of fluid moving through a porous medium. Darcy's experiment is an empirical law upon which groundwater science development continues to evolve. Figure 2.1 indicates the generalised experimental apparatus applied by Darcy.

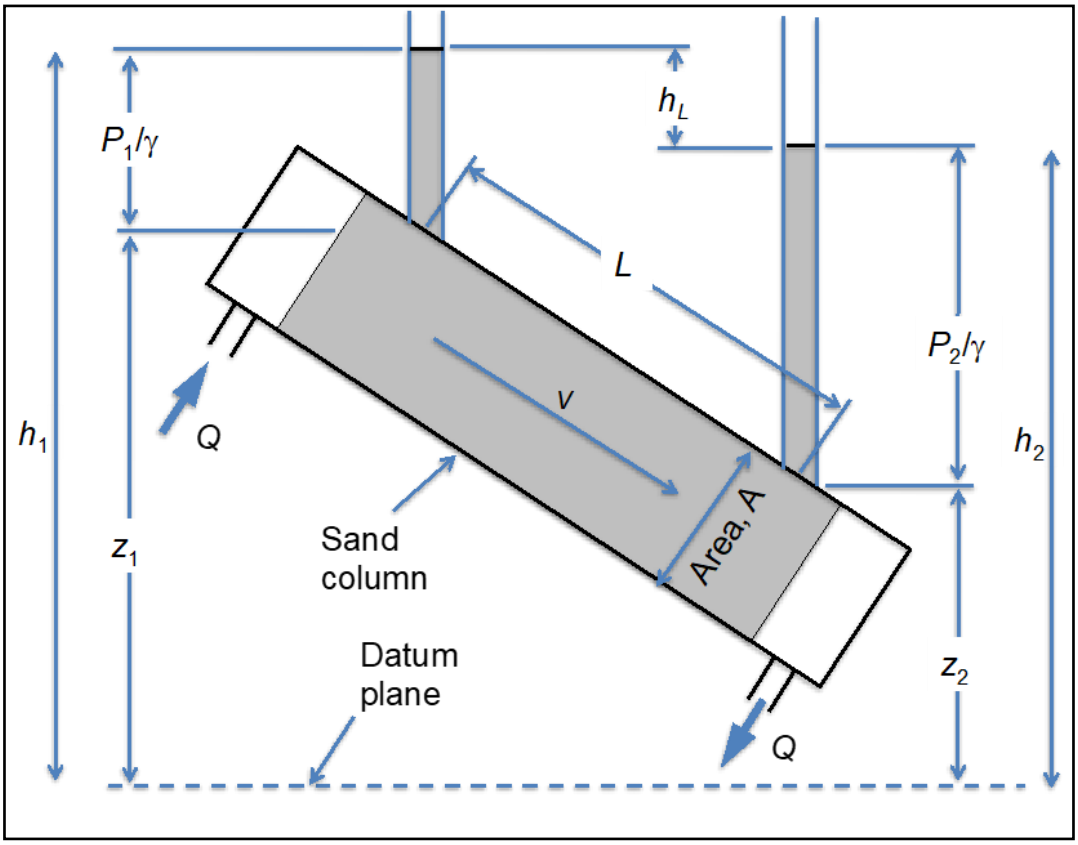

Figure 2.1 Darcy's Generalised Experiment

Darcy concluded that the discharge is proportional to the area, head difference and inversely proportional to the length (Darcy, 1856).

$$Q \propto A\frac{h_1 - h_2}{L}$$ equation 1

The proportionality constant in these relationships has been defined as hydraulic conductivity. Darcy's Law can then be written as:

$$Q = - KA\frac{\Delta h}{L}$$ equation 2

where:

- Q is the volume rate of flow (Length^3/Time)
- K is the hydraulic conductivity (Length/Time)
- A is the cross-sectional area to the direction of flow (Length^2)
- Negative indicates the direction of flow
- Δh is the hydraulic head loss (Length)
- Δh/L is the hydraulic gradient (dimensionless) and

- L is the length measured at Points 1 and 2

2.1.1.1 Pumping Test and Estimation of Aquifer Parameters

Pumping tests are an effective means of obtaining hydraulic parameters and determining aquifer boundaries. The tests can be a variable head, constant head, and recovery tests. A pumping test is a field study where a well is pumped at a regulated rate, and the pumped water response level is measured either from observation bores or the same pumping test well. The collected response data is used to estimate the hydraulic features of aquifers, assess the performance of the well, and identify the aquifer's boundaries (Floriancic et al., 2019). A pumping test such as the constant rate test aims to estimate the hydraulic features of a given aquifer system with specific reference to such features as hydraulic conductivity, transmissivity, and storage coefficient. Pumping tests are used in layered systems to estimate the features of aquitards. Pumping tests can identify and note no flow and recharge boundaries, which might also limit the aquifer's lateral extent (Nijp et al., 2019).

The following is the summary of the parameters that may be estimated from pumping tests:

- Hydraulic conductivity (m/d)
- Transmissivity (m^2/d)
- Storativity of the confined aquifer (Dimensionless) and
- Storage coefficient or Specific yield of the unconfined aquifer (%)

The hydraulic conductivity parameter can also be estimated from the transmissivity calculated as hydraulic conductivity multiplied by the thickness of the aquifer or screen thickness of the bore (Asfahani, 2021). It is essential to note that the pumping aquifer tests only provide an average hydraulic conductivity of the formation, given by the transmissivity divided by the aquifer thickness or screen length.

2.2 Aquifer Spatial Variability

The literature has substantial attention detailing the spatial heterogeneity of hydraulic properties on both small-scale and regional scales. In addition, a good understanding of scale dependence of hydraulic conductivity or transmissivity results from data and statistics (Corvi et al., 1992). Clauser (1992) and Brace (1980, 1984) detailed that the permeabilities of sedimentary (porous) and crystalline (fractures) rocks vary when considering the scale effect. The research also explored how crystalline (fractured) and sedimentary permeability vary depending on the size or volume of material under consideration. They concluded that the rocks exhibited a scale effect, which meant that the larger the characteristic volume analysed, the greater the permeability. Clauser (1992) pointed out that permeability in fractured controlled or compacted rock mass exhibited variable permeability. Another interesting finding based on the study done by Neumann (1990) concluded that both the

porous and fractured media appear to indicate scaling rules when dealing with the flow of groundwater and transport simulation.

Similarly, the study by Butler and Liu (1993) concluded that the hydraulic properties were sensitive to drawdown when they had reviewed the pumping test in heterogeneous media. In addition, it was found that constant pumping tests were insufficient to characterise the hydraulic properties of the non-uniform formation. They suggested using the step test method in aquifers exhibiting large variability.

Sánchez-Vila et al. (1994) further explored similar work done by Clauser (1992) that effective transmissivity or hydraulic conductivity often increases with an increased observation scale. In addition, the author highlighted that the multi lognormality appears not to be valid in many cases, although the transmissivity values display a log-normal distribution. Sánchez-Vila et al. (1994) also suggested that the scale dependence of transmissivity and hydraulic conductivity is influenced by spatial variability. Another research undertaken by Sánchez-Vila et al. (1999) supports spatial varying of transmissivity in heterogeneous aquifer but with constant storativity.

In the same vein, the research by Oliver (1993) also found that the small-scale effects cause the deviation of transmissivity and storage near the pumping well. The influence of late-time drawdown at distant observation wells depends on the location of the non-uniformity, according to Oliver (1993). Oliver (1993) also found that analysing the drawdown curve near the pumping was challenging because the similarity of the small-scale feature is like the scale effect of large-scale non-uniformity at large distances.

On the other hand, Delphomme (1979) examined the use of geostatistical techniques to characterise the predicted hydraulic heads spatially and filed transmissivity uncertainty. The intent of this work observed field condition that at what point does the log transmissivity data knowledge lessen the uncertainty and head values. The paper deduced that when determining transmissivity from head values, the importance of the spatial variability of T has to be kept in mind. Asfahani (2021) published a paper on the classification of transmissivity spatial variation using statistical testing for the Krasny transmissivity approach to assessing groundwater supply potential.

Riva et al. (2017) investigated how the flow and transport in porous are strongly affected when the medium is heterogeneous. The author also reviewed that there are statistical moments of flow variable affected strongly by the deviation of log conductivity distribution from the Gaussian technique. In the same vein, they deduced that manifestation of non-Gaussian is the frequency distribution of Y, which often exhibits mild peaks and light tails that become heavier. Sen (1999) pointed out other hydrogeological variables that indicate similar behaviours comprise the log permeability of porous and fractured geologic media.

Freeze (1975) suggested using log-normal PDF to fit hydraulic conductivity data. The implication is that one can anticipate that K might vary over several orders of magnitude. Sen (1999) proposed using probabilistic and statistical methods when assessing groundwater yields, where the random character inherent in the aquifer parameters were deduced. The author also agreed with Freeze on using log-normal distribution for aquifer parameters. Sen (1999) further found that aquifer properties are ideal within the radius of influence; however, the local scale properties fail to provide a reliable solution on a regional scale due to heterogeneities, discontinuities, and anisotropies in the rock mass.

2.3 Review of Statistical Methods

It has been argued that statistical descriptions are a helpful way of deriving quick information about aggregated datasets (de Smith, 2015). The same can be applied in the hydrogeological discipline with aquifer parameters with a large enough dataset, and probability density functions (PDFs) are regularly used to convey the distribution of the dataset. Sen (1999) proposed using statistical methods to describe the groundwater flow rates based on randomness using log-normal distribution. The subsequent sections describe the most used statistical techniques to draw insight from the data in this study.

2.3.1 Histogram

A histogram shows the frequency distribution of a given set of data, hence providing a glance at the pattern of distribution charted in given categories. A uniform histogram shape indicates consistency with the data, and the frequency of every class tends to be the same as that of the other classes. A right-skewed histogram is characterised by a peak located at the left of the centre alongside a relatively more gradual tapering on the graph's right side. A left-skewed histogram appears to have a tail of the distribution going off to the left.

2.3.2 Quantile-Quantile (QQ) Plots

The quantile-quantile plot intends to indicate whether two data sets come from the same distribution. QQ plots are also deployed to determine the distribution type for the case of a random variable. The QQ plot intends to compare the distribution of the data set to a theoretical log-normal distribution with the same parameters (de Smith, 2015). The distribution is deemed to have a longer tail to the left or left-skewed if the bottom of the QQ plot ends deviate from the straight line even if the upper part is not. As the upper part of the plot deviates from the straight line, even as the lower section follows the line, the curve is said to have a longer till to the right and is positively skewed.

2.3.3 Box plots

A box plot is a technique used in the graphical illustration of the groups of numerical data via their quartiles and the median. Box plots might as well bear lines that extend from boxes showing the variability outside the lower and upper quartiles. The data midpoint is the median mark and is

depicted using a line that tends to be dividing the box into toe percentiles, i.e., 25th and 50th percentiles. It can determine how symmetrical or how the data is grouped tightly and whether skewed or not.

2.3.4 Density plots

Density distribution plots or Kernel density plots determine how a variable is distributed with the dataset. It depicts a smoothed distribution of points on the numeric axis. A skewed density curve to the left shows the mean is lower than the median, while one skewed to the left insinuates the mean is larger than the median. The mean is equal to the median suppose the density does not have any skew.

2.4 Study Area

The Pilbara region is in the Hamersley Province of Western Australia. Its landscape and wetlands are paramount to supporting water supply for Aboriginal communities and groundwater-dependent ecosystems (GDEs). The Pilbara region in Western Australia is widely known for mining houses owned by multinational Iron Ore companies (Department of Water, 2010).

Figure 2.2 indicates the location of Rio Tinto Mine operations. The Rio Tinto operations span the Hamersley Province of Western Australia that host world-class high-purity hematite orebodies.

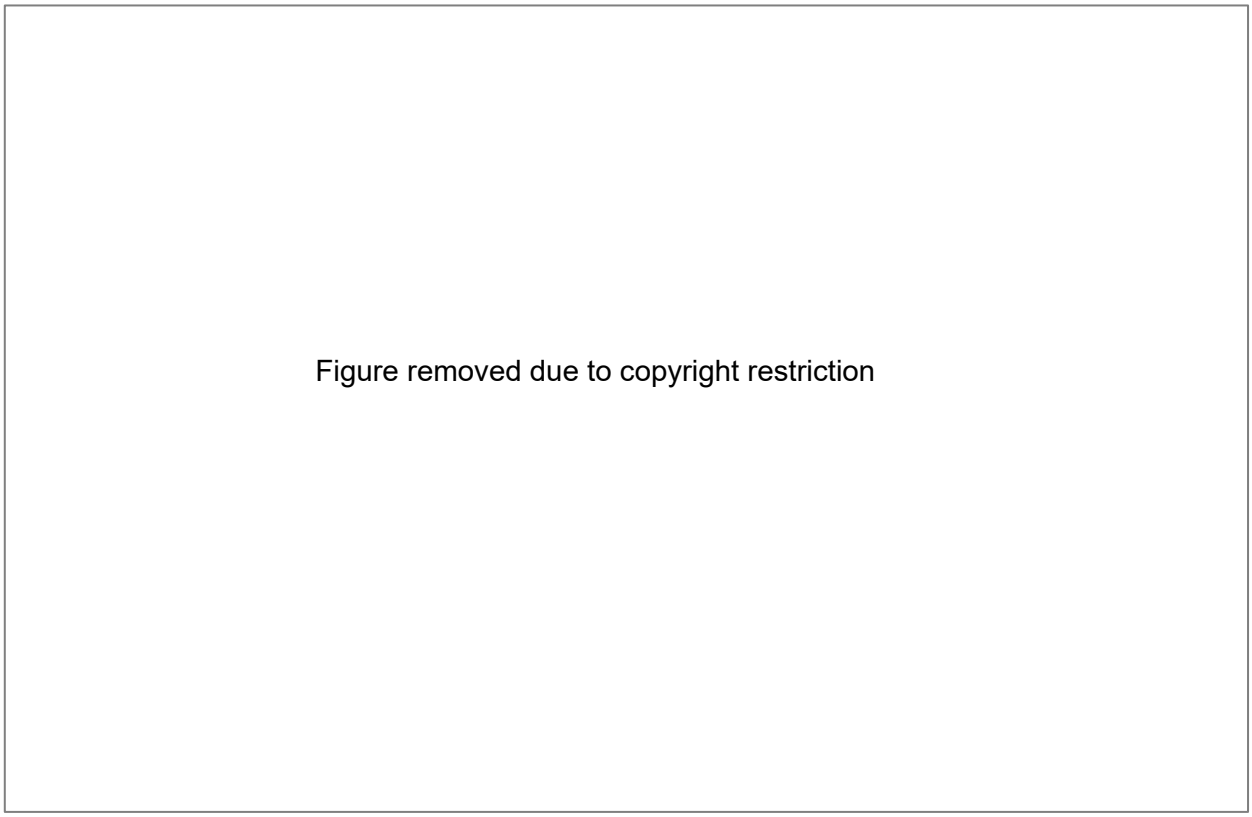

Figure 2.2 Location of Rio Tinto mine operations (Rio Tinto Iron Ore, 2010)

2.4.1 Climate

The Pilbara region is characterised by a semi-arid tropical climate resulting from tropical maritime and tropical continental air masses, receiving summer rainfall. Cyclones can occur during this period, bringing heavy rain that potentially destroys coastal and inland towns (BOM, 2021). The Pilbara region has a wide temperature range, with the maximum temperature rising to about 46 degrees Celsius (°C) during the summer and dropping to about 20°C in winter. Figure 2.3 indicates the average monthly temperatures between 1971 and 2021), and the annual historical rainfall is in Appendix 1 (BOM, 2021).

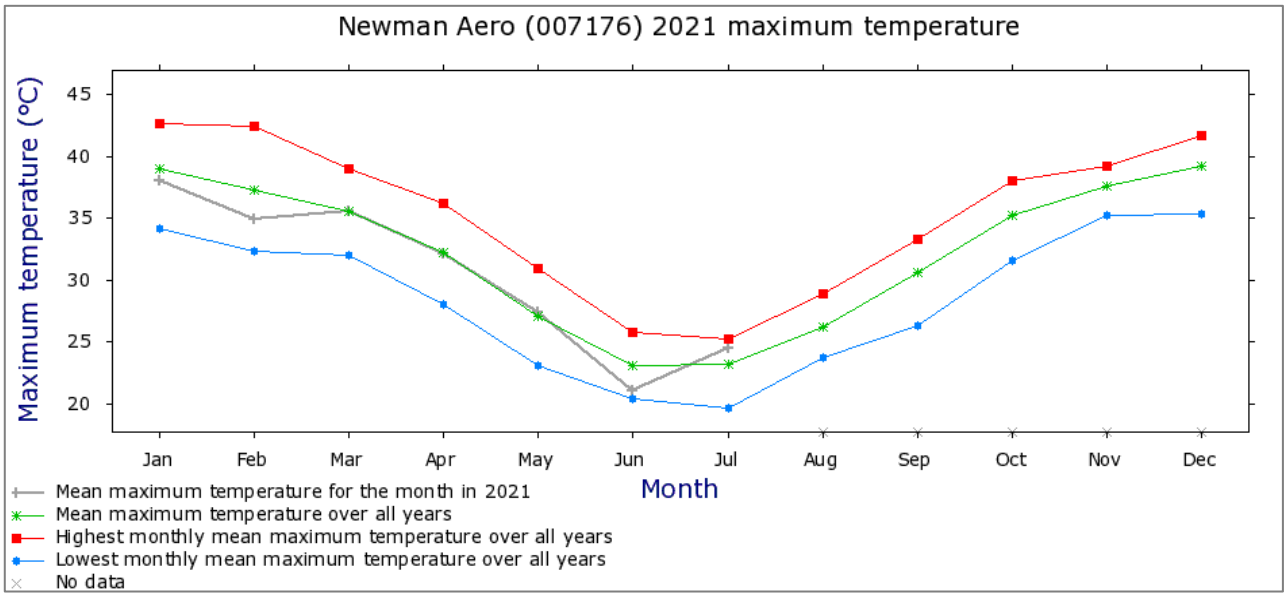

Figure 2.3 Newman Aero 1971 - 2021 Average Monthly Temperatures

2.4.2 Rainfall

The Pilbara region is grassland: hot (persistently dry), based on the modified Köppen classification system (BOM, 2021). Rainfall is typically associated with tropical low-pressure systems and thunderstorm activity from the summer's monsoonal troughs in northern Australia. Winters are typically dry, and mild through winter rain events can occur because of tropical cloud bands that intermittently affect the area. The mean monthly rainfall of 2018 to 2021 measured at Newman Aero varies between 27mm and 218 mm. The rainfall is also highly seasonal, with approximately 50% to 74% of the annual total occurring between December and April, and the wettest months are January and February (BOM, 2021).

Figure 2.4 indicates the average annual rainfall rates for Newman Aero from 2018 – 2021 (BOM, 2021).

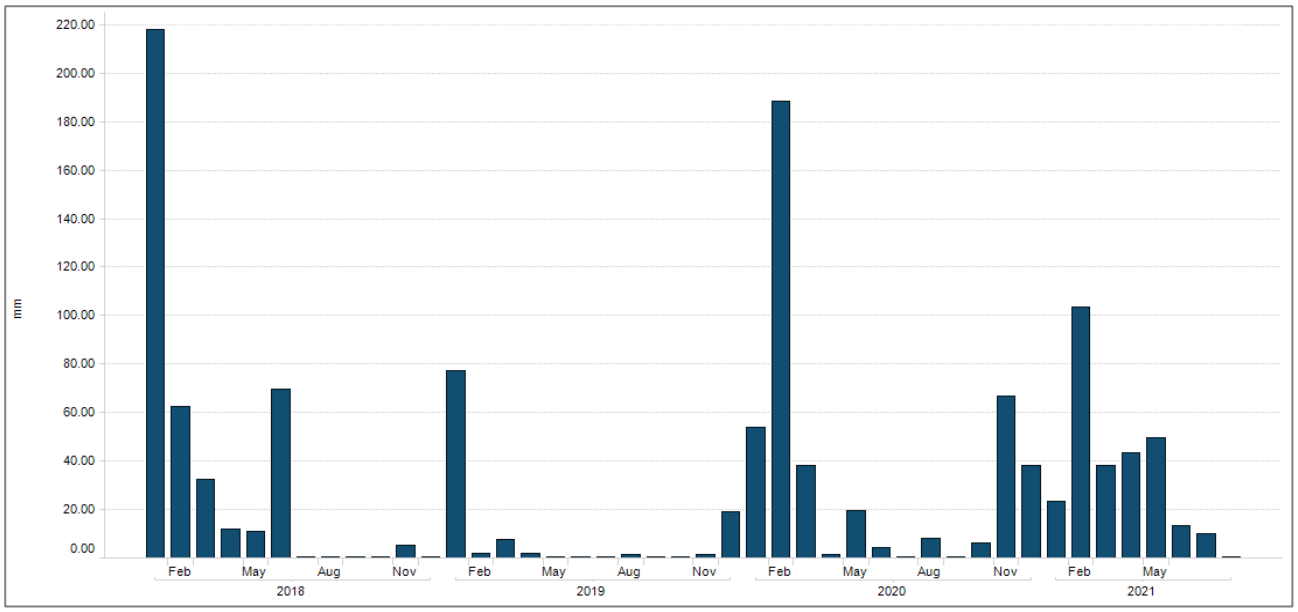

Figure 2.4 Neman Aero - Average Monthly Rainfall (mm) from 2018 – 2021

2.4.3 Regional geology and Hydrogeological Review

The Pilbara region is comprised of the depositional basin of Archaean to Lower Proterozoic rocks (Hamersley Group) which is underlaid by the Fortescue Group (Johnson & Wright, 2001). The Hamersley Group is divided into Brockman Iron Formation (BIF), Marra Mamba Iron Formation (MMIF), Boolgeeda Iron Formation, Woogarra Volcanic, Weeli Wolli Formation and more recent geology, the paleochannels and tertiary detrital units.

2.4.3.1 Brockman Iron Formation

The BIF is sub-divided into Yandicoogina Shale, Joffre Member, Whaleback Shale and Dales Gorge Member (Rio Tinto Iron Ore, 2010). The Yandicoogina Shale Member (~60 m thick) consists of interbedded chert and Shale, locally intruded by dolerite. The Joffre Member (~360 m thick) contains dominantly BIF units with minor thin Shale bands. BIF is sub-divided into six units (strands) named from the base upwards as J1 - J6. Strands J1, J3 and J5 contain more Shale than J2, J4 and J6. Then followed by the ~50m thick Whaleback Shale comprises four alternating macro-bands of Shale and BIF, whilst the upper zone consists of numerous alternating mesobands of chert and Shale. The Dales Gorge Member (~150 m) is an alternating assemblage of 17 BIF macro-bands (numbered DB0-16) and 16 shale macro-bands (DS1-16). These units persist throughout Hamersley Province (Rio Tinto Iron Ore, 2010).

The BIF is an important local aquifer, and the formation can be brittle, relatively resistant, and preserved as ridges that dominate the landscape (Dalstra, n.d.). The permeability in the BIF is typically associated with fractures and mineralised formation. In most cases, the ore body is the primary aquifer, and the weathered and fractured chert within the BIF form the local aquifer.

Figure 2.5 indicates the typical representative of the Brockman Iron Formation.

Figure 2.5 Typical examples of Banded Iron Formation (Hamersley Iron Pty Ltd, 2003)

Mount McRae Shale (MCS) Formation contains the Colonial Chert band, and BIF, Shale and dolomite dominate the lithology. The lithology that dominates the Mount Sylvia Formation (MTS) is Shale, Chert, Dolomite and BIF (Rio Tinto Iron Ore, 2010). The MCS and MTS are considered as low permeability on a regional scale.

2.4.3.2 Marra Mamba Iron Formation

The Marra Mamba Iron Formation is the principal ore-forming formation and is the oldest of the four Iron Formation in the Hamersley Group. The MMIF is subdivided into three Members: the Nammuldi, MacLeod, and Mount Newman (Rio Tinto Iron Ore, 2010).

The Nammuldi Member is the basal member of the Marra Mamba Iron Formation, which comprises Siliceous BIF, Chert, and interbedded Shale. Enrichment in the Nammuldi Member is rare and, when it occurs, is usually of low-grade Limonitic Goethite.

The MacLeod Member is the middle member consisting of BIF, Chert, Carbonate, and interbedded shales. The Macleod Member is frequently enriched to form low-grade Limonitic Goethite ore but contains a high proportion of Shale.

The Mount Newman Member is the upper member of the Marra Mamba Iron Formation and forms the main orebody. Shale bands separate the Newman into Upper (NE2) and Lower (NE1) strands. NE2 is further sub-divided into upper (N2U) and lower (N2L) units (Rio Tinto Iron Ore, 2010).

MMIF is an important local aquifer, especially the mineralised Mount Newman Members. The underlying lower Marra Mamba Iron Formation Members (MacLeod and Nammuldi Members).

Figure 2.6 indicates the representative structural geology setting for Marra Mamba Iron Formation.

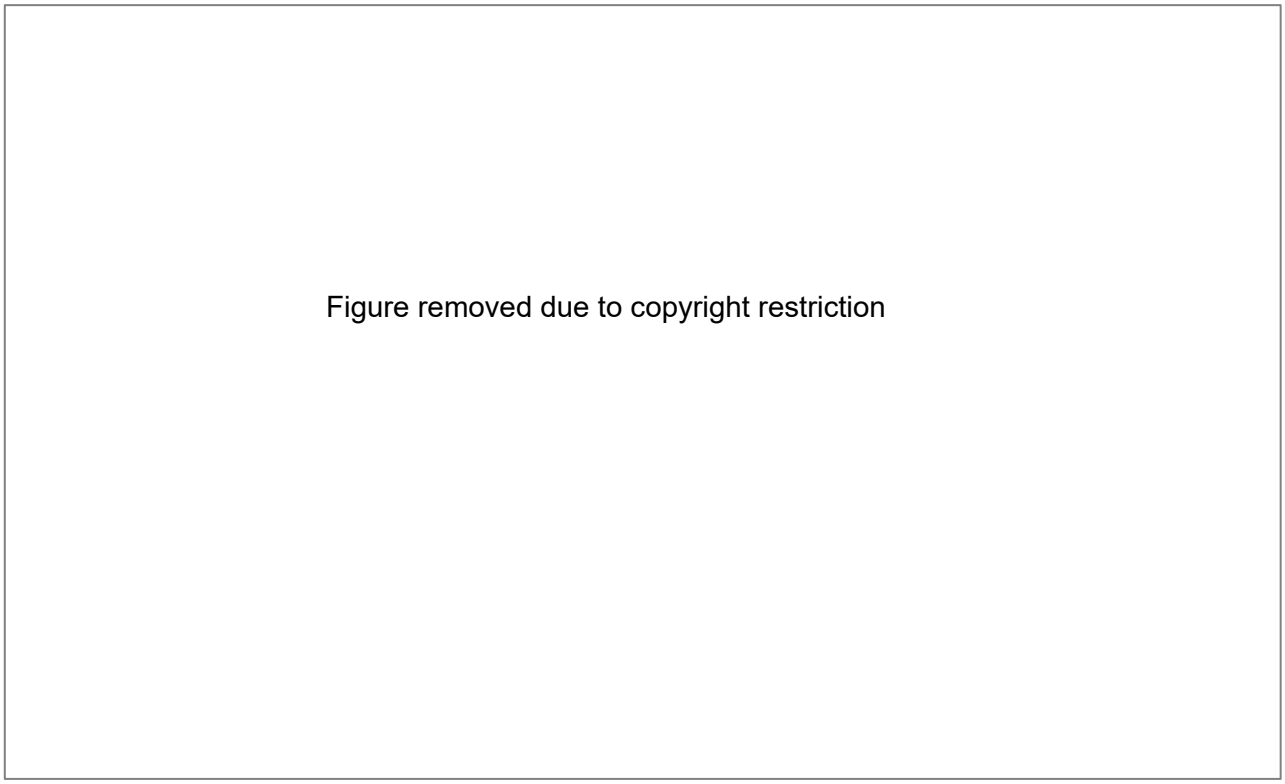

Figure 2.6 Typical of Marra Mamba Iron Formation (Hamersley Iron Pty Ltd, 2003)

2.4.3.3 Channel Iron Deposit

The Channel Iron Deposits setting occupies meandering paleochannels incised into Hamersley Basin. Channel iron deposits are characterised by fossilised wood fragments and the virtual absence of lithic rock fragments within the main ore zone (Morris, 1994). Paleochannel aquifers are highly porous, vuggy, and heterogeneous and behave similarly to a fractured rock aquifer (Dogramaci, 2009). The typical representation of the Channel Iron Deposit is as shown in Figure 2.7.

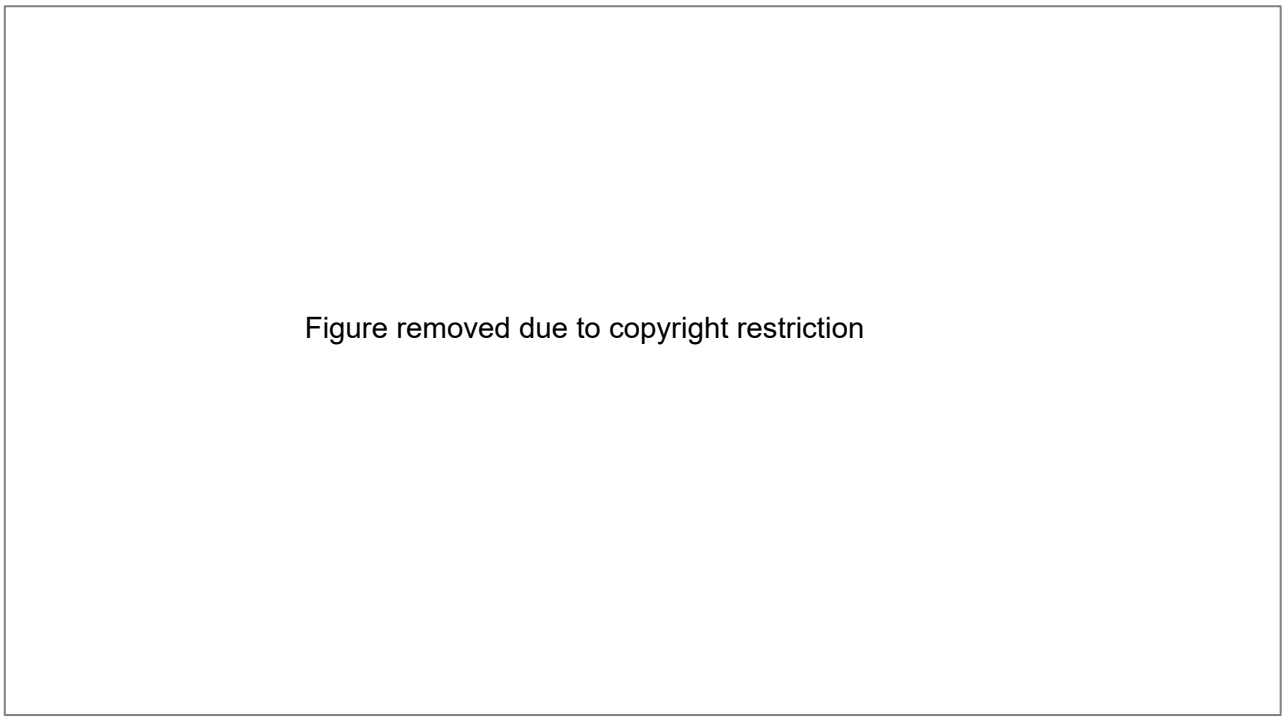

Figure 2.7 Typical examples of Channel Iron Deposit (Hamersley Iron Pty Ltd, 2003)

2.4.3.4 Tertiary & Quaternary settings

The Alluvial and tertiary detrital aquifers are sediment aquifers with high permeability depending on the matrix of sediments. The thickness of the alluvium varies between 2 and about 10m or more below ground level, and other detrital units are approximately 40m. These valley fill materials overlying the orebody or basement units consist of older lower detrital sediments along the present-day drainage and flood plain. The lower detrital unit invariably contains considerable proportions of sand and gravel with distinct pisolite pebbles and cobbles in a calcareous clay matrix (Rio Tinto Iron Ore, 2010).

Figure 2.8 indicates the tertiary and quaternary sediments (valley fill).

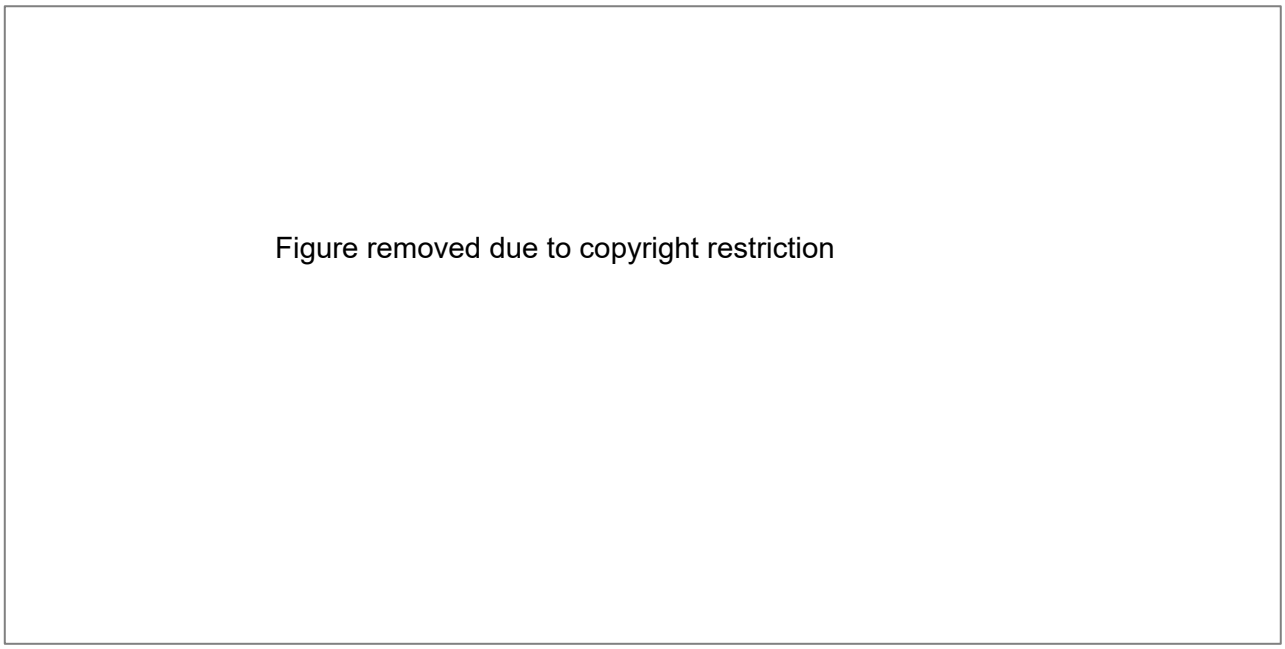

Figure 2.8 Typical example of Tertiary Detrital units (Hamersley Iron Pty Ltd, 2003)

2.4.3.5 Wittenoom Formation (WF)

In some literature, the Wittenoom Formation or Wittenoom Dolomite (WD). WF is subdivided into three members. The Wittenoom Formation is a dolomitic aquifer considered the primary groundwater supply target. Karstic features are because of the dolomite's susceptibility to chemical dissolution via percolation of surface water (Rio Tinto Iron Ore, 2014). Figures 2.4 and 2.6 indicate the Wittenoom Dolomite (WD) below the MTS within the stratigraphic sequence.

Figure 2.9 presents the Hamersley Group stratigraphic column (Rio Tinto Iron Ore, 2014).

Figure removed due to copyright restriction

Figure 2.9 Hamersley Group Stratigraphy (Rio Tinto Iron Ore, 2010)

Table 2.1 summarises the geology, groundwater potential commentaries on aquifers and aquitards. Information has been produced based on hydrogeological conceptual knowledge, supported by Rio Tinto Iron Ore (2014) and Johnson & Wright (2001).

The schematic cross-section of all aquifers in the Pilbara region is as indicated in Figure 10.

Table 2.1 Stratigraphic unit & hydrogeological description (Rio Tinto Iron Ore, 2010 & Johnson & Wright, 2001)

Age	Group	Stratigraphic Formation	Strands (i.e. Members)	Tags	Typical Lithologies	Hydrogeological Characteristics	Comments
Cainozoic							
Archaean to Early - Proterozoic							

Text removed due to copyright restriction

Figure removed due to copyright restriction

Figure 2.10 Schematic of major aquifers in the Pilbara (Rio Tinto Iron Ore, 2014)

CHAPTER 3 RESEARCH METHODOLOGY

3.1 Data aggregation

The dataset was sourced from various reports, internal Rio Tinto databases such as acQuire and EnviroSys, owned by acQuire Technology Solutions Pty Ltd (acQuire Technology Solutions, 2021). The reports and excel spreadsheets were other sources used during the data aggregation. The data aggregated from these sources include over 1000 aquifer tests. After excluding the data with many missing details or values (i.e., aquifer name, hydraulic conductivity values and screen depth) were not included to minimise biases and representation of the dataset. However, in some cases, the data with minimal missing values were included to draw insight from the data.

3.2 Data Cleaning

The data was then transformed into variable names and observations using the R programming language before importing and undertaking statistical / probability distributions. The R programming can import any text file, and then it is easy to use tidyverse packages (tidyverse is a collection of tidy-model packages) that can provide a consistent way to organise the data.

R programming software version 4.1.1 is a free environment developed for statistical computing (R Core Team, 2019). R programming is a user-friendly, large community, and the code is easy to follow for non-programmers. R software was written and executed using RStudio version 1.4 as an integrated development environment (IDE) for writing and running the code. R studio interface supports direct code tools for plotting, debugging, and easier workspace management (R Core Team, 2019). Another advantage of using RStudio IDE is that it is user-friendly for non-programmers to import libraries compared to packages in Python (R Core Team, 2019).

3.3 Data exploration analysis

This study involved data analysis and visualisation of statistical and probability distributions of the hydraulic dataset by deposit, aquifer type, and site. Notice that various graphical outputs of the hydraulic dataset are presented in the data analysis section. These statistical descriptions or graphical presentations include PDFs, QQ-plots, histograms, statistical analysis, and probability distributions.

The R-programming codes used to generate histograms, distributions, boxplot, QQ-plot, and other statistical analysis output are provided in Appendix 2. The graphical results of either histogram or QQ-plots are excellent at visualising the variation or covariation of aquifer type and by deposit or site. Histograms, box plots, and QQ-plots are the most informative visualising a univariate distribution (de Smith, 2015). In this study, multiple histograms were generated for aquifer types and locations. In addition to histograms, boxplots were also presented as detailed in the data analysis section. The

reason for using boxplots and QQ-plot is that it shows the core characteristics of the distribution of values within the dataset.

CHAPTER 4 DATA ANALYSIS

4.1 Statistical Data Analysis

This section details the statistical data analysis was undertaken using the R-programming package. The R-programming codes used for the data analysis are provided in Appendix 2, and the figures are presented in subsequent sections.

4.1.1 Summary statistics for Hydraulic Conductivity (K)

Table 4.1 presents the mean, maximum, minimum of hydraulic conductivity (K) of all the datasets; however, these results are likely to differ from the log-transformed data expected value.

Table 4.1 Summary statistics K (m/d)

K_mean	K_max	K_min	K_count
20.43	826.45	0.01	704

4.1.2 Histogram distribution of all K dataset

Figure 4.1 indicates the distribution of hydraulic conductivity on all the datasets.

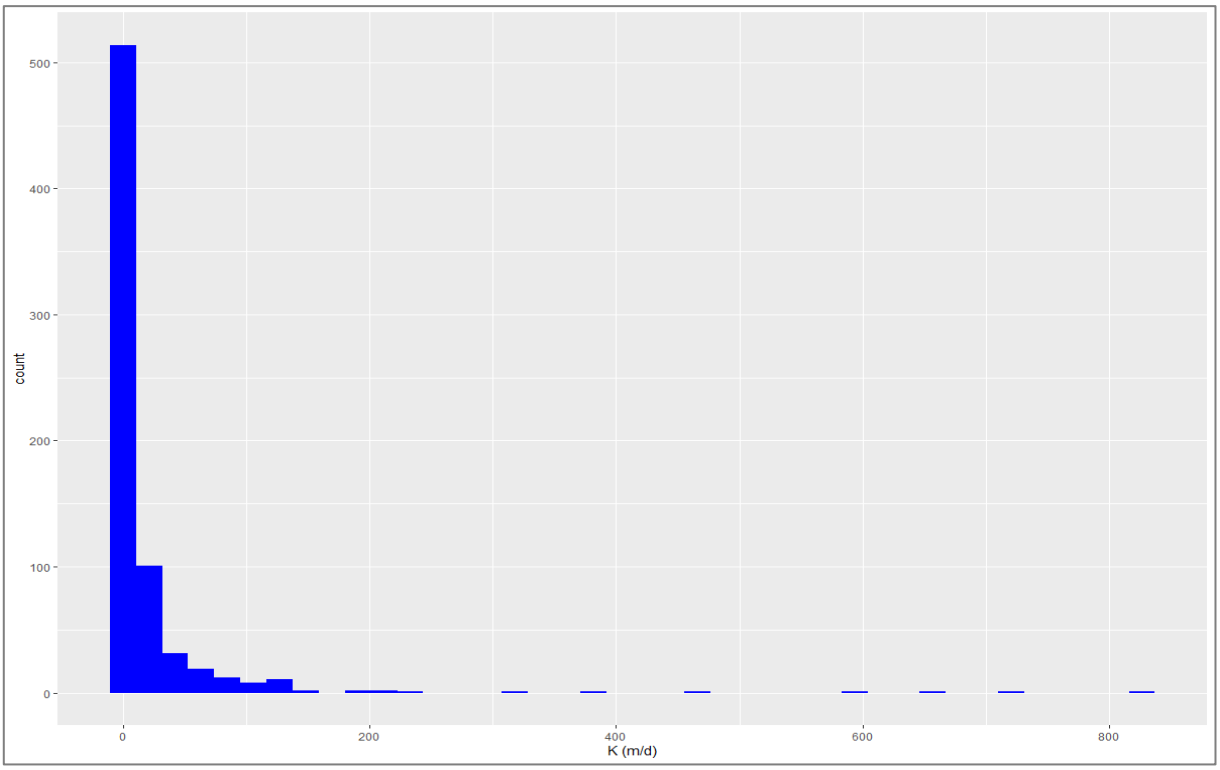

Figure 4.1 Histogram distribution of K (m/d) for all the datasets

Figure 4.1 shows that the distribution is log-normal for all the raw datasets without sub-dividing the parameters into either aquifer type or by location. Some minor datasets extend the x-axis to 800 (K

m/d). These nominal datasets on the x-axis were not removed to let the data speak for itself on the type of distribution. The hydraulic conductivity average value and the standard deviation of all datasets are about 20 m/d and 6.04 m/d, respectively.

4.1.3 Log transform and plot distribution

Figure 4.2 indicates that the distribution is log-normal, although outliers are on the distribution tails. The data were transformed logarithmically to reduce its skewness. If hydraulic conductivity is log-normally distributed, this should look like a normal distribution.

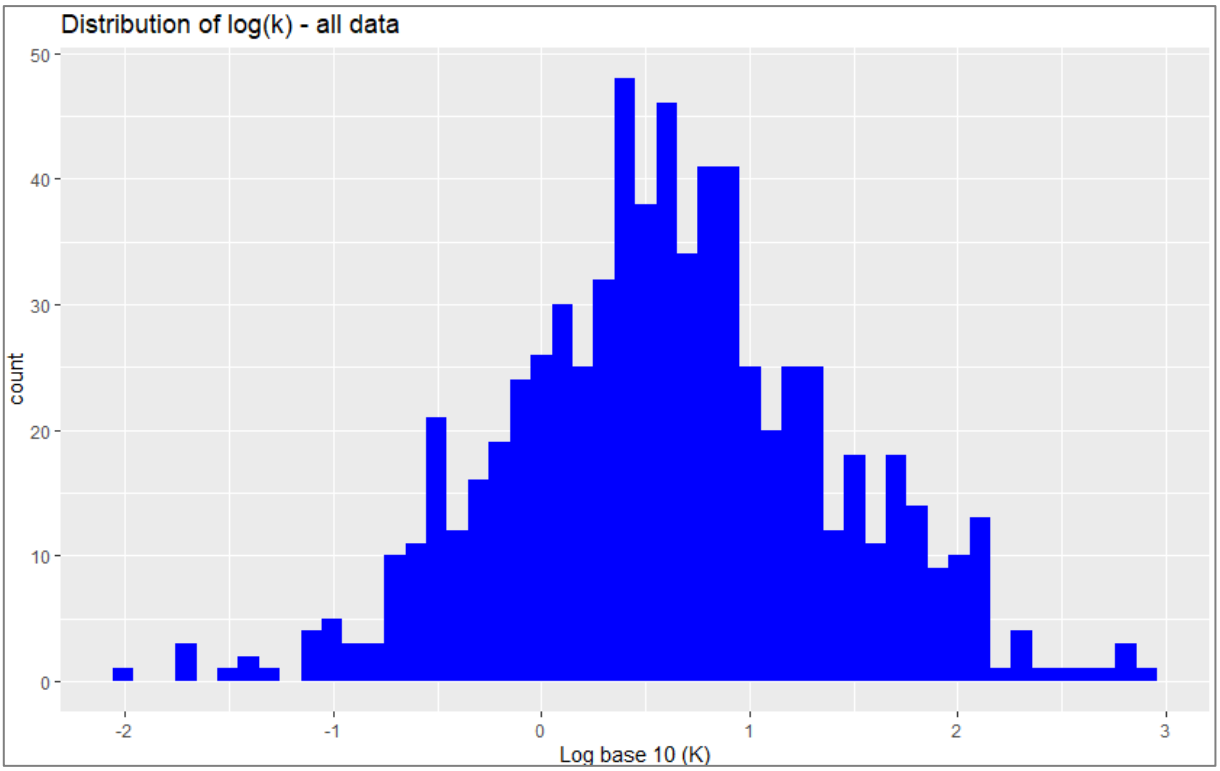

Figure 4.2 Distribution of log base 10 hydraulic conductivity for all the datasets

4.1.4 QQ plot for the dataset

Figure 4.3 shows the normal sample quartile versus theoretical quantile plot for the hydraulic conductivity for all the raw data sets.

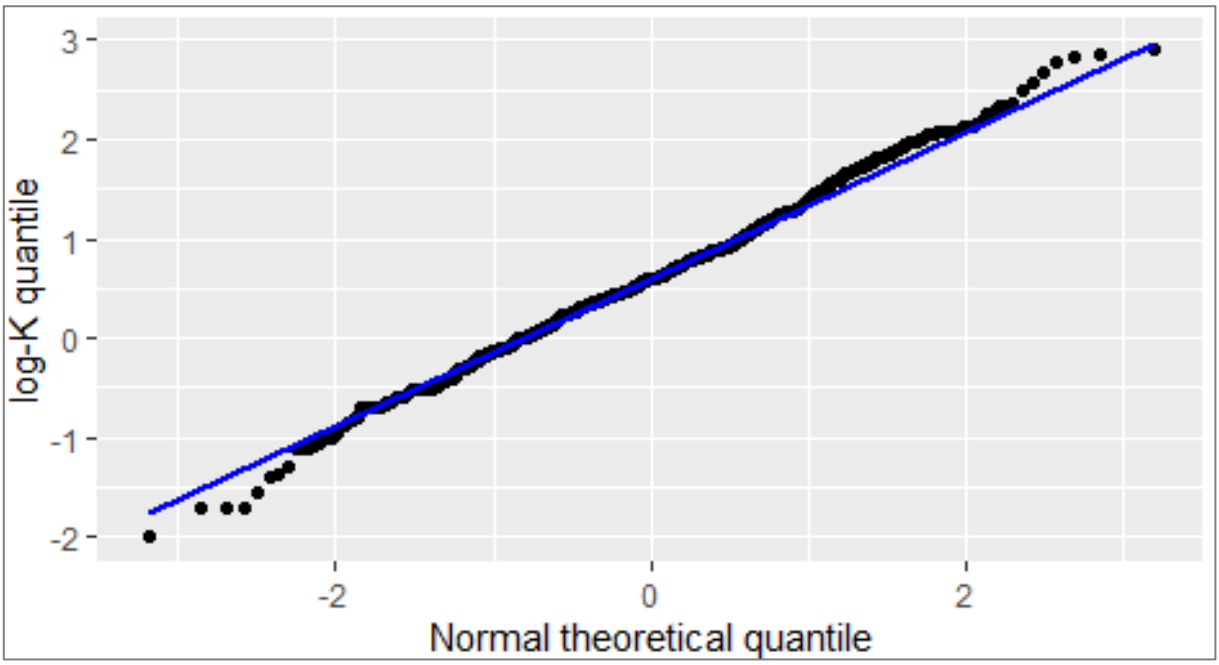

Figure 4.3 QQ plot of log-K for all datasets

Figure 4.3 indicates that the data conforms well to the theoretical distribution because most of the hydraulic conductivity dataset lies 45-degree to the blue line; however, there is a relatively small number of data points that fall in the few highest and few lowest quantiles (i.e., greater than log10 (K) 2 and less than log 10 (K) -2). Arguably, from the statistical point of view, the log (K) dataset is not precisely normal because of the data points that fall in the highest and lowest said quantiles but overall fits very well on the blue line.

4.2 Distribution of log-K by screened thickness, pumping test duration and discharge rates

This section examines bore completion characteristics such as screened thickness, duration of pumping test and discharge rate using histogram, QQ, box plots and density plots.

4.2.1 Histogram distributions of screened thickness, pumping test duration and discharge rates

Figure 4.4 illustrates a histogram of screened bores with their thickness. The screen thickness for some of the bores is less than 80m, and some are about 200m or more.

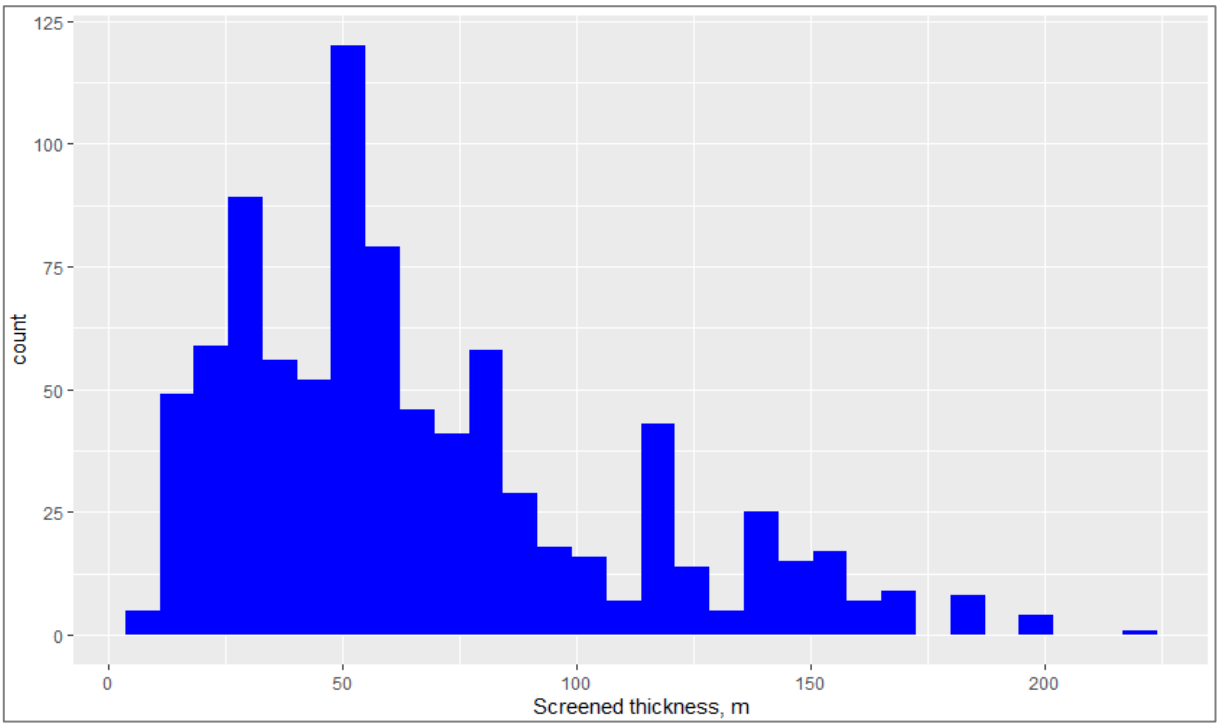

Figure 4.4 Histogram of the screened thickness (m) of bores

Figure 4.5 indicates the discharge rate data for 919 observations. The dataset is more easily viewed by taking the log transform as data spread out over several orders of magnitude (5 minutes to 120 000 minutes).

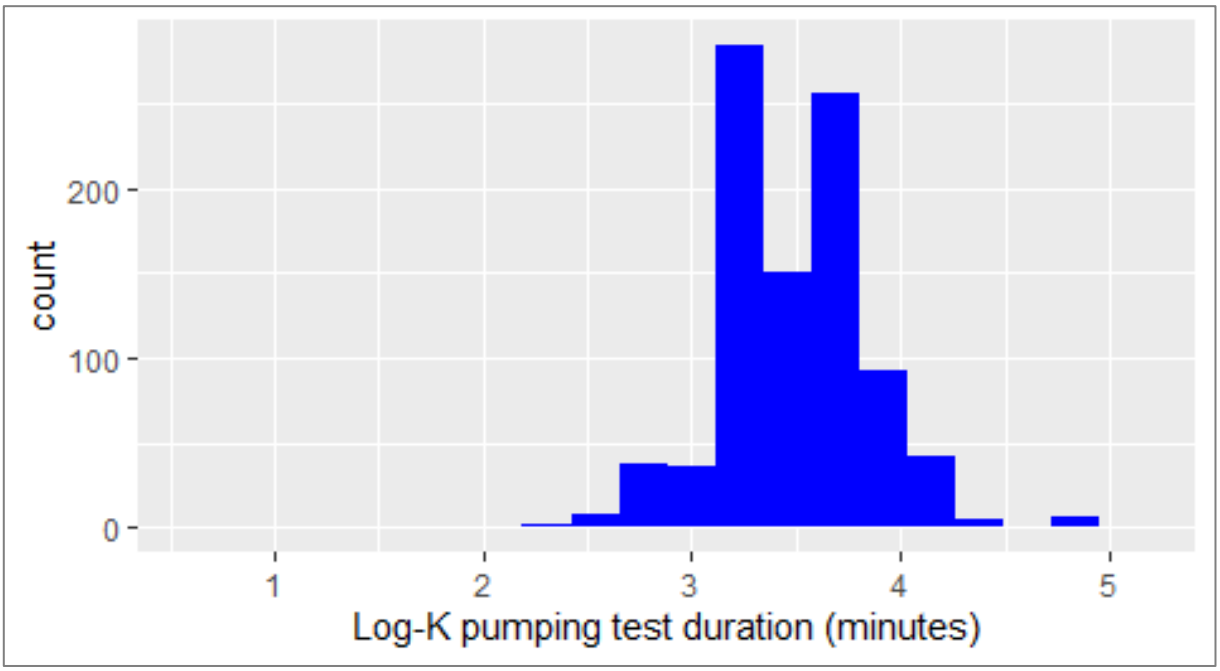

Figure 4.5 Histogram of log-K of pumping test duration in minutes

The discharge was also more easily viewed by taking the log transform as the data is spread out over several orders of magnitude (< 500 kL/day - 120,000 kL/day). Figure 4.6 indicates the histogram of discharge rate (kL/day).

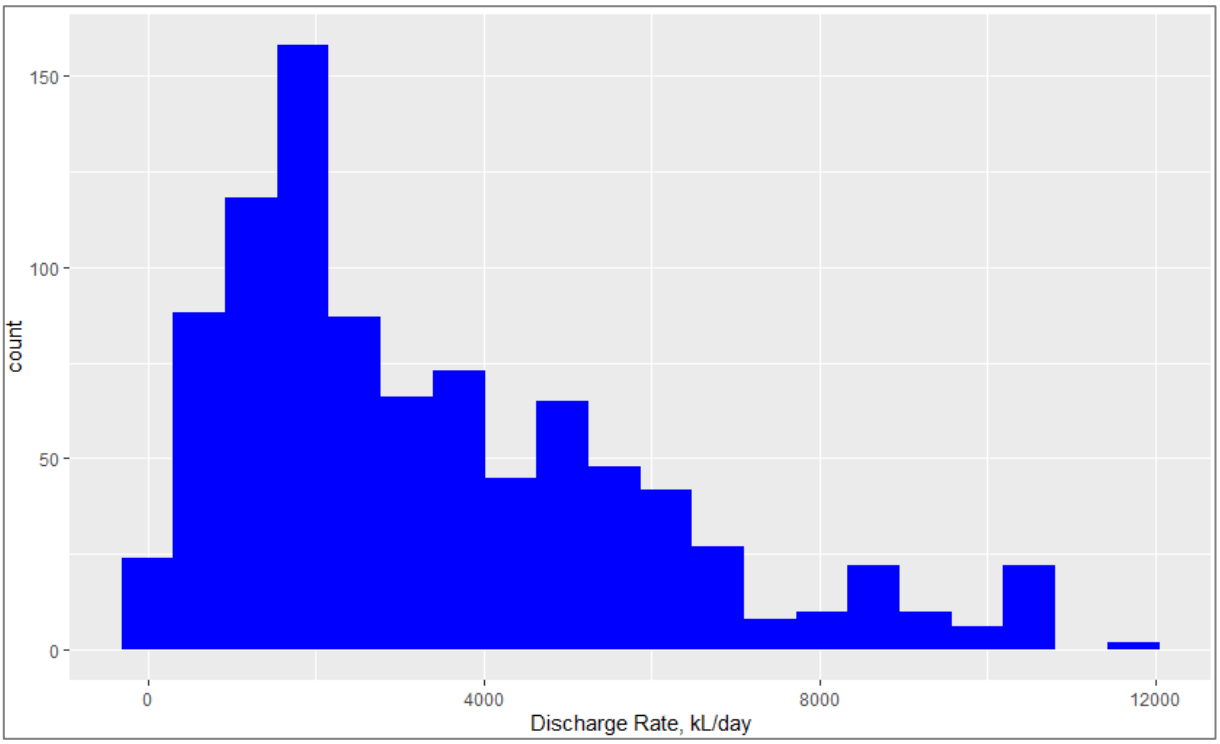

Figure 4.6 Discharge rate (kL/day)

4.2.2 QQ plots of log-K conditioned on screened thickness, pumping test duration and discharge rates

This subsequent section indicates the QQ plots of log (K) conditioned on screened thickness, pumping test duration and discharge rates. Figure 4.7 shows the QQ plot of log-transformed values conditioned on the screened thickness.

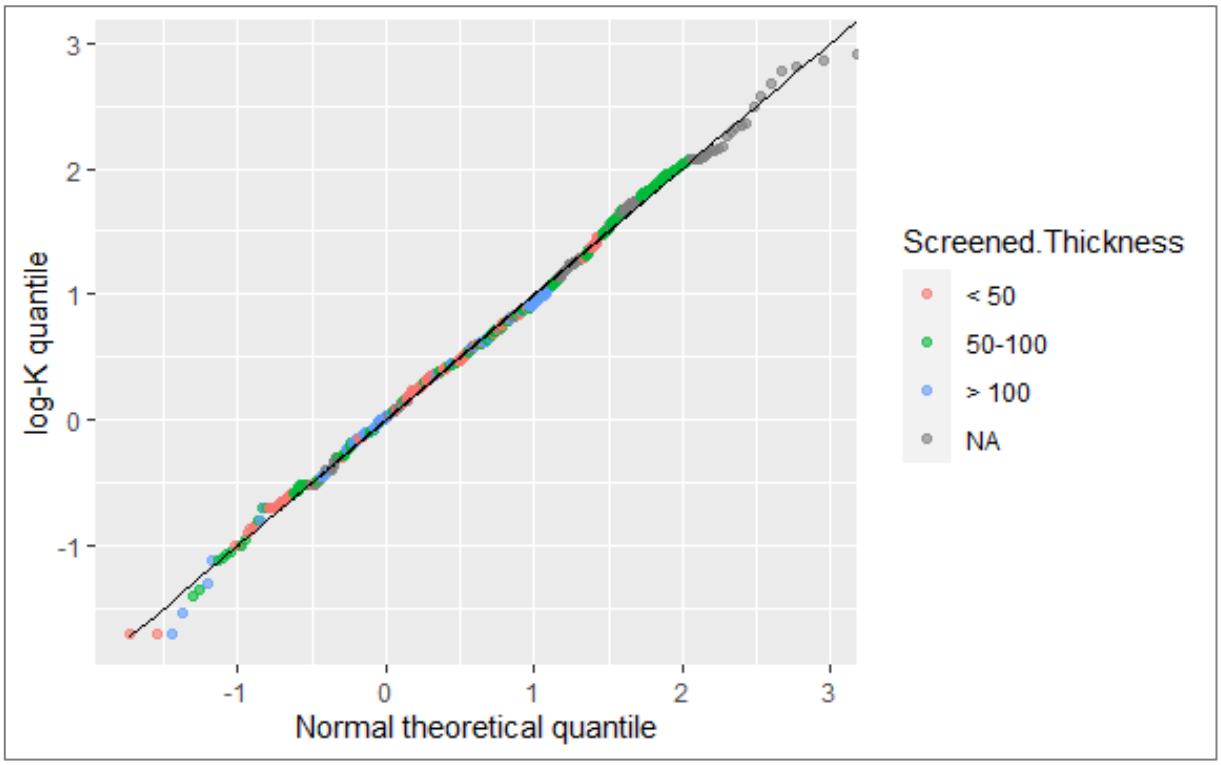

Figure 4.7 QQ-plot of log-transformed (K) values by screened thickness (m)

The QQ plot indicates the distribution match up along a straight line; though, a small number of data points fall in the few highest and few lowest quantiles due to random fluctuations at the extreme ends. Notice that the "NA" represents the data points with missing hydraulic conductivity values; thus, the data points were not conditioned on bore screened thickness.

Figure 4.8 indicates the QQ plot of log-transformed (K) conditioned by pumping test duration (hrs). The figure suggests that the pumping test duration points match along a straight line indicating that the quantiles match the base distribution.

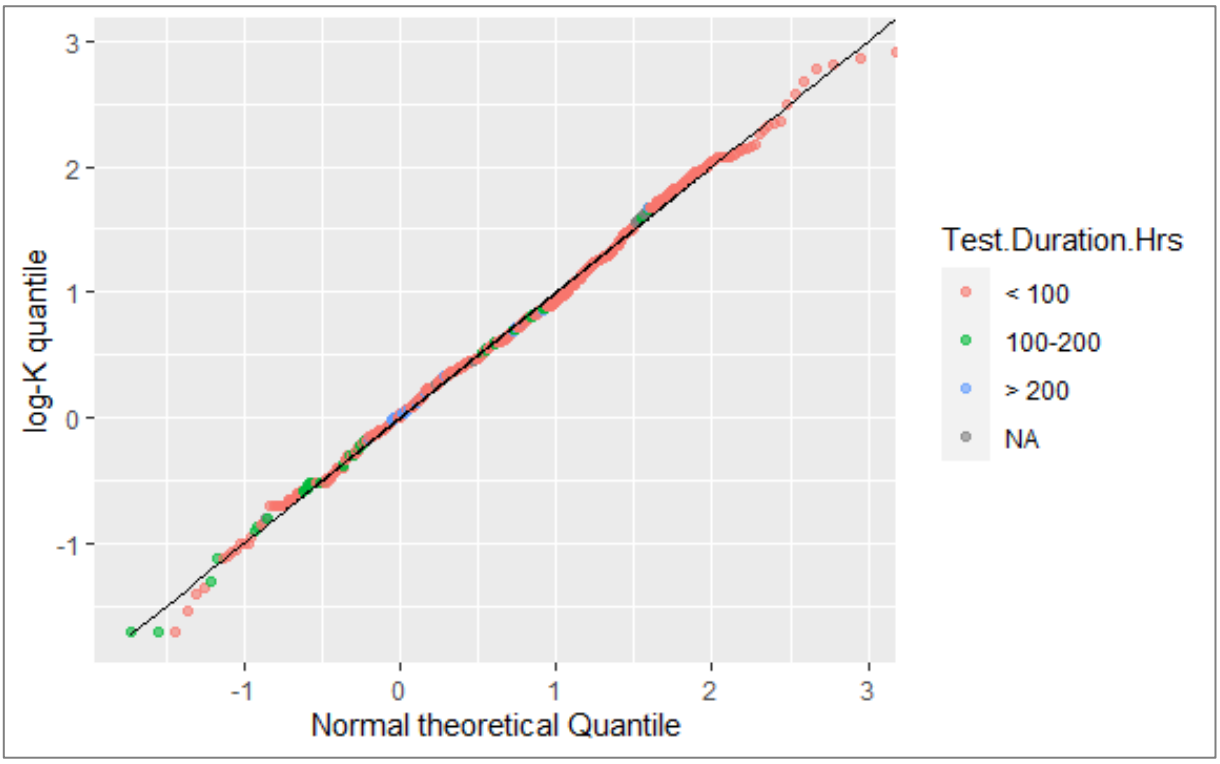

Figure 4.8 QQ-plot of log-transformed (K) conditioned on test duration (hrs)

Figure 4.9 shows the Q-Q plot of log-transformation (K) conditioned on discharge rate (kL/d). The figure indicates that the discharge rate points match a straight line that shows the quantiles match the base distribution. Notice that the “NA" data points had missing log (K) values to condition the test duration; the test duration or observation values were present.

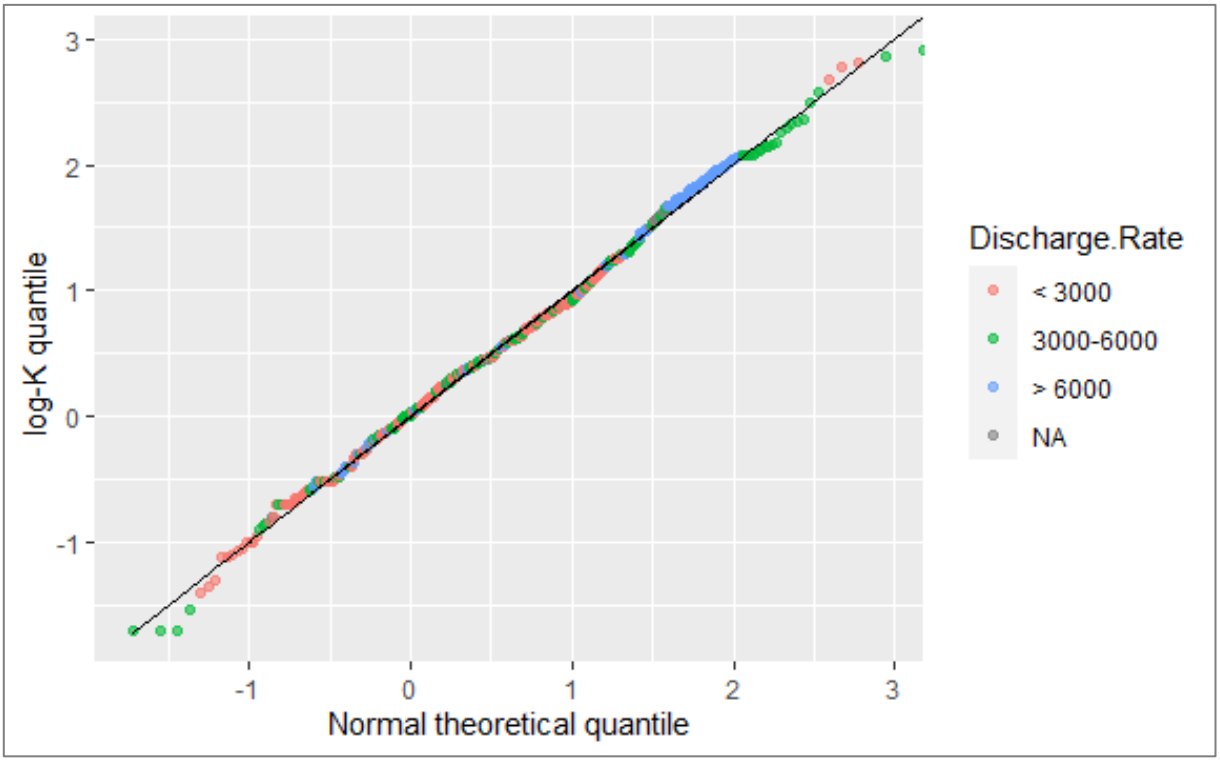

Figure 4.9 QQ-plot of log-transformed (K) values conditioned on discharge rate (kL/d)

4.2.3 Box plots of log-K by screened thickness, test durations and discharge rates

It is also informative to view distributions of log (K) by the bore completion characteristics (e.g., screened thickness, discharge rate and pumping test duration). The boxplots were used to show the distribution of data points. Figure 4.10 indicates the box plot of the log-transformed of bore screened thickness log(K) values.

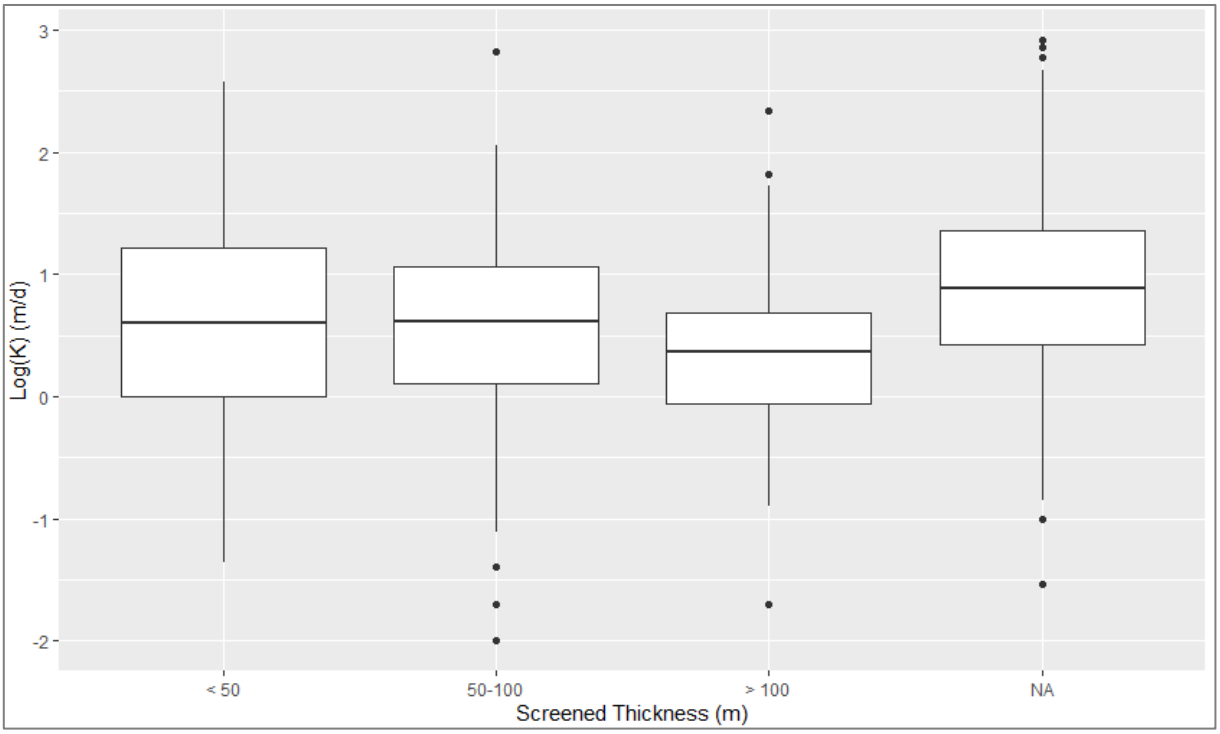

Figure 4.10 Box plot of log-transformed (K) conditioned on screened thickness (m)

From the box plot, the log (K) conditioned on screened thickness indicates that the:

- Screen thickness of < 50m is normally distributed because the plot is symmetrical.
- Screen thickness of 50m to 100m slightly negatively skewed
- Screen thickness of > 100m is negatively skewed
- Screen thickness indicating "NA" means there is erroneous or no data points for bore screen thickness. Though, the log (K) data points indicate no skewness.
- The box plot indicates that their outliers in the dataset

The differences in the distribution skewness could be associated with the averaged K values based on the screened length.

Figure 4.11 indicates the boxplot distribution of pumping test duration.

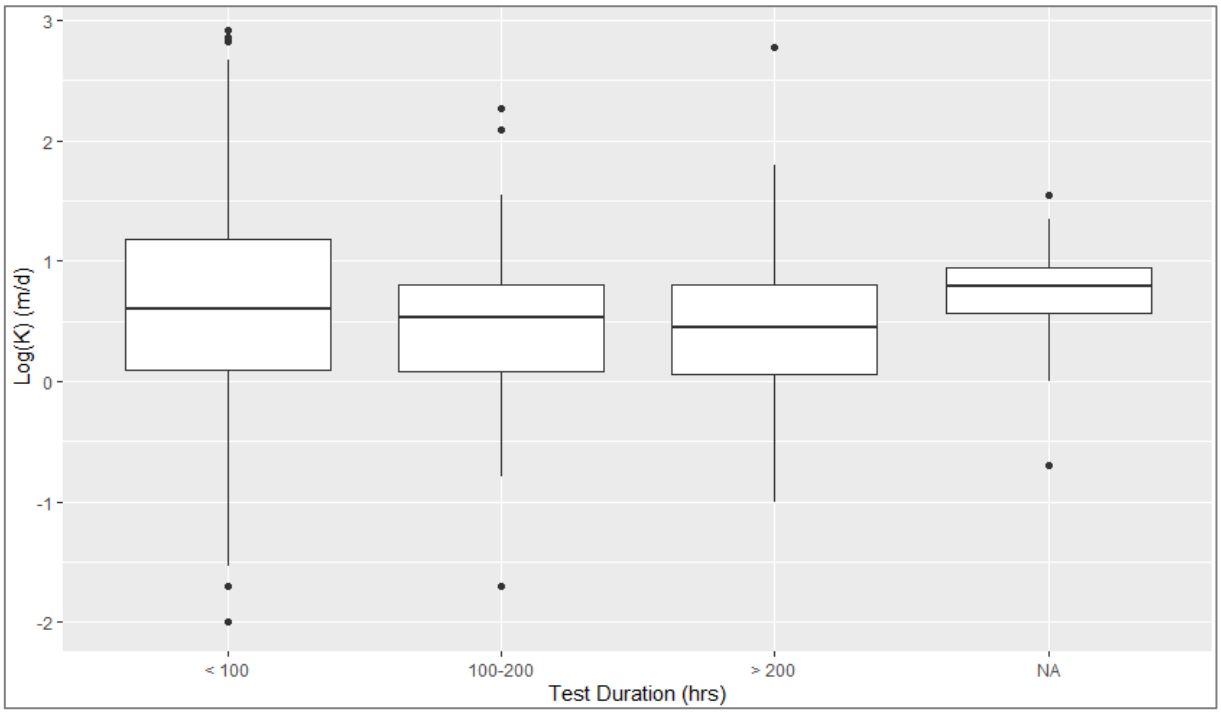

Figure 4.11 Distribution of log (K) by pumping test duration (hrs)

Figure 4.11 indicates that the pumping test duration < 100 hours with no skewness or normally distributed, while test duration of 100 to 200 hours shows left skewness. The pumping test duration of > 200 hours indicates that there is no skewness, while no test duration is available for "NA" but with available log (K) data points. The differences in skewness are attributed to test duration, the formation where the pumping test is undertaken. Figure 4.12 shows the boxplot distribution of log (K) data points conditioned on the discharged rate.

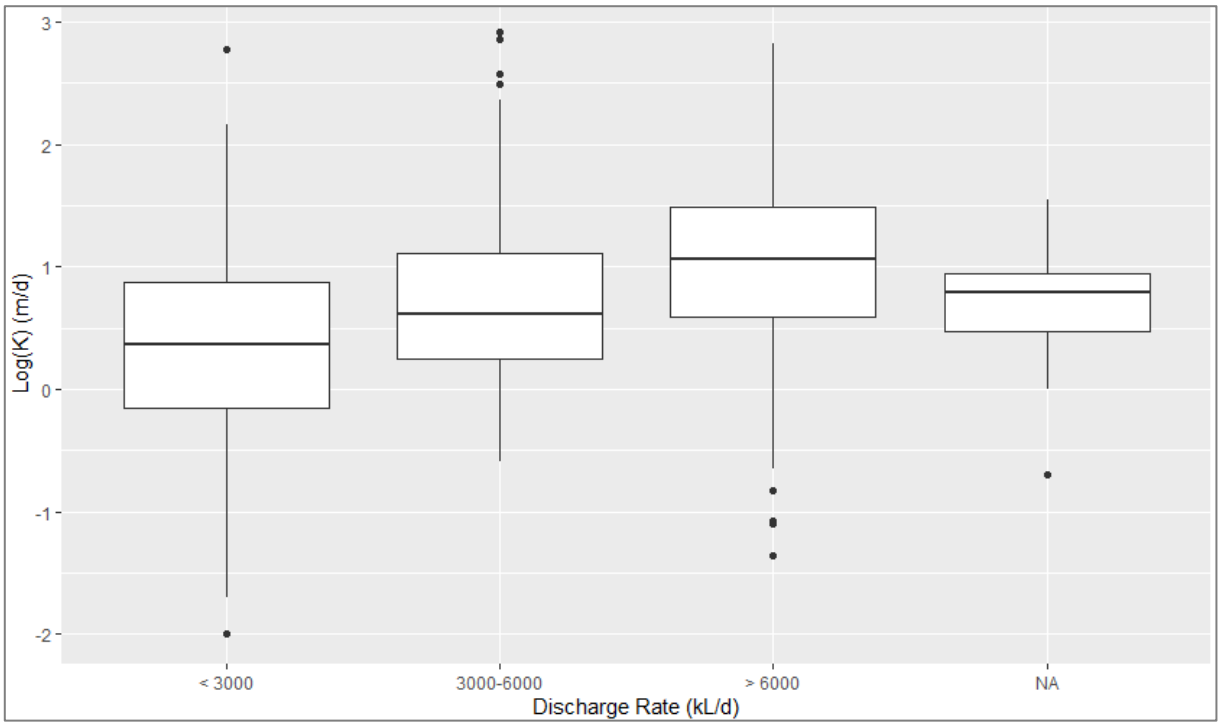

Figure 4.12 Boxplot of log-transformed (K) values on discharge rate (kL/d)

Figure 4.12 shows that the data points of log (K) when conditioned on discharge rates, it is observed that the:

- Discharge rate < 3000 kL/d is normally distributed
- Discharge rate 3000 to 6000 kL/d is positively (right) skewed
- Discharge rate > 6000 kL/d is slightly negatively (left) skewed and
- Discharge rate with "NA" means that log (K) data points were available with missing discharge rates; however, the distribution is normally distributed.

The difference in the distribution for discharge rates can be attributable to the hydraulic conductivity of the formation. It is also indicated that the lower K values have indicated the low discharge rate and higher values of K with a high discharge rate.

4.1.4 Density plots of log-K by screened thickness, test durations and discharge rates

This subsequent section will also compare the shape of the distribution of log (K) conditioned on various factors using the density plot. The latter normalises the frequency of log (K) values such that the y-axis of each distribution is on the same scale, highlighting differences in more subtle aspects of the distribution such as skewness and kurtosis (Figures 4.13 to 4.15)

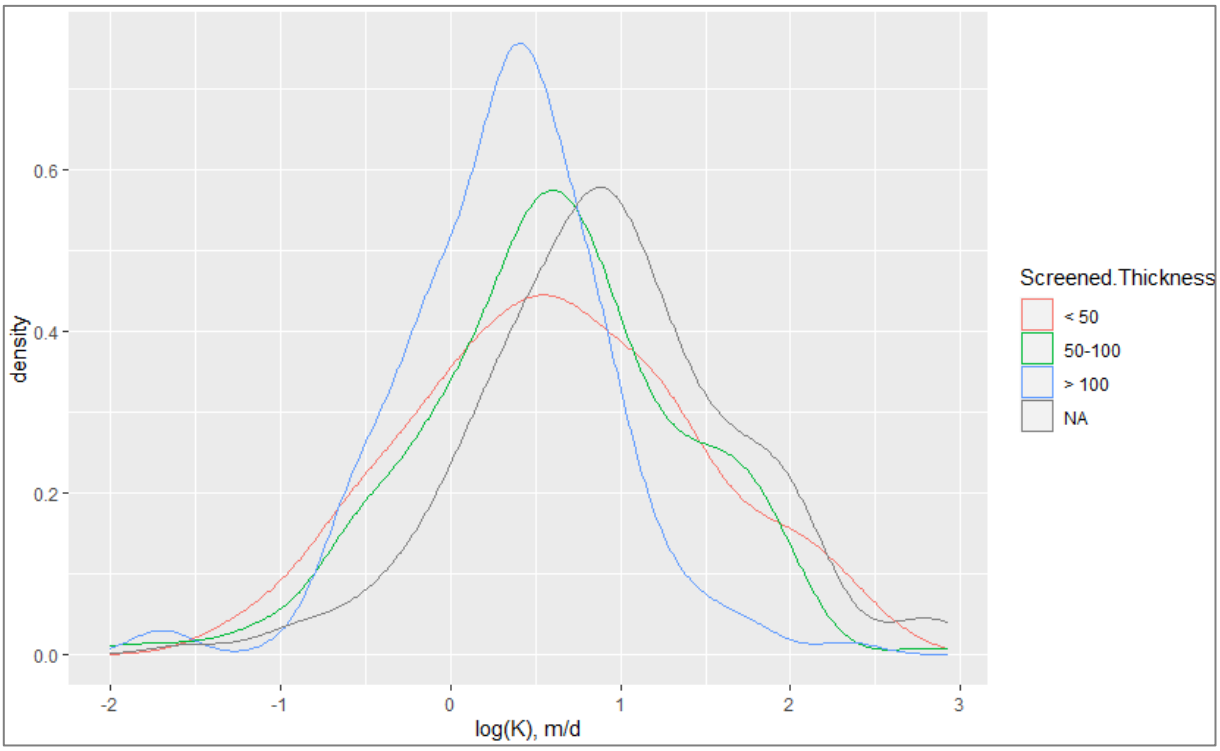

Figure 4.13 Density plot of log (K) conditioned on the screened thickness

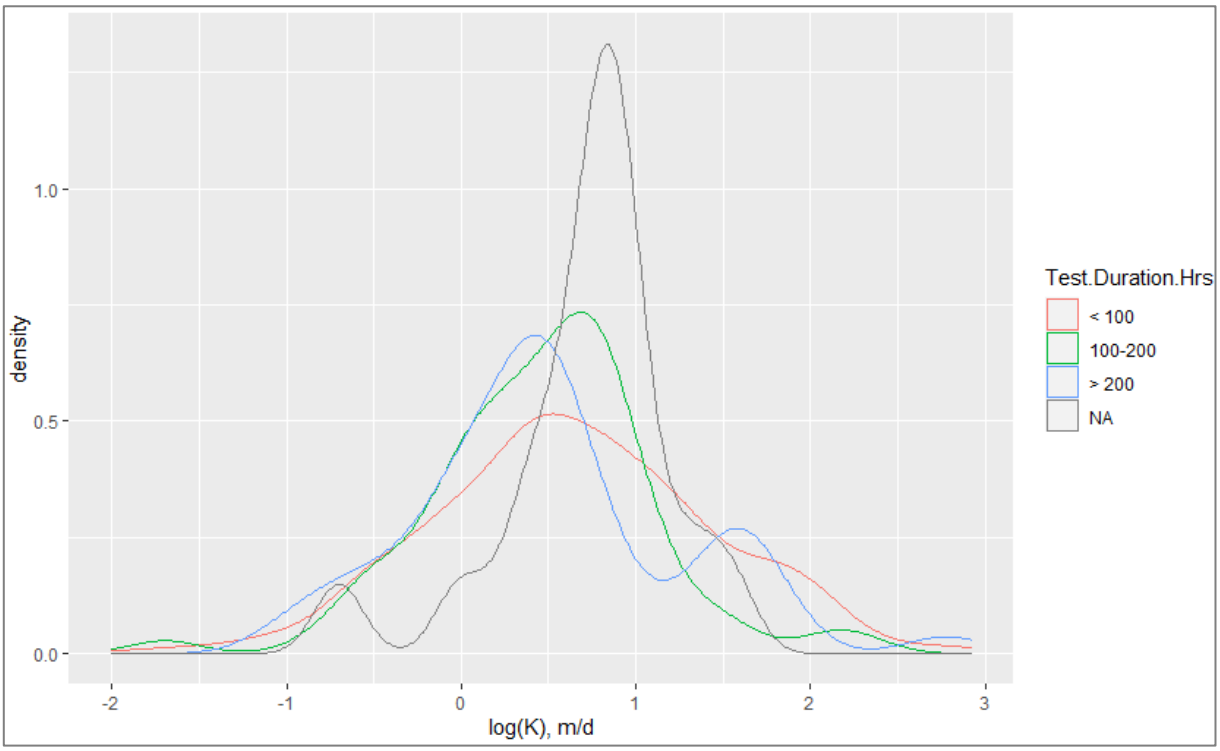

Figure 4.14 Density plot of log (K) conditioned on test duration

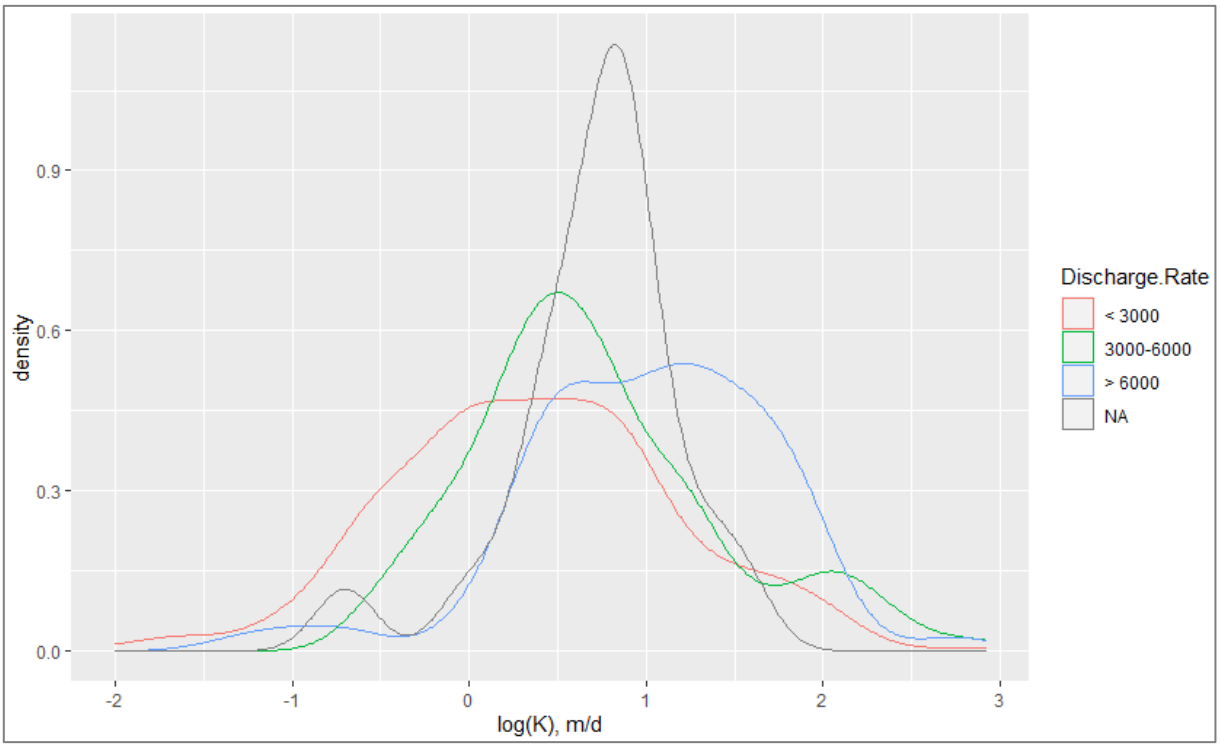

Figure 4.15 Density plot of log (K) conditioned on discharge rate

It is observed from Figure 4.15 the flatter and the broader distribution of log (K) for short test durations and consistent. Likewise, when the data is conditioned log(K) on discharge rate, it is observed that the distribution is generally right shifted for higher flow rates. The density plot distribution for low discharge rates is slightly left-skewed, while high discharge rates are right-skewed. On average, this

translates to lower hydraulic conductivity values when the discharge rate is low and higher hydraulic conductivity values when the discharge rate is high.

4.2 Distribution of log (K) by aquifer type

This section summarises the hydraulic conductivity relevant by aquifer type. Notice that certain groups are known to possess similar hydraulic properties when grouped by aquifer types:

- Low permeability shales, namely Bee Gorge, Mount Sylvia Formation and Mount McRae and Whaleback Shales
- Detrital units
- Robe Piezolite, Channel iron and Yarraloola Conglomerate

Table 4.2 indicates the statistical summary of hydraulic conductivity by aquifer code. The detailed list of aquifer codes is also presented in Appendix 3.

Table 4.2 Summary statistics of K by aquifer code

Aquifer Unit	Aquifer Code	Count	Mean log10_K	mean (K m/d)
Channel Iron Deposit	CID	283	1.034	10.8
Paraburdoo Karstic	Witt_K	168	0.511	3.25
Dale Gorge Mineralisation	DG_M	152	0.253	1.79
Marra Mamba Mineralisation	MM_M	97	0.869	7.4
Unknown	Unknown	69	0.84	6.92
Marra Mamba Unmineralisation	MM_UM	57	0.434	2.72
Dales Gorge Unmineralised	DG_UM	43	0.226	1.68
Newman Mineralised	MM_Mn	42	0.258	1.81
Detrital	Det	30	0.717	5.21
Low Phophrous Formation	LowP	23	0.078	1.2
West Angelas Mineralised	MM_Ma	22	-0.102	0.79
Calcrete/Silcrete	Calc	19	1.373	23.59
Alluvials	Alluv	11	0.137	1.37
Joffre Mineralised	J_M	11	-0.087	0.82
Fractured Rock	FracZ	8	0.251	1.78
Weeli Wolli Fresh	WW_F	6	-1.118	0.08
Dolerite Sill	Dol	5	-0.132	0.74
Turee Creek Group	TG	5	0.565	3.67
Wela Wolli Fresh	Witt_F	3	0.068	1.17

Table 4.2 average values are primarily within one order of magnitude (1 m/d to 10 m/d). The interquartile ranges of each aquifer type generally cover less than an order of magnitude. The CID and MM have noticeably larger interquartile ranges. CID, DG_M, LowP and Witt_K have very low permeability outliers.

Figures 4.16 and 4.17 indicates the boxplot and violin plots distribution of hydraulic conductivity by aquifer code.

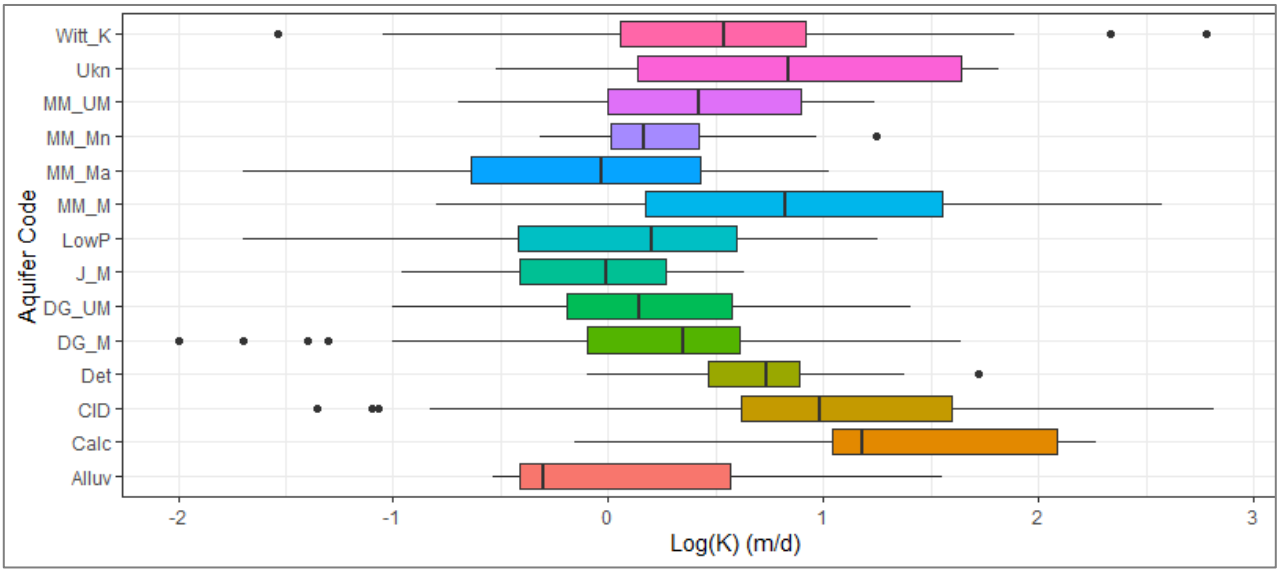

Figure 4.16 Boxplot distribution of log(K) by aquifer code

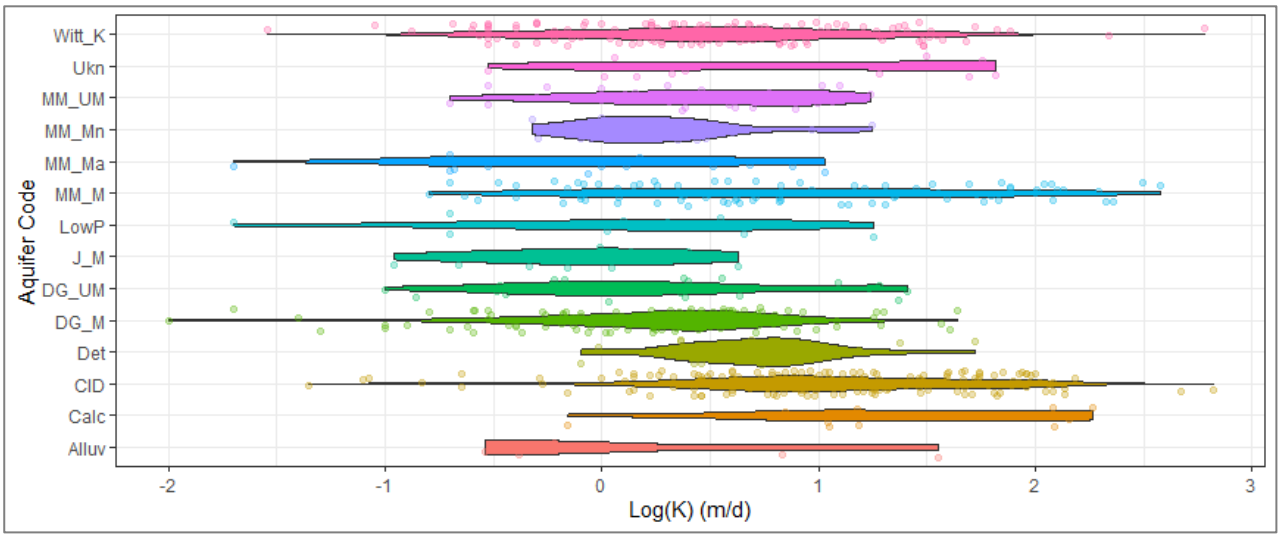

Figure 4.17 Violin plot of log(K) by aquifer code

Means are primarily within one order of magnitude (1 m/d to 10 m/d). Interquartile ranges of each aquifer type generally cover less than an order of magnitude. However, CID and MM have noticeably larger interquartile ranges. The CID, DG_M, LowP and Witt_K have very low permeability outliers. The plot can be deduced that the skinner the shape on each end represents a low probability of hydraulic conductivity, and the wider the middle indicates higher probability or data points concentration is around the median.

Figure 4.18 shows the aquifer code's histogram distribution of log (K) values.

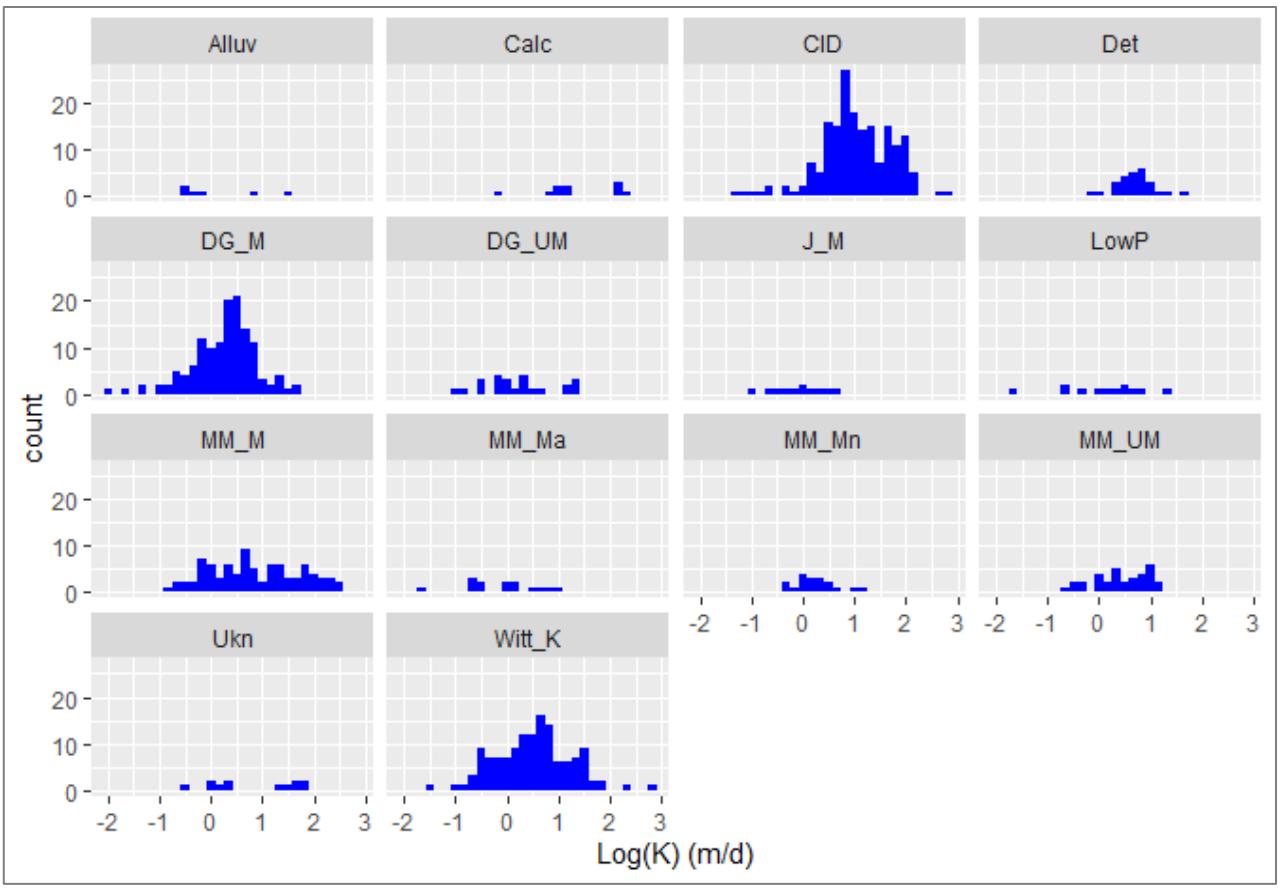

Figure 4.18 Histogram of log-transformed K values by aquifer code

Figure 4.18 indicates that a limited amount of data is broken down by aquifer type. However, it is noticeable that most aquifer types do not differ significantly in the aspects of their distributions visible here (range, mean). The bedded iron subgroups have distributions slightly skewed to the right.

4.2.1 Group by mineralised and un-mineralised formations

This section explores further banded iron groupings based on mineralised and un-mineralised formations or members. A complete list of aquifer codes used for this study is in Appendix 3 and Table 4.2.

- Mineralised bedded iron: DG_M, Witt_F, J_M, MM_M, MM_Mn, WW_M, MM_Ma
- Unmineralized bedded iron: DG_UM, J_F, MM_UM, WW_F, WBS

Figures 4.19 and 4.20 indicate the distribution of box plots and histograms for mineralised and un-mineralised formations.

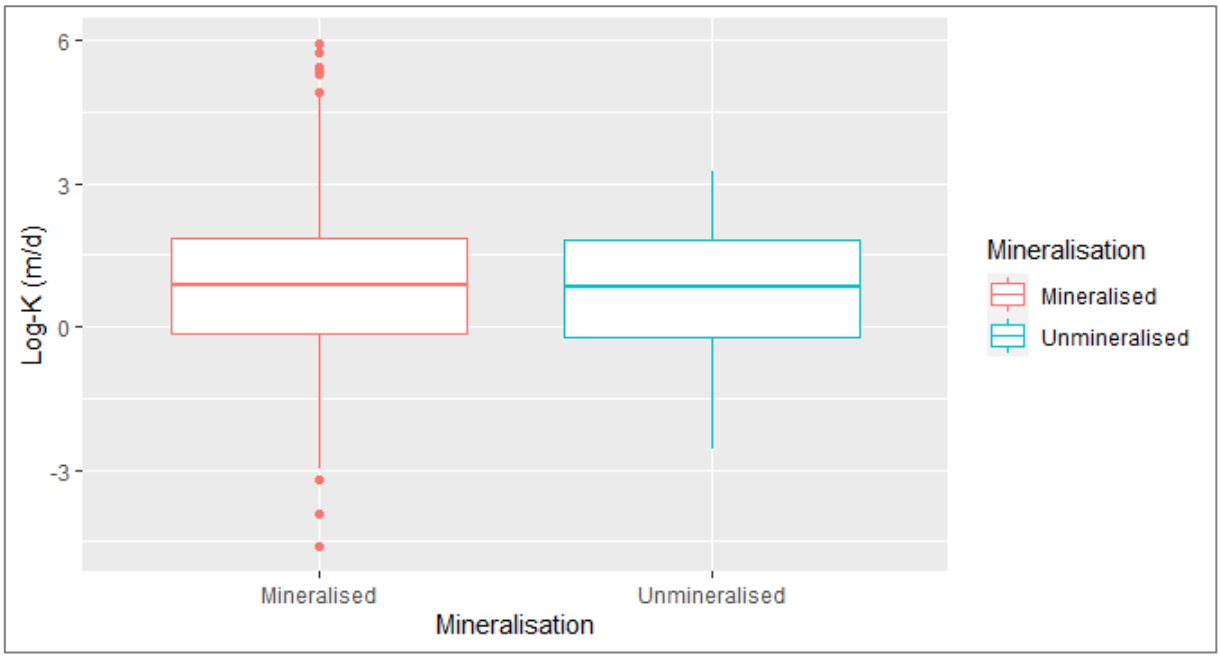

Figure 4.19 Box plot distribution of log(K) of BIF by mineralised & un-mineralised

The box plot indicates a normal distribution of mineralised and un-mineralised formations. However, there are outliers in the mineralised units compared to the unmineralised units due to aquifer spatial variability, and most bores are screened in the orebody.

Figure 4.20 indicates the distribution of hydraulic conductivity of the mineralised and un-mineralised formations.

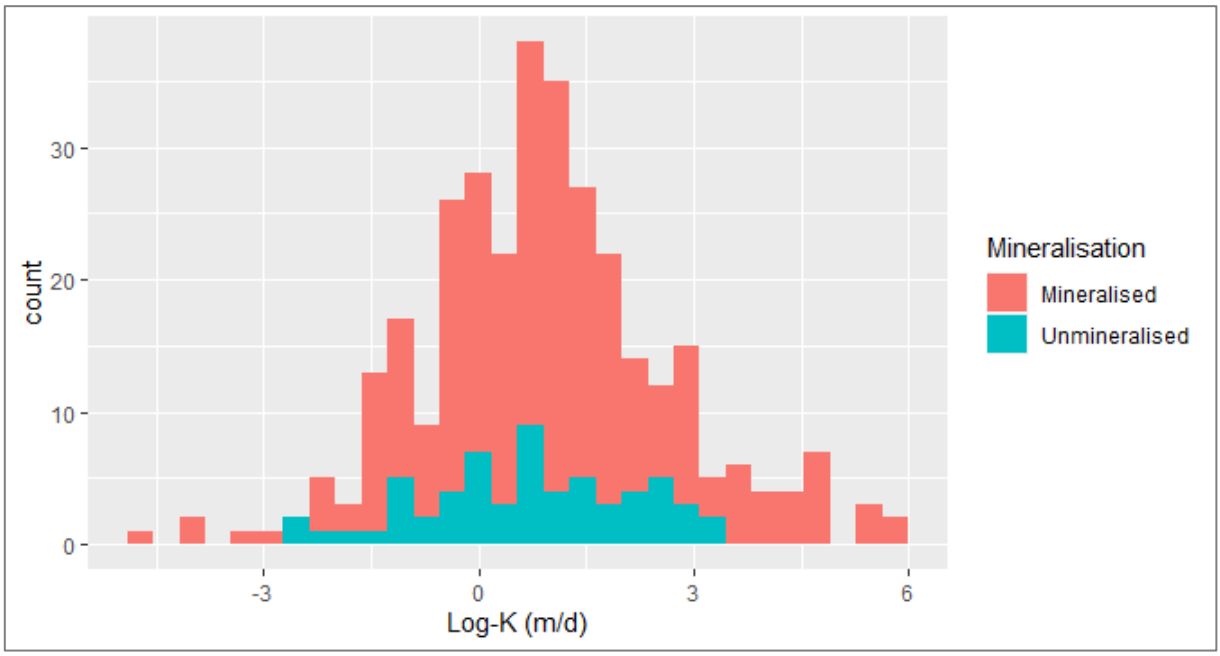

Figure 4.20 Histogram of log(K) of Bedded Iron formations by mineralised

Figure 4.20 indicates that the histogram of mineralised and un-mineralised for BIF look similar, which means the plot is normally distributed, and the log (K) average values are close. It is inferred that mineralised has no impact on hydraulic conductivity.

Figure 4.21 shows the density plot that would aid the comparison by normalising the different counts.

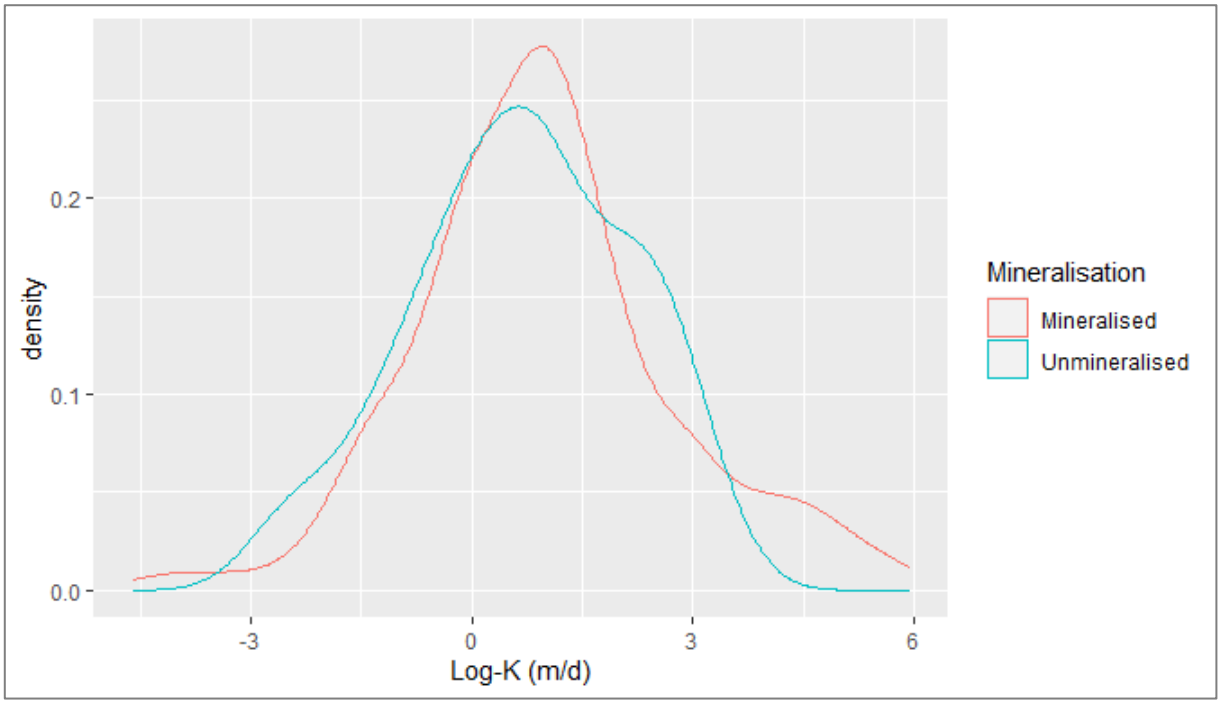

Figure 4.21 Comparison of mineralised & un-mineralised of log(K) of BIF

4.2.2 Fracture and unfractured rock

This section compared how fractured and unfractured formations or karstic material compared to the rest. Figures 4.22 – 4.24 indicates the normal distribution or unimodal, and the mean values of log (K) are close for both the fractured and unfractured rocks. The plot also shows that the median log (K) for unfractured is slightly higher than the fractured plot; thus, hydraulic conductivity has no impact on both aquifers. Notice that it is beyond the scope of this study to determine how fractured rocks are interconnected.

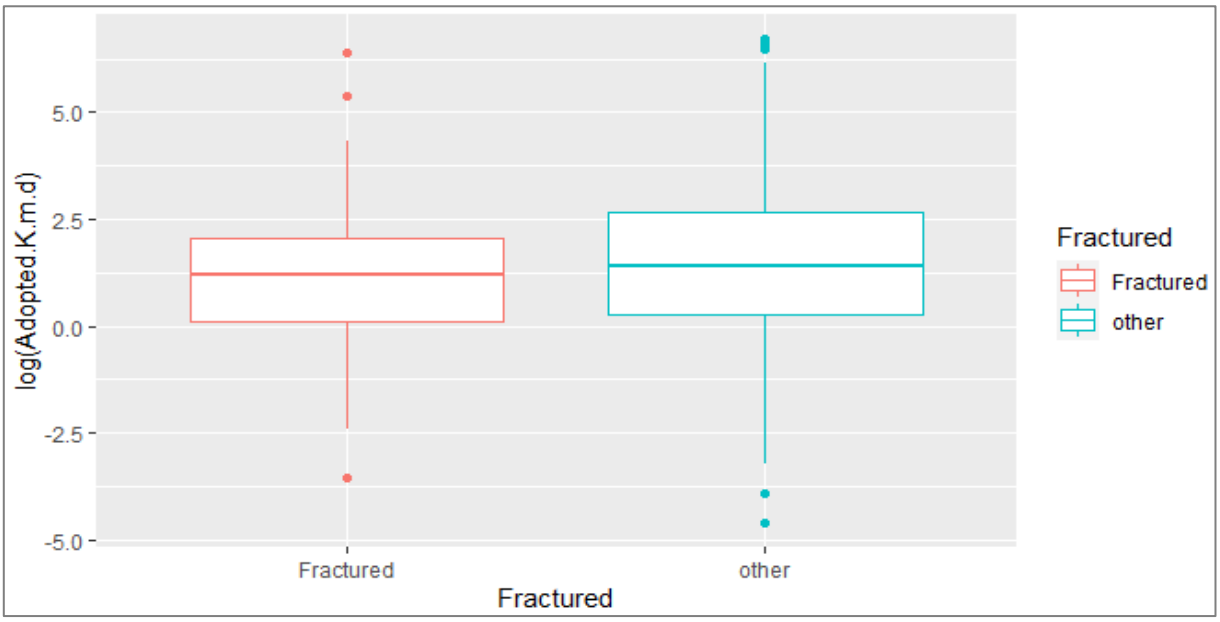

Figure 4.22 Box plot distribution of fractured versus unfractured rocks

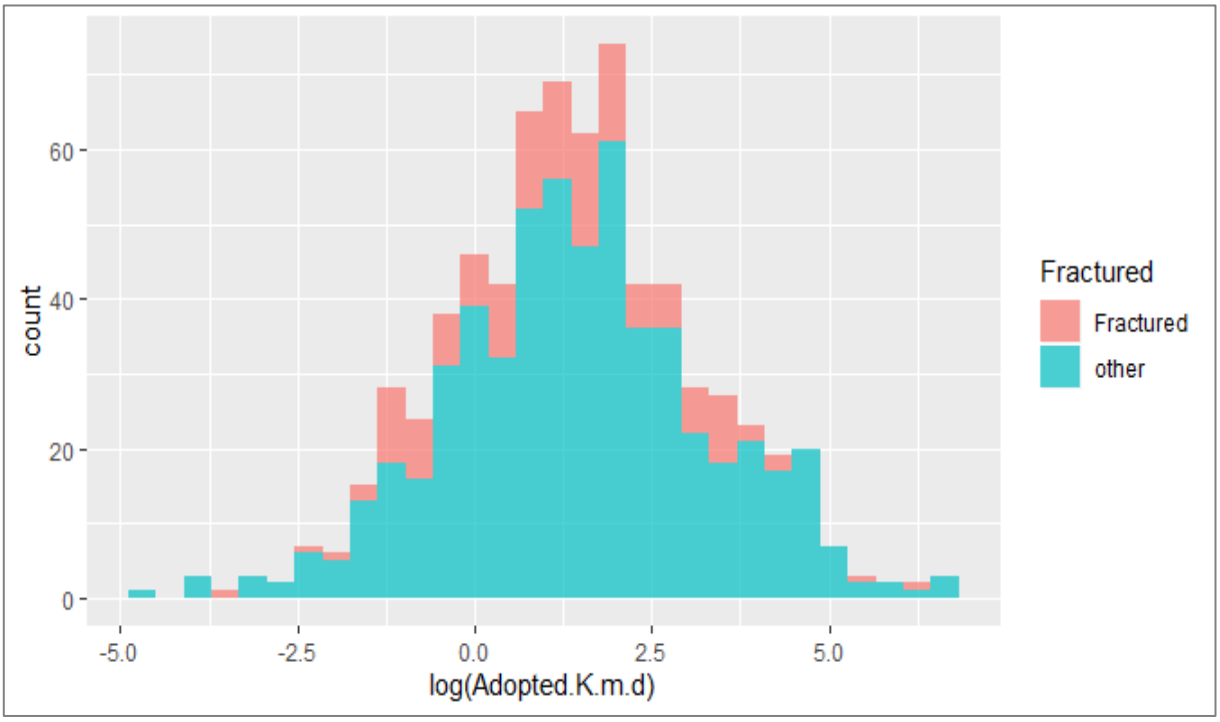

Figure 4.23 Histogram of distribution of log(K) of fractured vs unfractured rocks

The density plot is ideal for comparing by normalising the different counts, as illustrated in Figure 4.24).

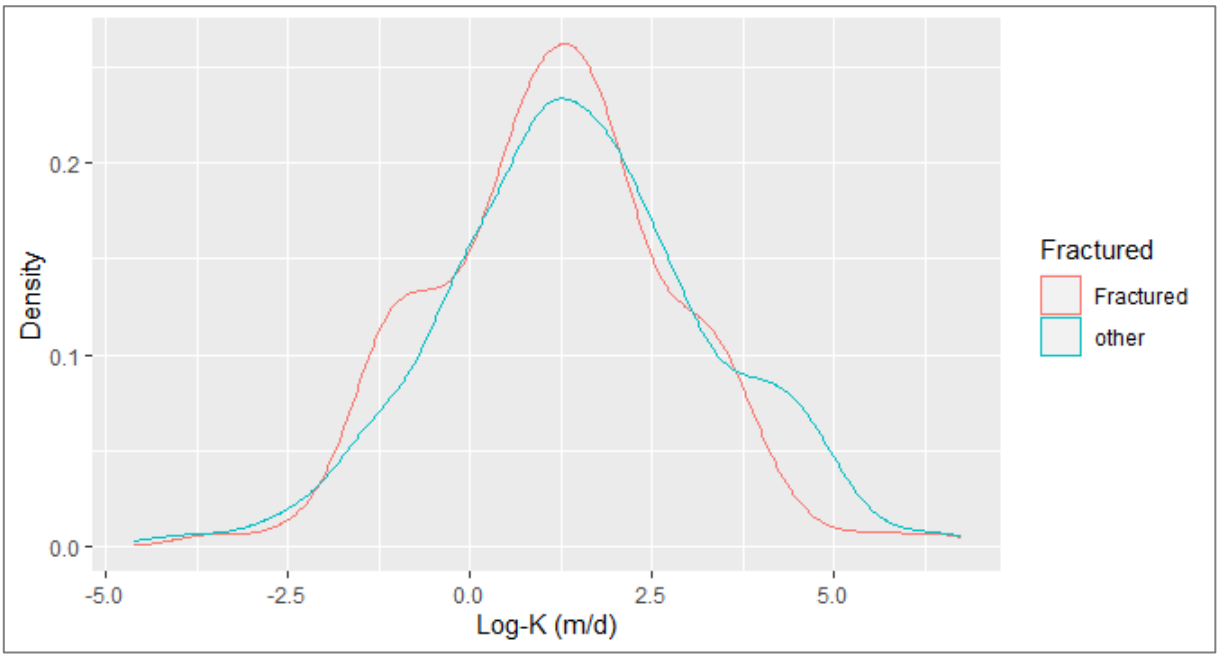

Figure 4.24 Density distribution of log(K) of fractured unfractured rocks

4.3 Distribution of log(K) by site

Table 4.3 presents the summary statistics for hydraulic conductivities for each site.

Table 4.3 Summary statistics of K by site

Site/Area	Log_K_mean (m/d)	K_mean (m/d)	K_max (m/d)	K_min (m/d)	K_Count
Brockman 1	-0.01	0.97	10	0.08	42
Brockman 2	-0.05	0.89	185	0.01	42
Brockman 4	0.37	2.33	12.2	0.09	63
Hope Downs 1	0.95	8.98	66.2	0.03	38
Hope Downs 4	0.29	1.96	143	0.1	87
Marandoo	1.12	13.15	600	0.23	93
Nammuldi	-0.05	0.88	18	0.02	68
Robe Valley	0.83	6.7	137	0.7	118
Section 10	0.86	7.29	44	0.69	20
Silvergrass East	0.35	2.24	24	0.48	35
Southern Fortescue	0.66	4.54	218	0.0.13	50
Tom Price	0.53	3.4	19.2	0.05	80
West Angelas	0.29	1.97	33.3	0.16	75
Yandicoogina	1.07	11.63	826	0.04	216

Figure 4.25 shows the spatial location of aquifer code and hydraulic conductivity values on each tested bore (i.e., aquifer type: K value in m/d). Appendix 4 shows the detailed spatial location of all the hydraulic conductivities by aquifer deposit and aquifer type.

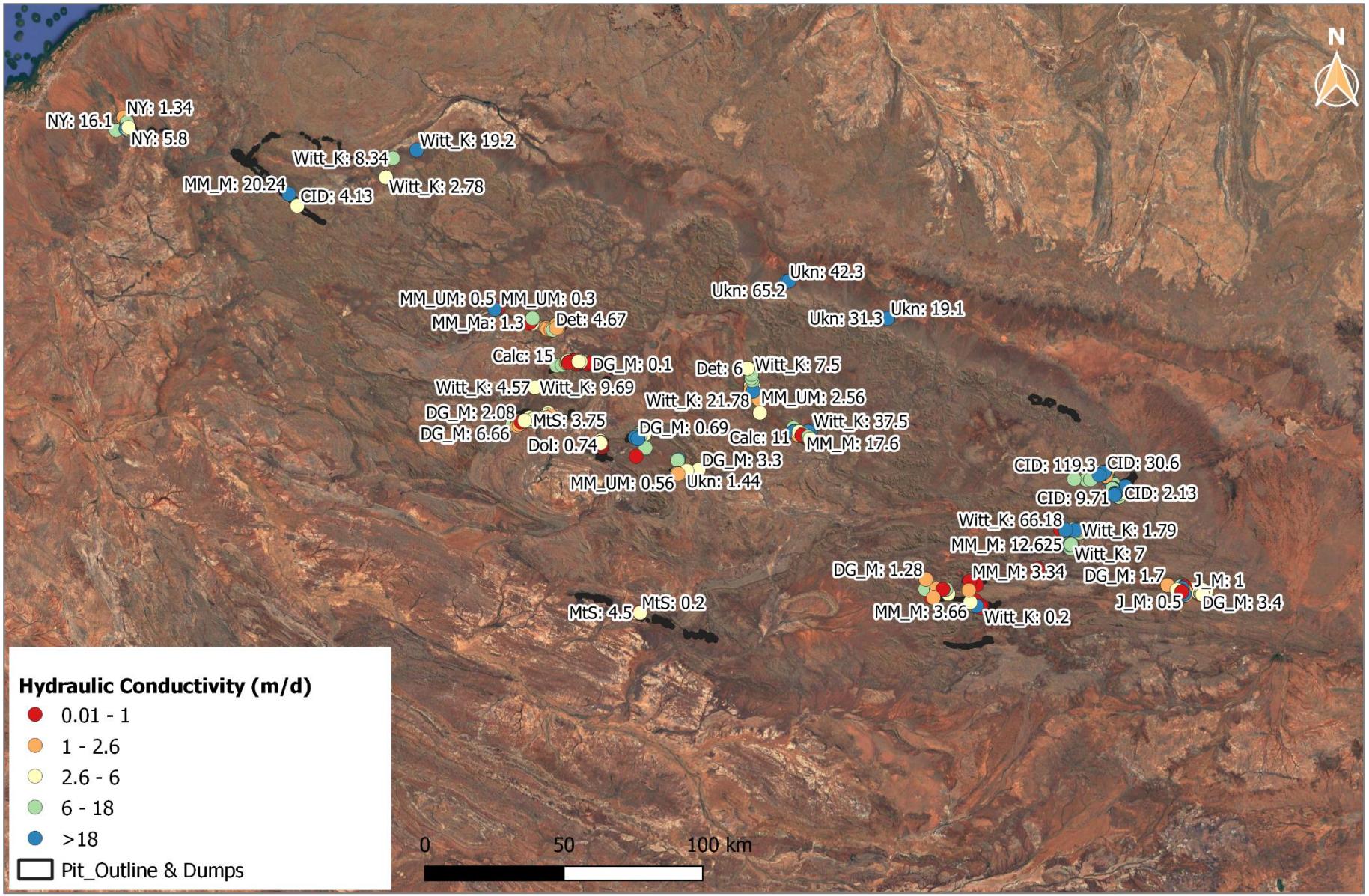

Figure 4.25 Spatial location of aquifer type & hydraulic conductivities (m/d) sample points on each bore

Figures 4.26 and 4.27 indicate site log-K aquifer values. These figures are presented separately for better visualisation or to avoid masking the plots and texts. It can be observed from the plots that hydraulic conductivity values vary spatially across all the sites due to the heterogeneity of the formation.

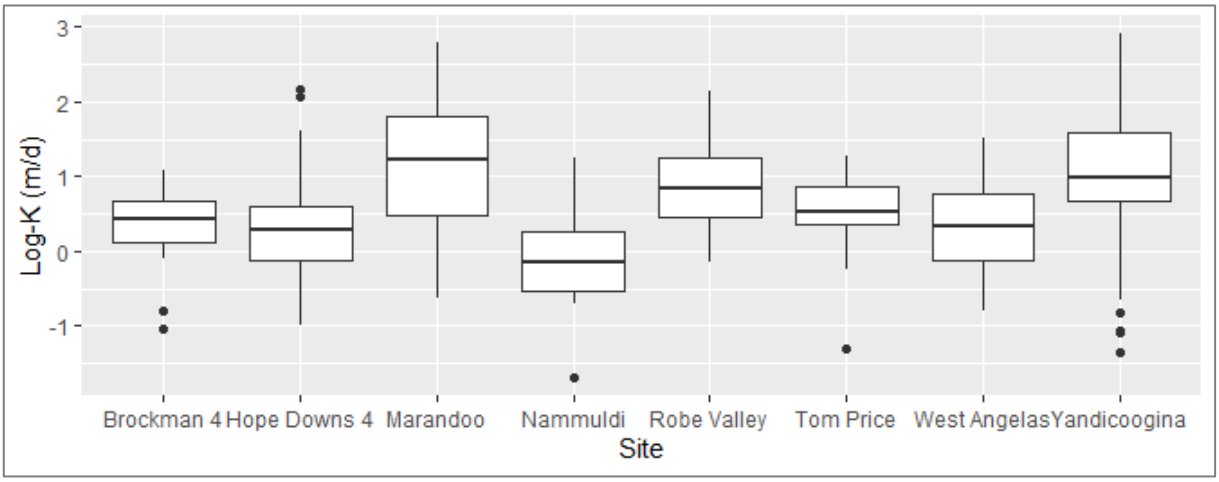

Figure 4.26 Box plot of log-transformed (K m/d) values by site

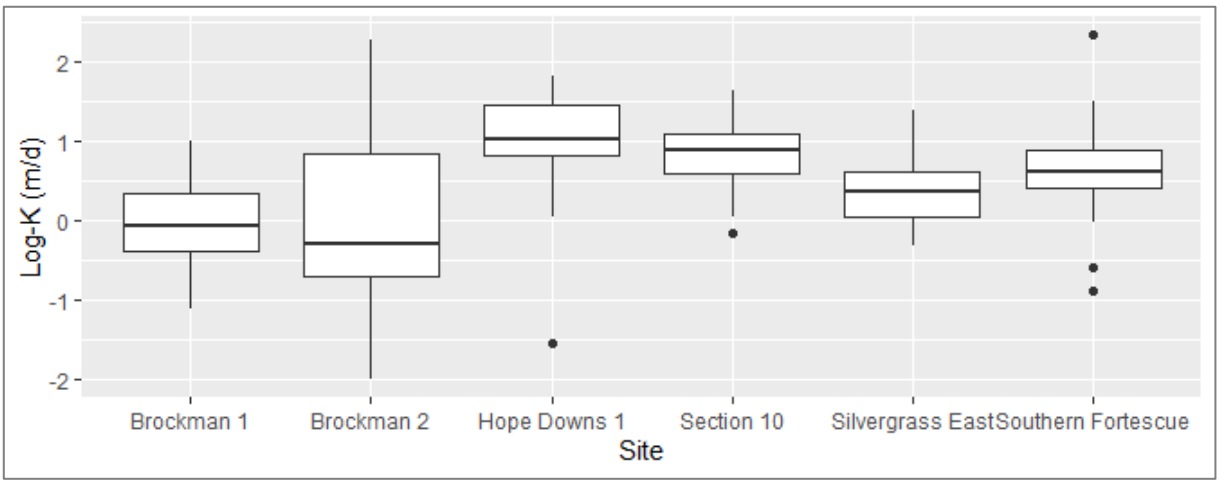

Figure 4.27 Box plot of log-transformed (K m/d) values by site

Figure 4.28 shows log (K) distributions by site and aquifer type.

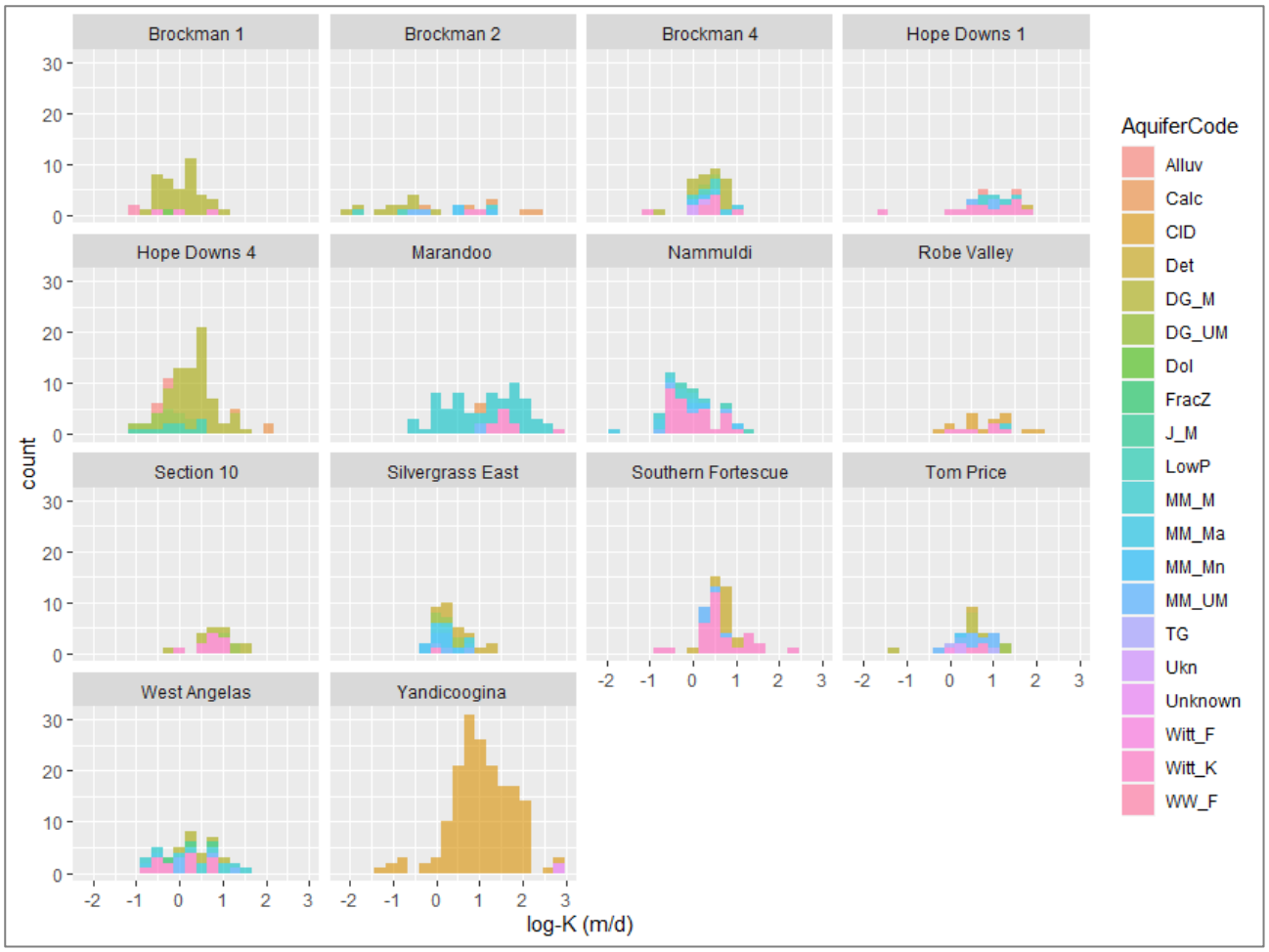

Figure 4.28 Histogram of log(K) values by site and aquifer type

The distribution for most aquifer types is consistent (e.g., for Yandicoogina and West Angelas), while other deposits such as Marandoo indicate multimodal distribution, suggesting spatial variability with the aquifer properties. These observed trends indicate higher K values for the Channel Iron and valley-fill deposits (Yandicoogina / Robe Valley) compared to the other deposits (Brockman, Marra Mamba deposits). Historically the CID deposit has generally been subjected to more aquifer test programmes due to risks associated with dewatering impact on creeks.

Figures 4.29 to 4.31 indicate the QQ and box plots for log-K sample points of mineralised and un-mineralised Marra Mamba formations by site and aquifer type. The log-K sample points displayed appear to contain lower values than the total sample count in Table 4.3. Lower displayed values result from fewer mineralised and un-mineralised Marra Mamba sample points. In addition, the total count of samples in Table 4.3 is based on the site, without explicitly differentiating mineralised or un-mineralised Marra Mamba units.

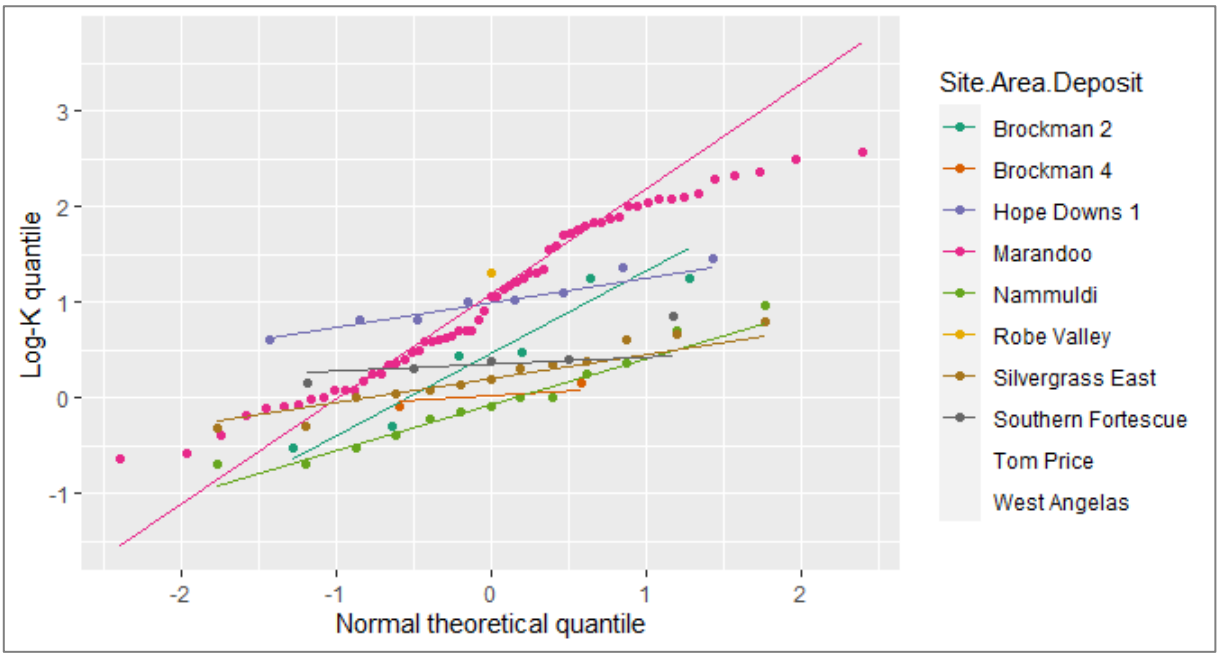

Figure 4.29 QQ-plot log-K Values of Marra Mamba by Site

Figure 4.29 indicates a significant deviation of hydraulic conductivities associated with all the sites due to aquifer spatial variability. There is no compelling reason to think the data points might behave differently because of formation heterogeneity. Butler and Liu (1993) also observed from their study that there are non-uniformities in the aquifer. It is not sufficient to characterise the aquifer properties in the real heterogeneous aquifer that exhibits large variability.

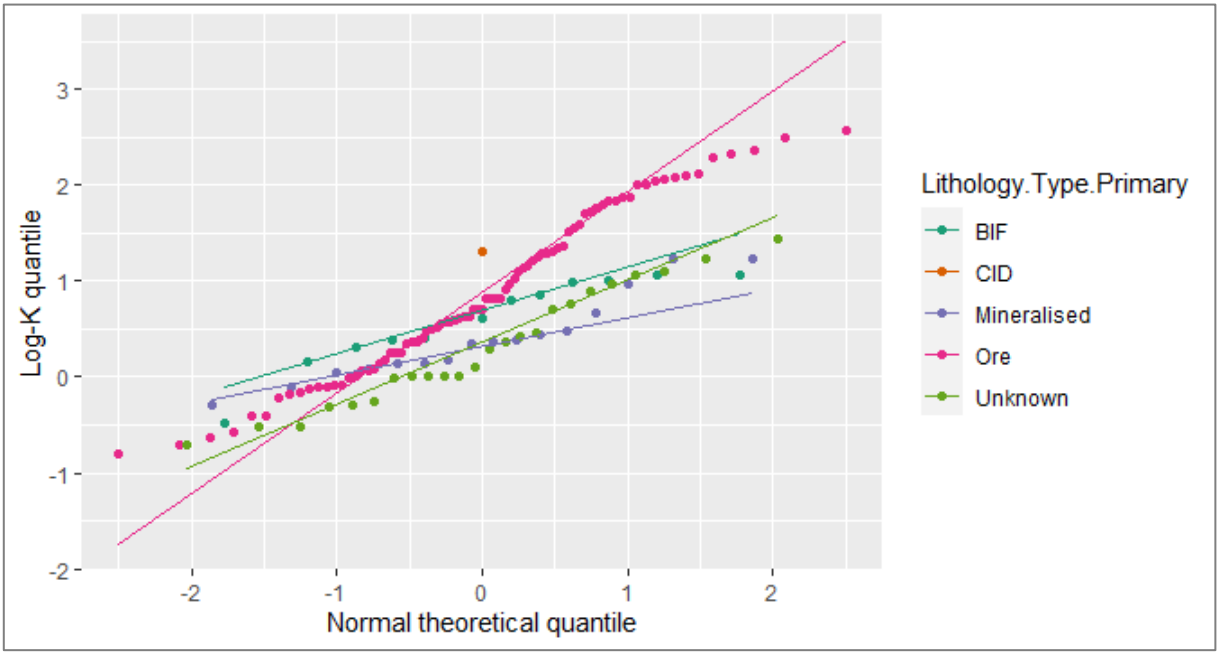

Figure 4.30 QQ-plot log(K) for Marra Mamba aquifer by aquifer type

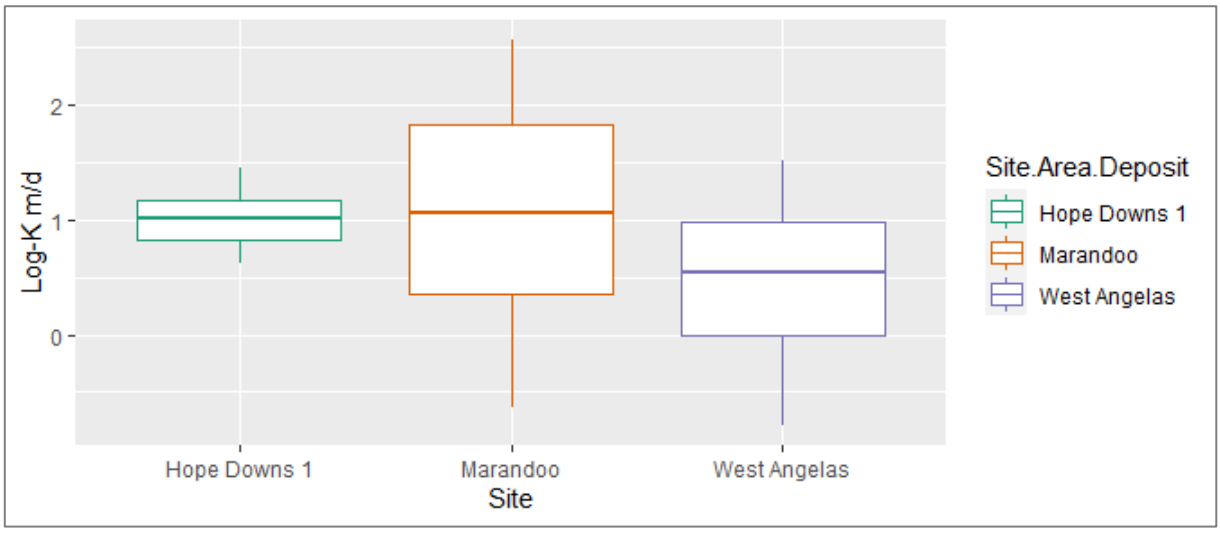

Figure 4.31 Box plot comparison of Marra Mamba sites

The Marandoo site has the highest median hydraulic conductivity compared to other sites in similar settings, and the plot is positive-skewed. The West Angelas site indicates a negative-skewed dataset; however, this can be attributed to the hydraulic conductivity variability despite the sites sitting in the same geology.

CHAPTER 5 DISCUSSION AND CONCLUSIONS

5.1 Discussion of Results

Based on the analysis results, the average hydraulic conductivities of aquifer types are mostly within approximately one order of magnitude (1m/d to 10m/d). Notice that the expected value of a log-normally distributed data set is not the mean. Instead, it is the mean of the log-transformed values. In the same vein, the histogram and density plots for aquifer types indicated that the distribution is normal and consistent (e.g., for Yandi and West Angelas). Other deposits such as Marandoo show multimodal distribution, suggesting spatial variability or scale effect with hydraulic conductivities. It is consistent with Brace (1980, 1984) findings that both the porous and fractured rocks exhibit a scale effect. It is observed that there are higher K values for deposits in the Channel Iron and valley-fill deposits. Historically the CID deposit has generally been subject to more pumping tests due to associated dewatering impact on creeks when mining than other deposits.

The assumption of log normality of the K distribution generally holds when plotting the data on a QQ plot. The distribution indicated some deviations in the highest and lowest quartiles but fit very well on the theoretical line. According to Sánchez-Vila et al. (1994), spatial variability influences the scale dependence of transmissivity and hydraulic conductivity. On the contrary, Oliver (1993) argued that a small-scale effect near the pumping bore could be influenced by the late-time drawdown effect of a large-scale non-uniformity or heterogeneous media.

In addition, individual interquartile ranges of K for each aquifer type generally cover less than an order of magnitude; however, CID and MM have noticeably larger interquartile ranges. The CID, DG_M, LowP and Witt_K have very low permeability outliers. The distribution of hydraulic conductivity (K) does not differ between aquifer types, mineralised/unmineralised groups, or even between fractured/unfractured rocks.

It is also observed from the hydraulic parameters by sites that Yandicoogina, Marandoo and Hope Downs 4 have larger hydraulic conductivity ranges. Marandoo has a noticeably higher mean hydraulic conductivity than the rest of the sites. The higher mean values could be attributed to drilling into the same aquifer. It is also observed that the Yandicoogina hydraulic conductivity values when plotting by site and aquifer type right-shifted distribution (i.e., on average, higher values).

5.2 Conclusions

The research objectives of this study were to determine how aquifer parameters vary across iron ore deposits, geographically and by aquifer type. Secondly, determine the distribution of hydraulic aquifer parameters across the Pilbara region by aquifer type and deposit.

The study also indicated that the lognormality of the distribution of hydraulic conductivities (K) generally holds or is consistent by aquifer type across geographical space, consistent with conceptual knowledge.

5.3 Limitations and Future Work

Limitations specific to this study are summarised as follows:

5.3.1 Limitations

- There is a lack of reasonable data quality, particularly aquifers' logging and differentiating between mineralised / unmineralised.
- The K values are the average of actual hydraulic conductivities, ranging over many orders of magnitude. It can also be argued that the long screens mean that the K values are averaged. The analysed data showed a correlation between screen length and K values.
- There is also variability in detail of aquifer testing undertaken by hydrogeologists. Some bore analyses are based on a single production/monitoring bore assessment type, whilst others are from more extensive test programmes with more than one observation bores.

5.3.2 Future Works

To enhance knowledge of hydraulic parameters, the following are suggestions for future work based on the findings of this thesis.

- Look at the early-stage pumping test (very local - the immediate area around the well) versus the long term bore response (more prominent geological structures come into play, homogenisation of the area, averaging out complexity). There is a need to reconcile further or validate the yields against geology or whether geology is not a factor.

- Significant amounts of data are sitting in reports, which is highly impractical, and It is challenging to yield insights due to decentralised data. It must also be improved with the automated extraction of relevant data from PDF files. For example, airlift yields from well completion logs

- Further study is required to correlate airlift yields with permeability, transmissivity, and other relevant hydraulic parameters. It would validate that an approach that combines statistical learning with well-completion data is viable.

- The pumping test results must be against observed responses during dewatering/water supply abstraction to confirm the initial assessment, with updates to conceptualisation and modelling as necessary.

REFERENCES

acQuire Technology Solutions, 2021. *acQuire & EnviroSys,* Perth: acQuire Technology Solutions Pty Ltd.

Asfahani, J., 2021. Classification of the transmissivity spatial variations of a Quaternary aquifer in the semi-arid Region Khanasser Valley, Northern Syria. *Journal of African Earth Sciences,* 104269(182), pp. 1-13.

BOM, 2021. *Bureau of Meteorology.* [Online]
Available at: http://www.bom.gov.au/climate/dwo/index.shtml
[Accessed 05 July 2021].

Brace, W., 1980. Permeability of crystalline and argillaceous rocks. *Int. J. Rock Mech. Min. Sci. and Geomech. Abstr./t,* Issue 241.

Brace, W., 1984. Permeability of crystalline rocks: New in situ measurements. *Geophys. Res,* 4327(B6), p. 89.

Butler, J. & Liu, W., n.d. Pumping tests in non-uniform aquifers: the radially asymmetric case. *Water Resources Research,* Volume 29, p. 259–269.

Clauser, C., 1992. Permeability of Crystalline Rocks. *EOS, Transitions, American Geophysical Union,* 73(21), pp. 233-240.

Corvi et al., 1992. Reservoir characterisation using expert knowledge, data and statistics. *Oilfield Review*, January, pp. 25-39.

Cresswell, J.W, 2018. *Research Design: Qualitative, Quantitative and Mixed Method Approaches.* 4th ed. Thousand Oaks (CA): Sage.

Dalstra, H., n.d. Structural Controls on High-grade Iron ores hosted by Banded Iron Formation: A global perspective. *Reviews in Economic Geology,* Volume 15, pp. 73-106.

de Smith, M., 2015. *Statistical Analysis Handbook: A Comprehensive Handbook of Statistical Concepts,* Winchelsea, UK: The Winchelsea Press.

Delhomme, J., 1978. Kriging in the hydrosciences. *Advan. Water Resour,* 1(15), pp. 251-266.

Department of Water, 2010. *Pilbara regional water plan 2010 - 2030,* Perth, Western Australia: Department of Water.

Department of Water, 2010. *Pilbara regional water plan 2010 - 2030,* Perth, Western Australia: s.n.

Dogramaci, K., 2009. *Yandicoogina Hydrogeological Field Program Report - Bore Installation and Test Pumping,* s.l.: RTIO-PDE-0074001.

Floriancic, M., 2019. Spatial variability in specific discharge and streamwater chemistry during low flows: Results from snapshot sampling campaigns in eleven Swiss catchments. *Hydrological Processes,* 39(22), pp. 2847-2866.

Freeze, R., 1975. A stochastic conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resource,* 11(5), pp. 725-741.

Hamersley Iron Pty Ltd, 2003. *Geology and Mineralogy of the Hamersley Province Iron Ores, the Year 2000 Update,* s.l.: Confidential Internal Report.

Johnson, S. & Wright, A., 2001. Central Pilbara Groundwater Study, Water and Rivers Commission. *Hydrogeological Record Series,* p. 102.

Kabacoff, R., 2015. *R in Action: Data Analysis and Graphics with R.* Greenwich, USA: Manning Publications Co.

Kolterman, C. & Gorelick, S., 1996. Heterogeneity in sedimentary deposits: A review of structure imitating, process-Imitating, and descriptive approaches. *Water Resource Research,* 39(9), pp. 2617-2658.

Loaiciga, H., 2015. Probability Distributions in Groundwater Hydrology: Methods and Applications. *J. Hydrol. Eng,* Volume 5, p. 20.

Morris, R., 1994. *Detrital Iron Deposits of the Hamersley Province,* s.l.: CSIRO Division of Exploration and Mining GDSR 2536.

Neumann, S., 1990. Universal scaling of hydraulic conductivities and dispersivities in geologic media. *Water Resour. Res,* 28(8), p. 1749.

Nijp, J., 2019. High-resolution peat volume change in a northern peatland: Spatial variability, main drivers, and impact on ecohydrology. *Ecohydrology,* 12(6), p. 2114.

Nilsson, B. H. A. R. J. C. a. T., 2007. Uncertainty in geological and hydrogeological data. *Hydrol. Earth Syst. Sci,* Volume 11, pp. 1551-1561.

Oliver, D., 1993. The influence of non-uniform transmissivity and storativity on drawdown. *Water Resources Research,* Issue 29, p. 169–178.

Osnes, H., 1995. Stochastic analysis of head spatial variability in bounded rectangular. *Water Resour Res,* Issue 31, pp. 2981-90.

Punch, K. F., 2009. *Introduction to Research Methods in Education.* London: Sage Publications.

R Core Team, 2019. *R: A Language and Environment for Statistical Computing.* [Online]
Available at: https://www.R-project.org/
[Accessed 05 June 2021].

Rio Tinto - Iron Ore, 2021. *Google EC - Internal Rio Tinto,* Perth: Rio Tinto.

Rio Tinto Iron Ore, 2010. *Green Book - Geology and Mineralogy of the Hamersley Province Ores,* Perth: Rio Tinto Iron Ore.

Rio Tinto Iron Ore, 2014. *Blue Book - Hydrogeology of the Hamersley Basin,* Perth: Water Resource Evaluation Department.

Riva, M., Neuman, S. & Guadagnini, A., 2017. Theoretical analysis of non-Gaussian heterogeneity effects on subsurface flow and transport. *Water Resour. Res,* 53(10.1002/2016WR019353), p. 2998–3012.

Sanchez-Vila, X., Carrera, J. & Girardi, P., 1994. Scale effects in transmissivity. *Journal of Hydrology,* pp. 1-22.

Sue, Z., 1999. Simple Probabilistic and Statistical Risk Calculations in an Aquifer. *GROUNDWATER,* 5(37), pp. 748 - 754.

APPENDIX 1: NEWMAN AERO ANNUAL RAINFALL

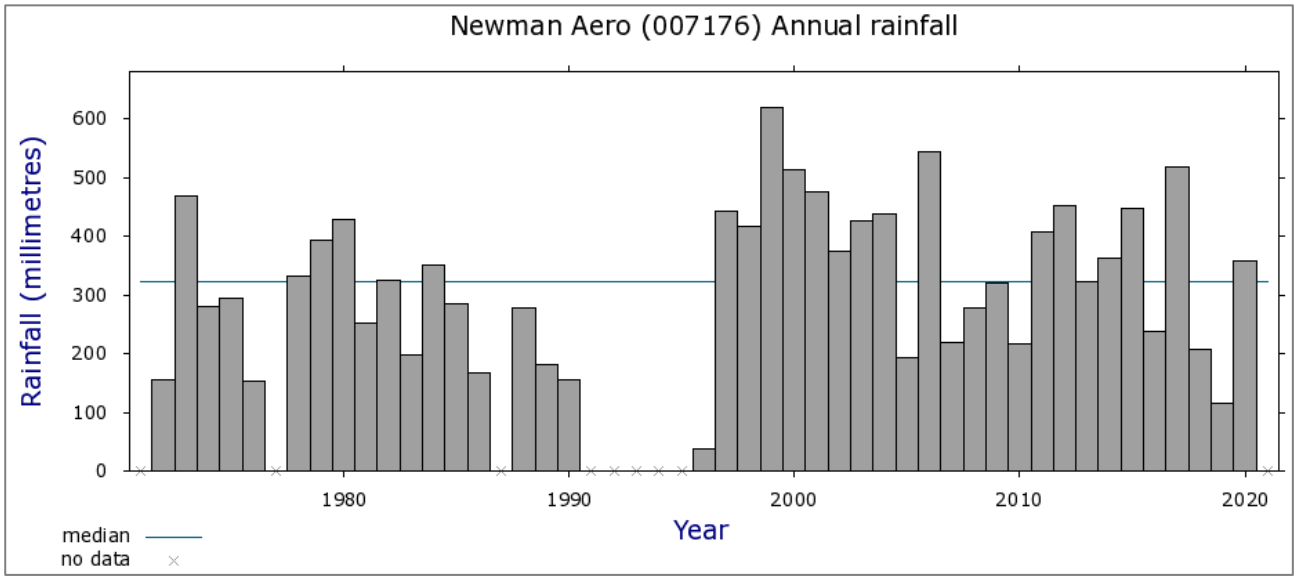

Figure 4.32 Newman Aero historical annual rainfall dating from 1971 - 2021

APPENDIX 2: R-PROGRAMMING CODES

The following steps are the extracts of the R-programming codes used for the analysis, and the figures presented in the analysis section are also sequential as the code here. The code is presented in steps, so it can quickly be followed.

Step 1: Set the importing library and load the data file

```
library(tidyverse)
path <- "./"
gw <- read.table(paste0(path, "PB_cleaned.csv"), header=T, sep=",")
```

Step 2: Summary statistics for hydraulic conductivity (K m/d)

```
gw %>%
  filter(is.numeric(Adopted.K.m.d) & !is.na(Adopted.K.m.d)) %>%
  summarise(K_mean = mean(Adopted.K.m.d, na.rm = T),
            K_max = max(Adopted.K.m.d, na.rm = T),
            K_min = min(Adopted.K.m.d, na.rm = T),
            K_count = n())
```

```
##     K_mean    K_max K_min K_count
## 1 20.42789 826.4463  0.01     704
```

Step 3: Plotting distribution of K (m/d) on all the dataset

```
ggplot(data = gw, aes(x=Adopted.K.m.d)) +
  geom_histogram(bins=50) +
  ggtitle("Distribution of K - all data") +
  labs(x = "K (m/d)")
```

Step 4: **Plotting distribution log of hydraulic conductivity**

```
ggplot(data=gw, aes(x=log10(Adopted.K.m.d))) +
  geom_histogram(bins=50) +
  ggtitle("Distribution of log(k) - all data") +
  labs(x = "Log(K) (m/d)")
```

Step 5: Plot of QQ-plot of log (K)

```
ggplot(gw, aes(sample = log10(Adopted.K.m.d))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("qq-plot of log(K) - all data")
```

Step 6: Screened thickness, test during and discharge

```
print(paste("Number of observations with a known screened thickness: ", gw %>%
filter(!is.na(PB.Screened.Thickness.m)) %>% summarise(count=n())))
```

```
## [1] "Number of observations with a known screened thickness:  871"
```

```
gw %>%
  ggplot(aes(x=PB.Screened.Thickness.m)) +
  geom_histogram(bins=30) +
  labs(title='Histogram of Screened Thickness, m', x='Screened thickness, m')
```

Step 7: Histogram of pumping test during (minutes)

```
print(paste("Number of observations with a known pump test duration: ", gw %>%
filter(!is.na(Duration.of.Test.min)) %>% summarise(count=n())))
```

```
## [1] "Number of observations with a known pump test duration:  921"
```

```
gw %>%
  ggplot(aes(x=log10(Duration.of.Test.min))) +
  geom_histogram(bins=20) +
  labs(title='Histogram of Log Pump Test Duration, minutes', x='Log Pump Test Duration, minutes')
```

Step 8: Histogram distribution of discharge rate

```
print(paste("Number of observations with a known discharge rate: ", gw %>%
filter(!is.na(Discharge.Rate.KL.d.m3.day)) %>% summarise(count=n())))
```

```
## [1] "Number of observations with a known discharge rate:  919"
```

```
gw %>%
  ggplot(aes(x=(Discharge.Rate.KL.d.m3.day))) +
  geom_histogram(bins=25) +
  labs(title='Histogram of Discharge Rate, kL/day', x='Discharge Rate, kl/day')
```

Step 9: QQ-plot of log(K) screened by the thickness

```
gw <- gw %>%
  mutate(Screened.Thickness = case_when(
    PB.Screened.Thickness.m < 50 ~ "< 50",
    between(PB.Screened.Thickness.m, 50, 100) ~ "50-100",
    PB.Screened.Thickness.m > 100 ~ "> 100"
    )) %>%
  mutate(Screened.Thickness = factor(Screened.Thickness, levels=c("< 50", "50-100", "> 100")))

gw <- gw %>%
  mutate(Test.Duration.Hrs = case_when(
    Duration.of.Test.min < 6000 ~ "< 100",
    between(Duration.of.Test.min, 6000, 12000) ~ "100-200",
    Duration.of.Test.min > 12000 ~ "> 200"
```

```
  )) %>%
  mutate(Test.Duration.Hrs = factor(Test.Duration.Hrs, levels=c("< 100", "100-200", "> 200")))

gw <- gw %>%
  mutate(Discharge.Rate = case_when(
    Discharge.Rate.KL.d.m3.day < 3000 ~ "< 3000",
    between(Discharge.Rate.KL.d.m3.day, 3000, 6000) ~ "3000-6000",
    Discharge.Rate.KL.d.m3.day > 6000 ~ "> 6000"
    )) %>%
  mutate(Discharge.Rate = factor(Discharge.Rate, levels=c("< 3000", "3000-6000", "> 6000")))

num_quantiles = nrow(gw %>% select(Adopted.K.m.d) %>% na.omit())

qq_data <- gw %>%
  filter(!is.na(Adopted.K.m.d)) %>%
  mutate(Log.Adopted.K.quantile = Adopted.K.m.d %>%
           log10() %>%
           quantile(seq(1./num_quantiles, 1., 1./num_quantiles))) %>%
  mutate(Theoretical.quantile = qnorm(
    seq(1./num_quantiles, 1., 1./num_quantiles),
    mean = mean(Adopted.K.m.d %>% log10()),
    sd = sd(Adopted.K.m.d %>% log10())
    ))

# qq-plot conditioned on the screened thickness
qq_data %>%
  ggplot(aes(x=Theoretical.quantile, y=Log.Adopted.K.quantile)) +
    geom_point(aes(colour = Screened.Thickness), alpha=0.6) +
    geom_line(aes(x=Theoretical.quantile, y=Theoretical.quantile)) +
    labs(x = 'theoretical', y = 'sample', title = 'qq-plot of log(K) by Screened Thickness (m)')
```

Step 10: QQ-plot conditioned on test duration

```
# qq-plot conditioned on test duration
qq_data %>%
  ggplot(aes(x=Theoretical.quantile, y=Log.Adopted.K.quantile)) +
    geom_point(aes(colour = Test.Duration.Hrs), alpha=0.6) +
    geom_line(aes(x=Theoretical.quantile, y=Theoretical.quantile)) +
    labs(x = 'theoretical', y = 'sample', title = 'qq-plot of log(K) by Test Duration (hrs)')
```

Step 11: QQ-plot conditioned on discharge rate

```
# qq-plot conditioned on discharge rate
qq_data %>%
  ggplot(aes(x=Theoretical.quantile, y=Log.Adopted.K.quantile)) +
    geom_point(aes(colour = Discharge.Rate), alpha=0.6) +
    geom_line(aes(x=Theoretical.quantile, y=Theoretical.quantile)) +
    labs(x = 'theoretical', y = 'sample', title = 'qq-plot of log(K) by Discharge Rate (kL/d)')
```

Step 12: Distribution of log K by Thickness

```
gw %>%
  group_by(Screened.Thickness) %>%
  ggplot(aes(x=Screened.Thickness, y=log10(Adopted.K.m.d))) +
  geom_boxplot() +
  ggtitle("Distribution of log(K) by Bore Screened Thickness (m)") +
  labs(x="Screened Thickness (m)", y="Log(K) (m/d)")
```

Step 13: Distribution of log K by pumping test during (hrs)

```
gw %>%
  group_by(Test.Duration.Hrs) %>%
  ggplot(aes(x=Test.Duration.Hrs, y=log10(Adopted.K.m.d))) +
  geom_boxplot() +
  ggtitle("Distribution of log(K) by Pump Test Duration (hrs)") +
  labs(x="Test Duration (hrs)", y="Log(K) (m/d)")
```

Step 14: Distribution of log K by discharge rate

```
gw %>%
  group_by(Discharge.Rate) %>%
  ggplot(aes(x=Discharge.Rate, y=log10(Adopted.K.m.d))) +
  geom_boxplot() +
  ggtitle("Distribution of log(K) by Discharge Rate (kl/d)") +
  labs(x="Discharge Rate (kL/d)", y="Log(K) (m/d)")
```

Step 15: Density plot of log K conditioned on the screened thickness

```
gw %>%
  ggplot(aes(x=log10(Adopted.K.m.d))) +
  geom_density(aes(colour=Screened.Thickness)) +
  labs(title='Density plot of log(K) conditioned on Screened Thickness', x='log(K), m/d')
```

```
## Warning: Removed 354 rows containing non-finite values (stat_density).
```

Density plot of log K conditioned on test duration

```
gw %>%
  ggplot(aes(x=log10(Adopted.K.m.d))) +
  geom_density(aes(colour=Test.Duration.Hrs)) +
  labs(title='Density plot of log(K) conditioned on Test Duration', x='log(K), m/d')
```

```
## Warning: Removed 354 rows containing non-finite values (stat_density).
```

Step 16: Density plot of log K conditioned on discharge rate

```
gw %>%
  ggplot(aes(x=log10(Adopted.K.m.d))) +
```

```
  geom_density(aes(colour=Discharge.Rate)) +
  labs(title='Density plot of log(K) conditioned on Discharge Rate', x='log(K), m/d')
```

```
## Warning: Removed 354 rows containing non-finite values (stat_density).
```

Step 17: Distribution of log K by Aquifer Type

```
riotinto.aquifers <- read.table(paste0(path, "AquiferCodes.csv"), header=T, sep=",")
print(riotinto.aquifers)
```

Step 18: Aquifer grouping

```
# make groupings
LowP <- c('BeeG', 'MtS', 'McRS', 'WBS') # low permeability group
detritals <- c('Det')
cid <- c('CID', 'TP', 'NY') # Robe piezolite, channel iron and Yarraloola conglomerate

gw <- gw %>%
  mutate(AquiferCode = case_when(as.character(Lithology.Aquifer) %in% LowP ~ 'LowP',
                                 as.character(Lithology.Aquifer) %in% cid ~ 'CID',
                                 as.character(Lithology.Aquifer) == "-999" ~ 'Unknown',
                                 TRUE ~ as.character(Lithology.Aquifer)))

gw %>%
  group_by(AquiferCode) %>%
  summarise(count = n(), mean_log_K=mean(log(Adopted.K.m.d), na.rm=T)) %>%
  arrange(-count) %>%
  print(n=Inf)
```

Step 19: Plotting of log K by aquifer group (code)

```
gw %>%
  group_by(AquiferCode) %>%
  filter(n() >= 10) %>%
  ggplot(aes(x=AquiferCode, y=log10(Adopted.K.m.d))) +
  geom_boxplot() +
  ggtitle("Distribution of log(K) by Aquifer Code") +
  labs(x="Aquifer Code", y="Log(K) (m/d)")
```

Step 20: Histogram of log K by aquifer type

```
gw %>%
  group_by(AquiferCode) %>%
  filter(n() >= 10) %>%
  ggplot(aes(x=log10(Adopted.K.m.d))) +
    geom_histogram(bins=30) +
    facet_wrap(~AquiferCode) +
```

```
  ggtitle('Distribution of log(K) by Aquifer Type') +
 labs(x="Log(K) (m/d)")
```

Step 21: Grouping by mineralised formations

```
# make groupings
Mineralised <- c('DG_M', 'Witt_F', 'J_M', 'MM_M', 'MM_Mn', 'WW_M', 'MM_Ma') # bedded iron mineralised group
Unmineralised <- c('DG_UM', 'J_F',  'MM_UM', 'WW_F', 'WBS')  # bedded unmineralised group

gw <- gw %>%
 mutate(Mineralisation = case_when(as.character(Lithology.Aquifer) %in% Mineralised ~ 'Mineralised',
                                   as.character(Lithology.Aquifer) %in% Unmineralised ~ 'Unmineralised',
                                   TRUE ~ 'other'))

gw %>%
 filter(Mineralisation %in% c('Mineralised', 'Unmineralised')) %>%
 group_by(Mineralisation) %>%
 summarise(mean_log_k = mean(log(Adopted.K.m.d), na.rm=T))
```

Step 22: Distribution of mineralised log K of BIF

```
gw %>%
 filter(Mineralisation %in% c('Mineralised', 'Unmineralised')) %>%
 ggplot(aes(x=log(Adopted.K.m.d))) +
  geom_histogram(aes(fill=Mineralisation), bins=30) +
  ggtitle('Distribution of log(K) of Bedded Iron formations by Mineralisation')
```

Step 23: Distribution of log K of BIF mineralised

```
gw %>%
 filter(Mineralisation %in% c('Mineralised', 'Unmineralised')) %>%
 ggplot(aes(x=log(Adopted.K.m.d))) +
  geom_density(aes(colour=Mineralisation)) +
  ggtitle('Distribution of log(K) of Bedded Iron formations by Mineralisation')
```

Step 24: Making groupings for fractured rock

```
# make groupings
Fractured <- c('Dol_FracZ', 'FracZ', 'Witt_K') # fractured rock / karstics

gw <- gw %>%
 mutate(Fractured = case_when(as.character(Lithology.Aquifer) %in% Fractured ~ 'Fractured',
                              TRUE ~ 'other'))

gw %>%
```

```
  ggplot(aes(x=Fractured, y=log(Adopted.K.m.d))) +
    geom_boxplot(aes(colour=Fractured)) +
    ggtitle('Distribution of log(K) of fractured vs unfractured rock')
```

Step 25: Distribution of log K of fractured vs unfractured rock

```
gw %>%
  ggplot(aes(x=log(Adopted.K.m.d))) +
    geom_histogram(aes(fill=Fractured), bins=30, alpha=0.7) +
    ggtitle('Distribution of log(K) of fractured vs unfractured rock')
```

Step 26: Distribution of log K of fractured vs unfractured rock

```
gw %>%
  ggplot(aes(x=log(Adopted.K.m.d))) +
    geom_histogram(aes(fill=Fractured), bins=30, alpha=0.7) +
    ggtitle('Distribution of log(K) of fractured vs unfractured rock')
```

Step 27: Normalised Density plots

```
gw %>%
  ggplot(aes(x=log(Adopted.K.m.d))) +
    geom_density(aes(colour=Fractured)) +
    ggtitle('Distribution of log(K) of Fractured vs Unfractured Rock')
```

Step 28: Aquifer by Site

```
# by site
gw %>%
  group_by(Site.Area.Deposit) %>%
  summarise(K_mean = mean(Adopted.K.m.d, na.rm = T),
            K_max = max(Adopted.K.m.d, na.rm = T),
            K_min = min(Adopted.K.m.d, na.rm = T),
            K_count = n()) %>%
  filter(K_count >= 20)


sites <- gw %>%
  group_by(Site.Area.Deposit) %>%
  summarise(count = n()) %>%
  filter(count > 50) %>%
  select(Site.Area.Deposit)

gw %>%
  filter(Site.Area.Deposit %in% sites$Site.Area.Deposit) %>%
  ggplot(aes(x=Site.Area.Deposit, y=log10(Adopted.K.m.d))) +  # , colour=Site.Area.Deposit
  geom_boxplot() +
```

```
  scale_color_brewer(palette="Dark2") +
  ggtitle("Distribution of log(K) by Site")
```

Step 29: Distribution of log K by site and aquifer type

```
gw %>%
  filter(Site.Area.Deposit %in% sites$Site.Area.Deposit) %>%
  ggplot(aes(x=log10(Adopted.K.m.d))) +
    geom_histogram(aes(fill=AquiferCode), bins=50, alpha=0.6) +
    facet_wrap(~Site.Area.Deposit) +
    ggtitle('Distribution of log(K) by Site and Aquifer Type')
```

Step 30: Fitting PDF to major formation – Marra Mamba aquifer

```
gw %>%
  filter(Lithology.Aquifer %in% c("MM_M", "MM_UM")) %>%
  ggplot(aes(sample = log10(Adopted.K.m.d))) +
  stat_qq() +
  stat_qq_line() +
  ggtitle('qq-plot log(K) - all Marra Mamba Aquifers')
```

Step 31: QQ-plot of hydraulic conductivity of Marra Mamba aquifer by site

```
gw %>%
  filter(Lithology.Aquifer %in% c("MM_M", "MM_UM")) %>%
  ggplot(aes(sample = log10(Adopted.K.m.d), colour=Site.Area.Deposit)) +
  stat_qq() +
  stat_qq_line() +
  scale_color_brewer(palette="Dark2") +
  ggtitle('qq-plot log(K) - Marra Mamba Aquifers by Site')
```

Step 32: QQ-plot log hydraulic conductivity of Marra Mamba aquifer by type

```
gw %>%
  filter(Lithology.Aquifer %in% c("MM_M", "MM_UM")) %>%
  ggplot(aes(sample = log10(Adopted.K.m.d), colour=Lithology.Type.Primary)) +
  stat_qq() +
  stat_qq_line() +
  scale_color_brewer(palette="Dark2") +
  ggtitle('qq-plot log(K) - Marra Mamba Aquifers by Type')
```

Step 33: Boxplot of Marra Mamba log K by site

```
gw %>%
  filter(Lithology.Aquifer %in% c("MM_M", "MM_UM") & Site.Area.Deposit %in% c("Marandoo",
"West Angelas", "HD1")) %>%
  ggplot(aes(x=Site.Area.Deposit, y=Overall.Ranking, colour=Site.Area.Deposit)) +
```

```
geom_boxplot() +
scale_color_brewer(palette="Dark2")
```

APPENDIX 3: AQUIFER CODES

Table 4.4 indicates an extract from the R-programming representing the aquifer codes used in R-programming software. Notice that the codes were shortened to avoid being unreadable or not easy to follow; however, it is easy to modify within the codes.

Table 4.4 Aquifer Codes

Aquifer Type Code	Code
Alluvials	Alluv
Bee Gorge	BeeG
Calcrete/Silcrete	Calc
Channel Iron Deposit	CID
Dales Gorge Unmineralised	DG_UM
Dales Gorge Min	DG_M
Detrital_TD1	Det
Detrital_TD2	Det
Detrital_TD3	Det
Dolerite Dyke Faulted	Dol_FracZ
Dolerite Sill	Dol
Fault Zone	FracZ
Fractured Rock	FracZ
Paraburdoo Fresh	Witt_F
Joffre Fresh	J_F
Joffre Min	J_M
Paraburdoo Karstic	Witt_K
Marra Mamba Mineralised	MM_M
Marra Mamba UnMin	MM_UM
McRaeShale	MCS
Mt. Sylvia Formation MtS	MtS
Newman Mineralised	MM_Mn
Robe Pisolite CID	TP
Weeli Wolli Fresh	WW_F
Weeli Wolli Min	WW_M
West Angelas Mineralised	MM_Ma
Yarraloola Conglomerate	NY
Turee Creek Group	TG
Unknown	Ukn
Whaleback Shale	WBS

APPENDIX 4: HYDRAULIC CONDUCTIVITIES BY DEPOSIT & AQUIFER TYPE

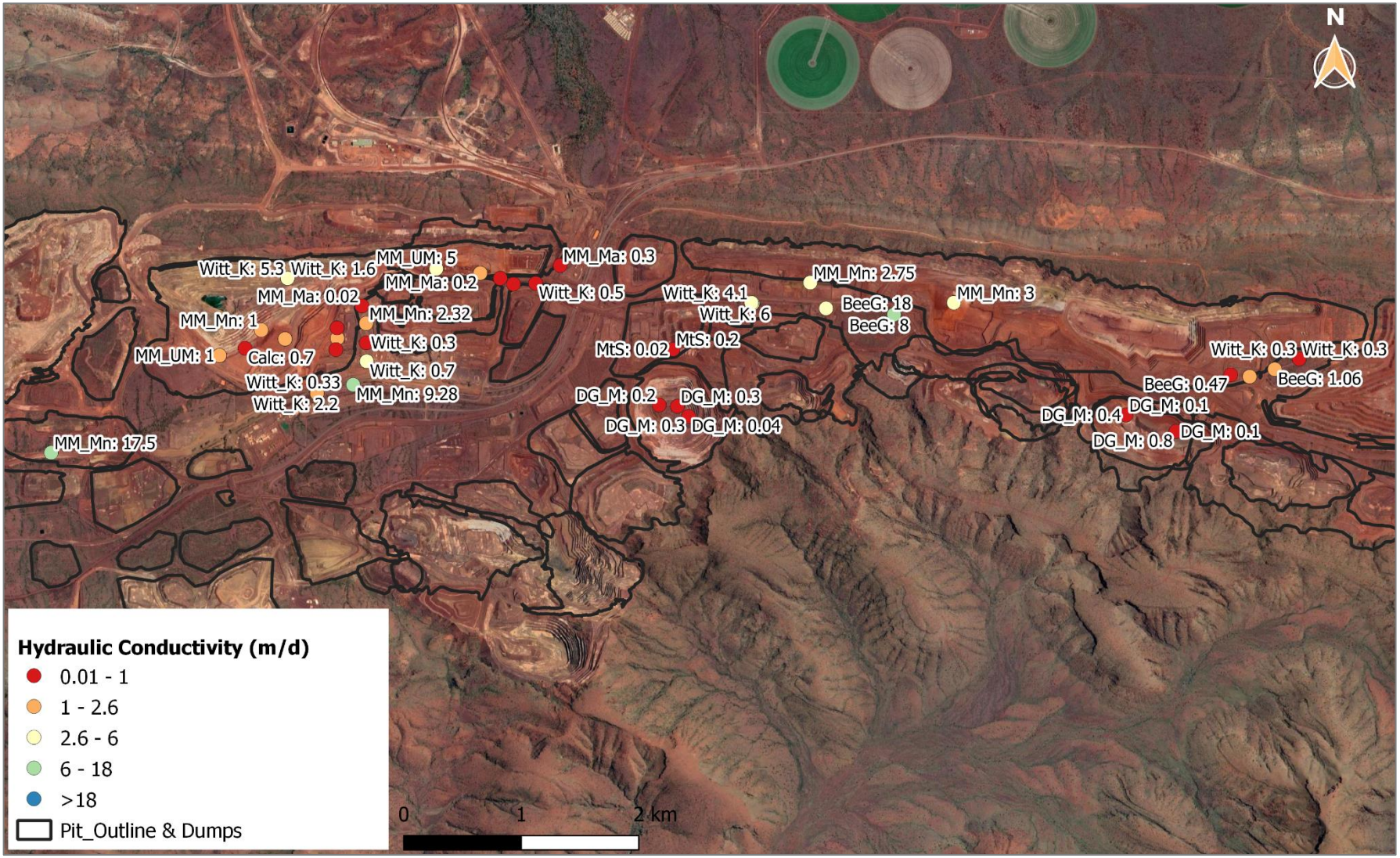

Figure 4.33 Brockman Syncline 2 (BIF Deposit) & Nammuldi (MMIF Deposit) – Aquifer Code: K (m/d) on each tested well

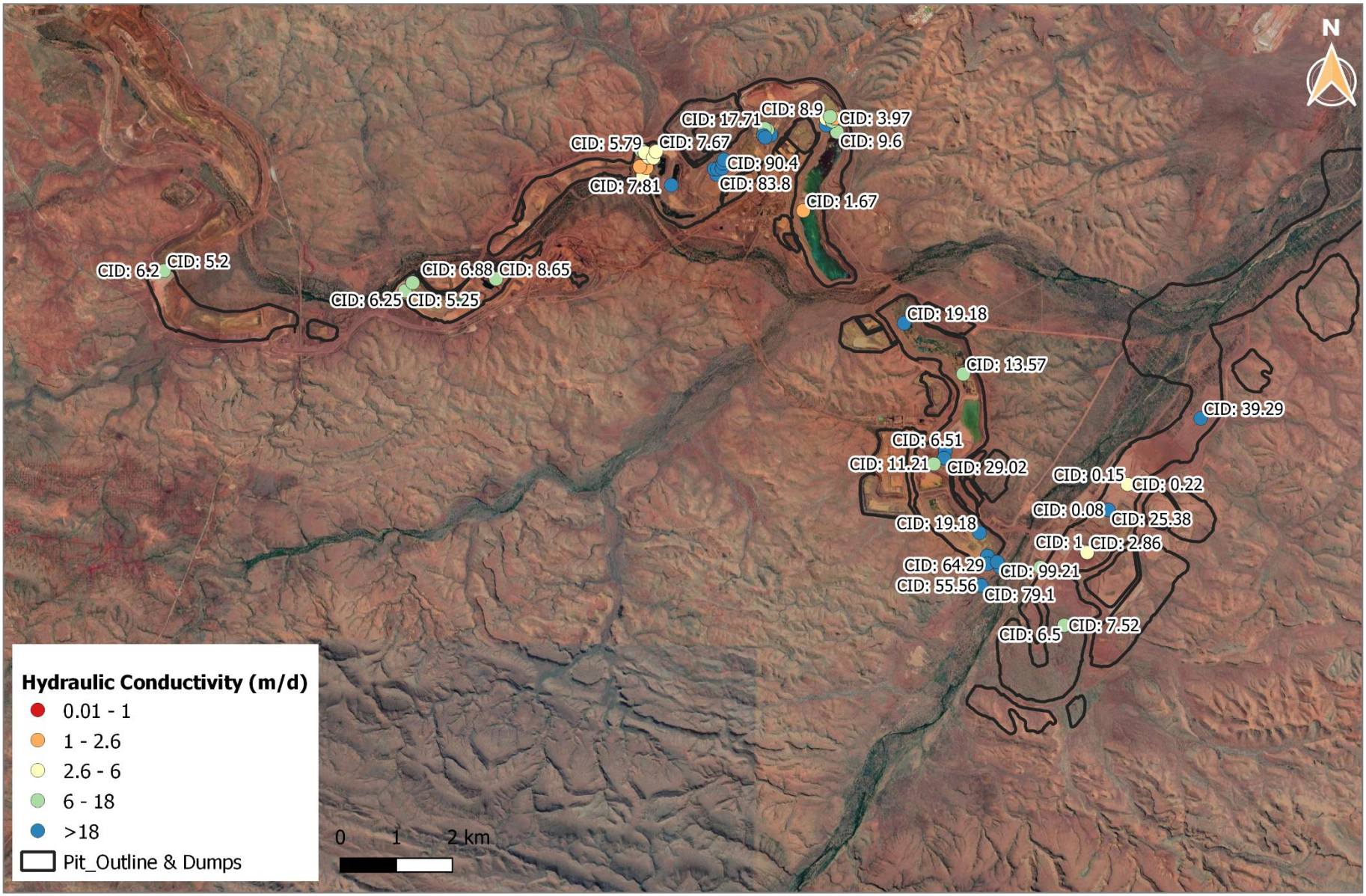

Figure 4.34 Yandicoogina (CID Deposit) – Aquifer Code: K (m/d) on each tested well

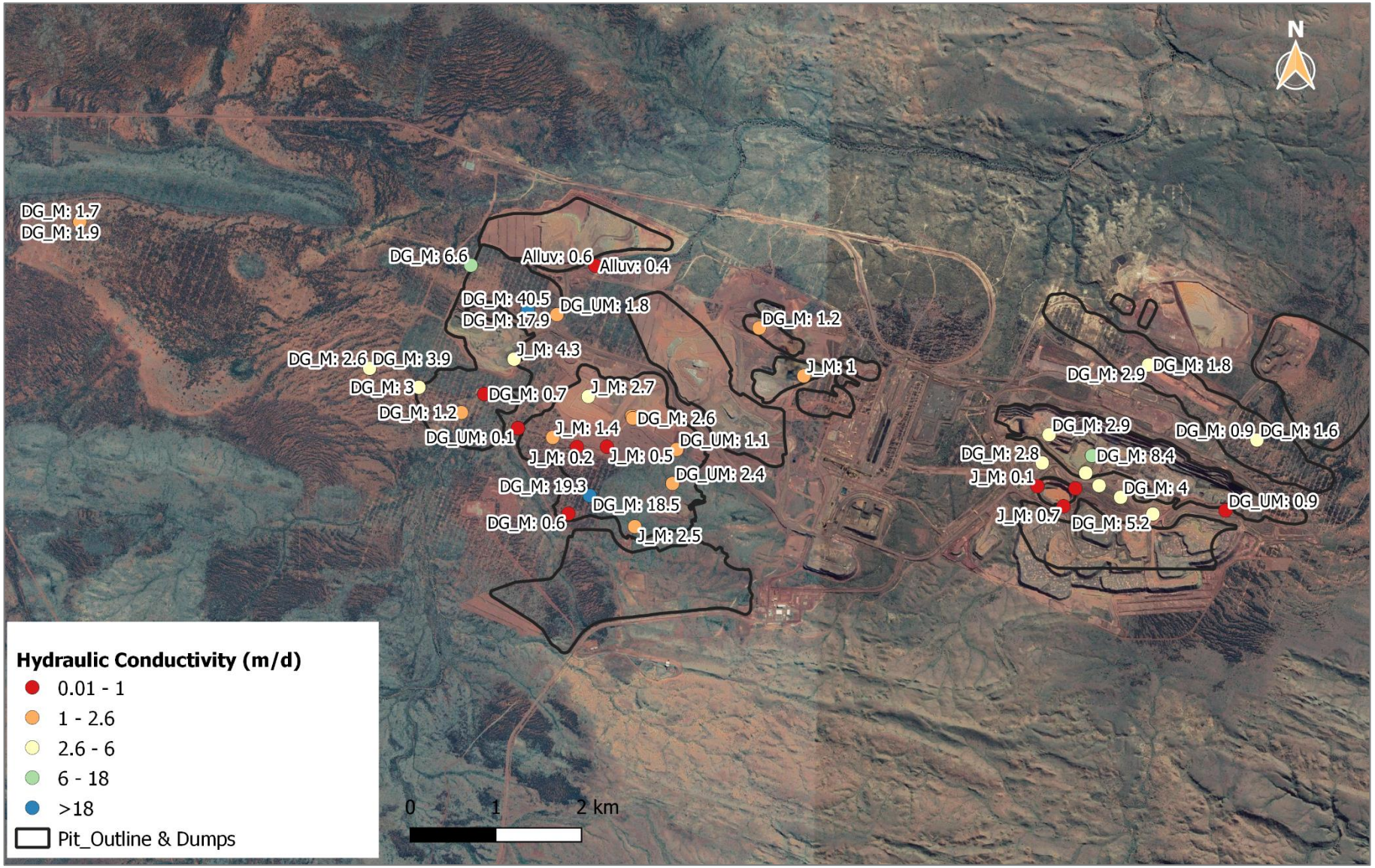

Figure 4. 35 Hope Down 4 (BIF Deposit) – Aquifer Code: K (m/d) on each tested well

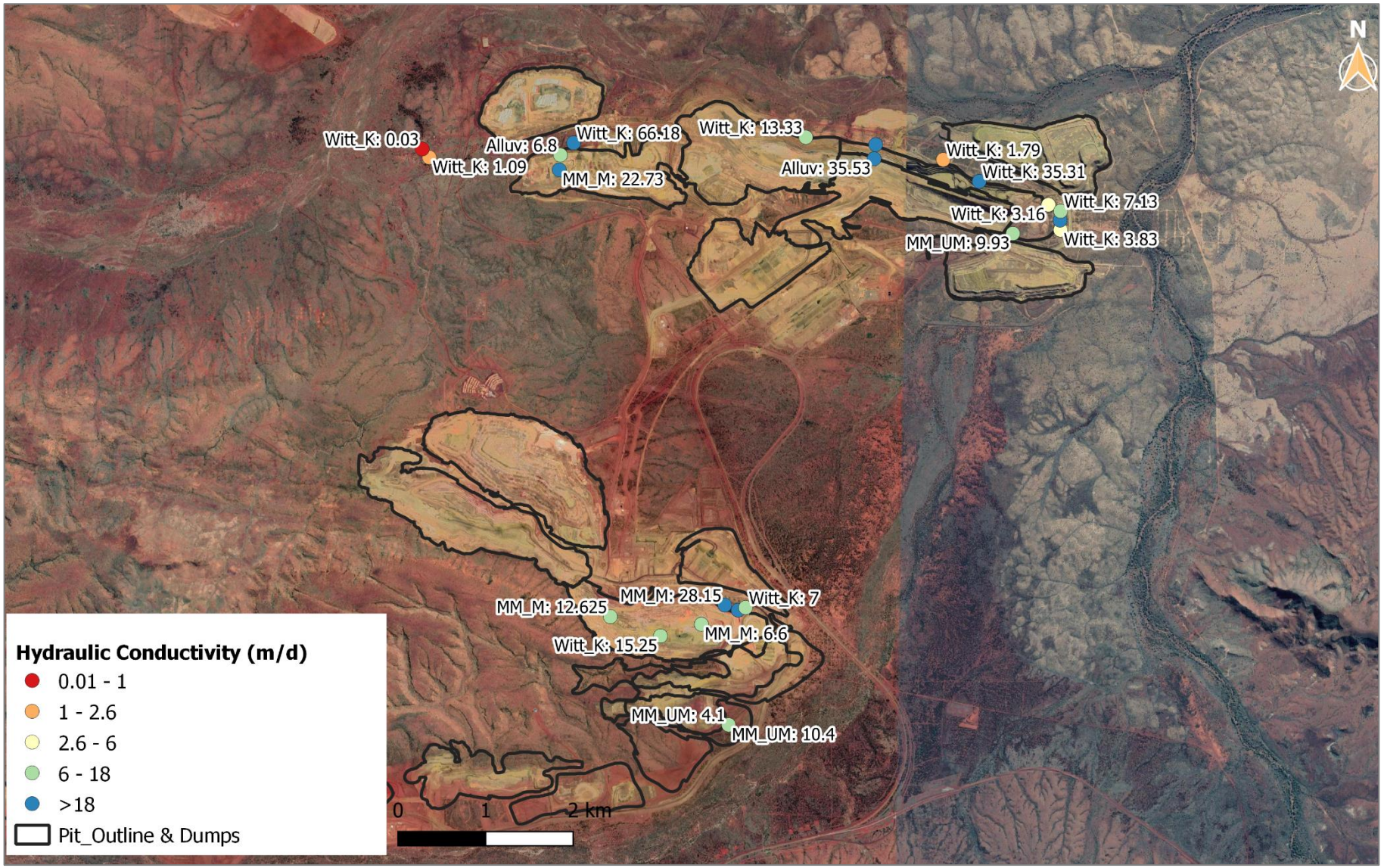

Figure 4.36 Hope Downs 1 (MMIF Deposit) – Aquifer Code with hydraulic Conductivities (m/d) on each tested well

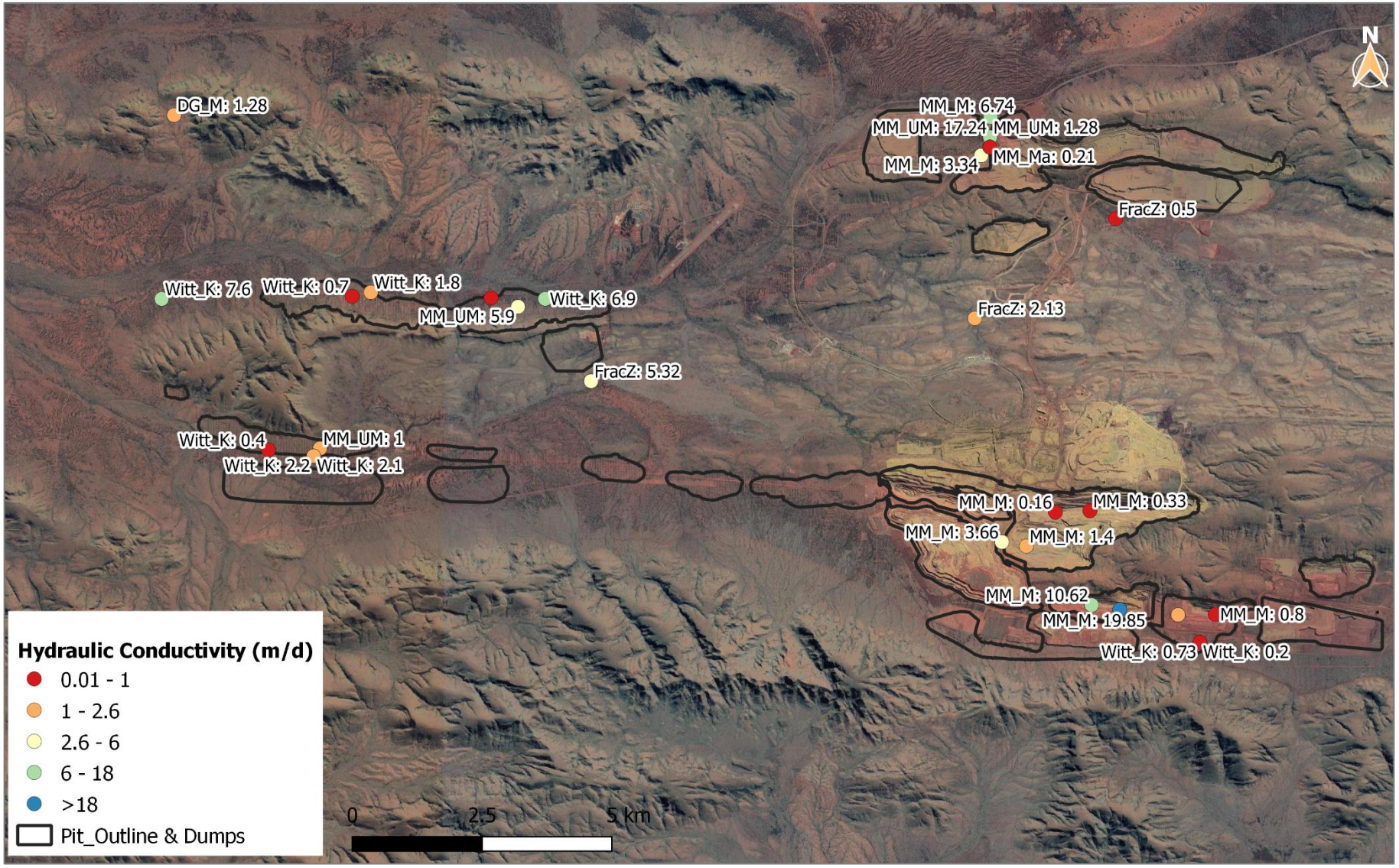

Figure 4.37 West Angelas (MMIF Deposit) – Aquifer Code: K (m/d) on each tested well

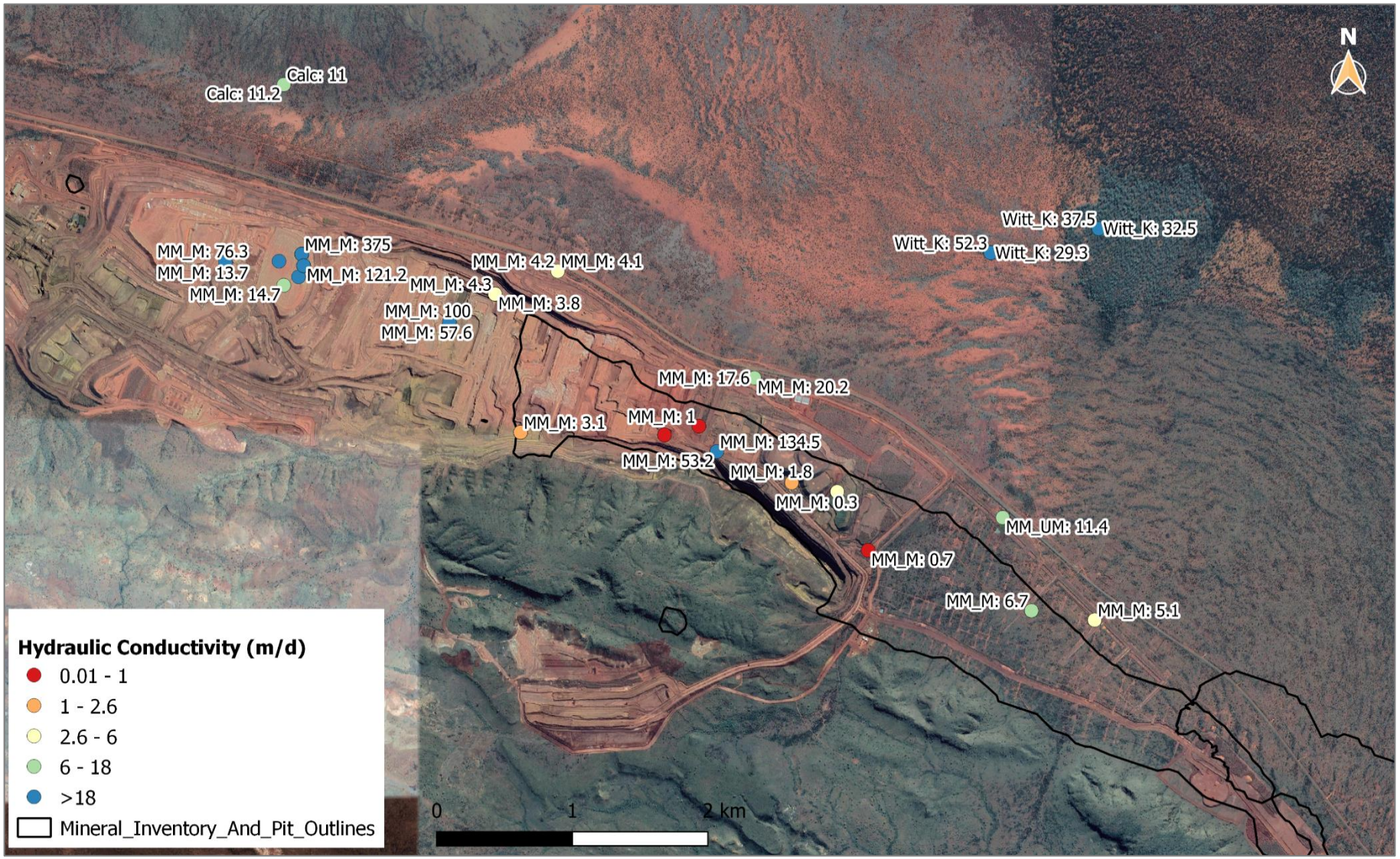

Figure 4.38 Marandoo (MMIF Deposit) – Aquifer Code with hydraulic Conductivities (m/d) on each tested well

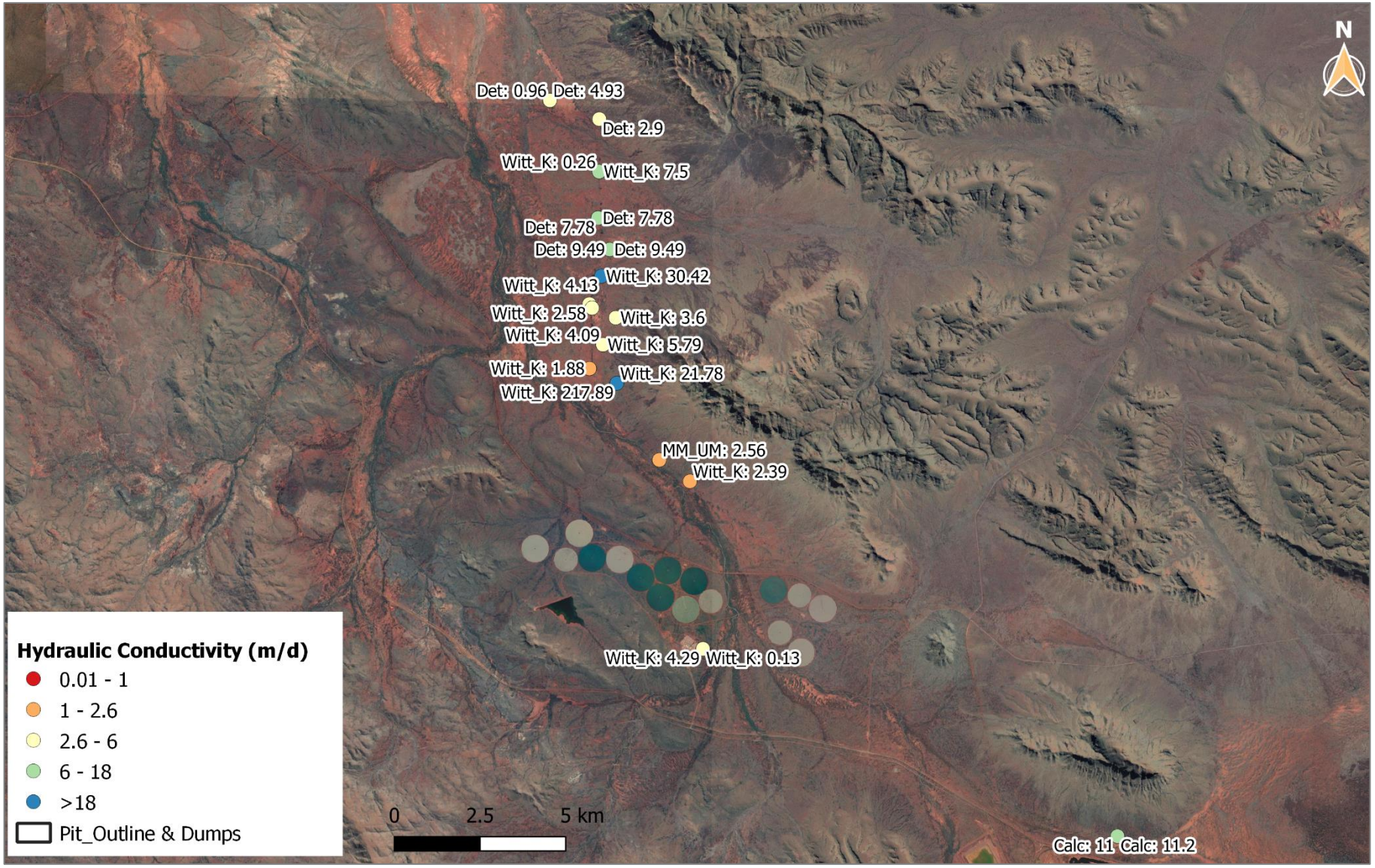

Figure 4.39 Southern Fortescue (MMIF Deposit) - Aquifer Code: K (m/d) on each tested well

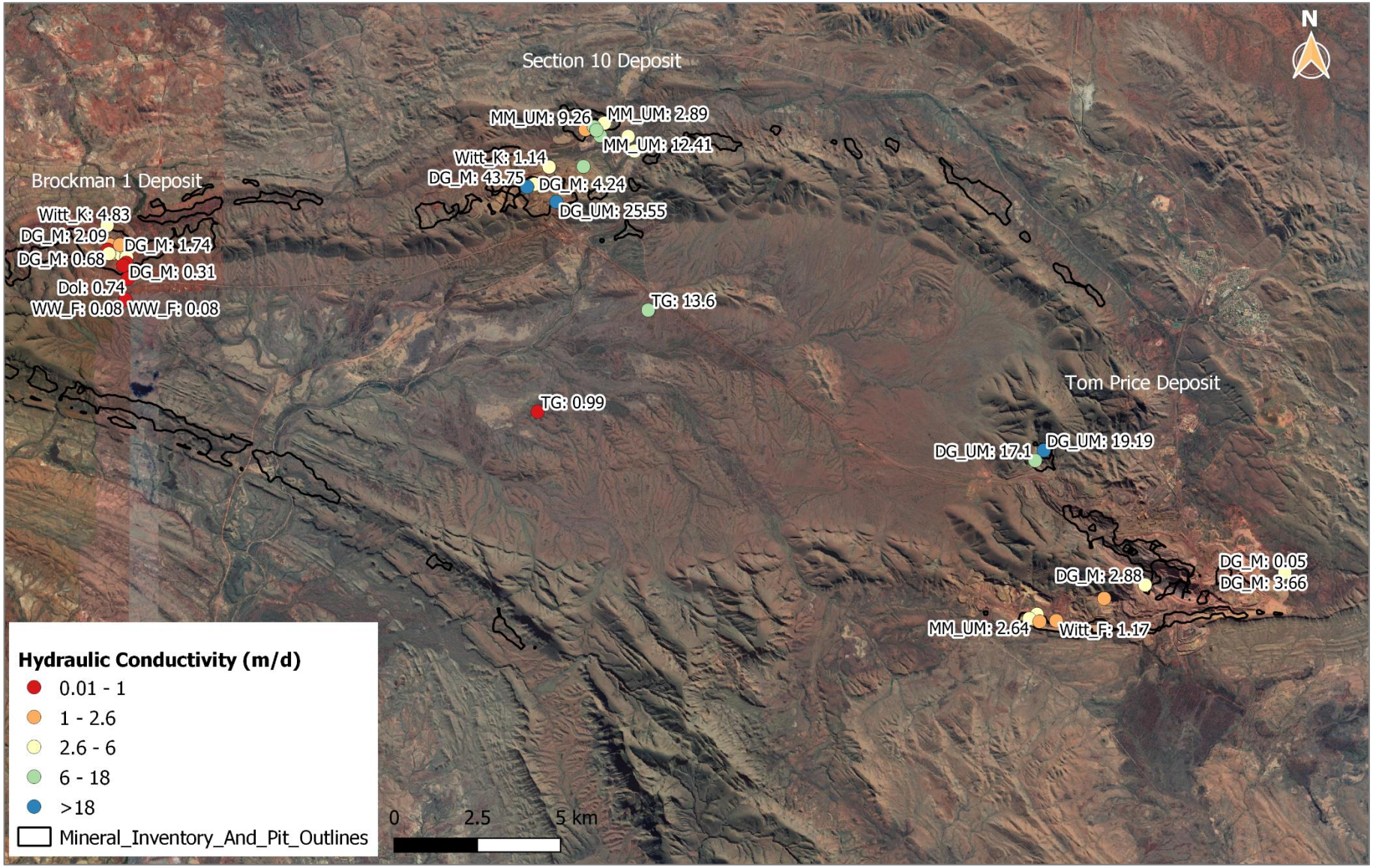

Figure 4.40 Tom Price, Section 10 & Brockman 1 (BIF & MMIF Deposits) – Aquifer Code: K (m/d) on each tested well

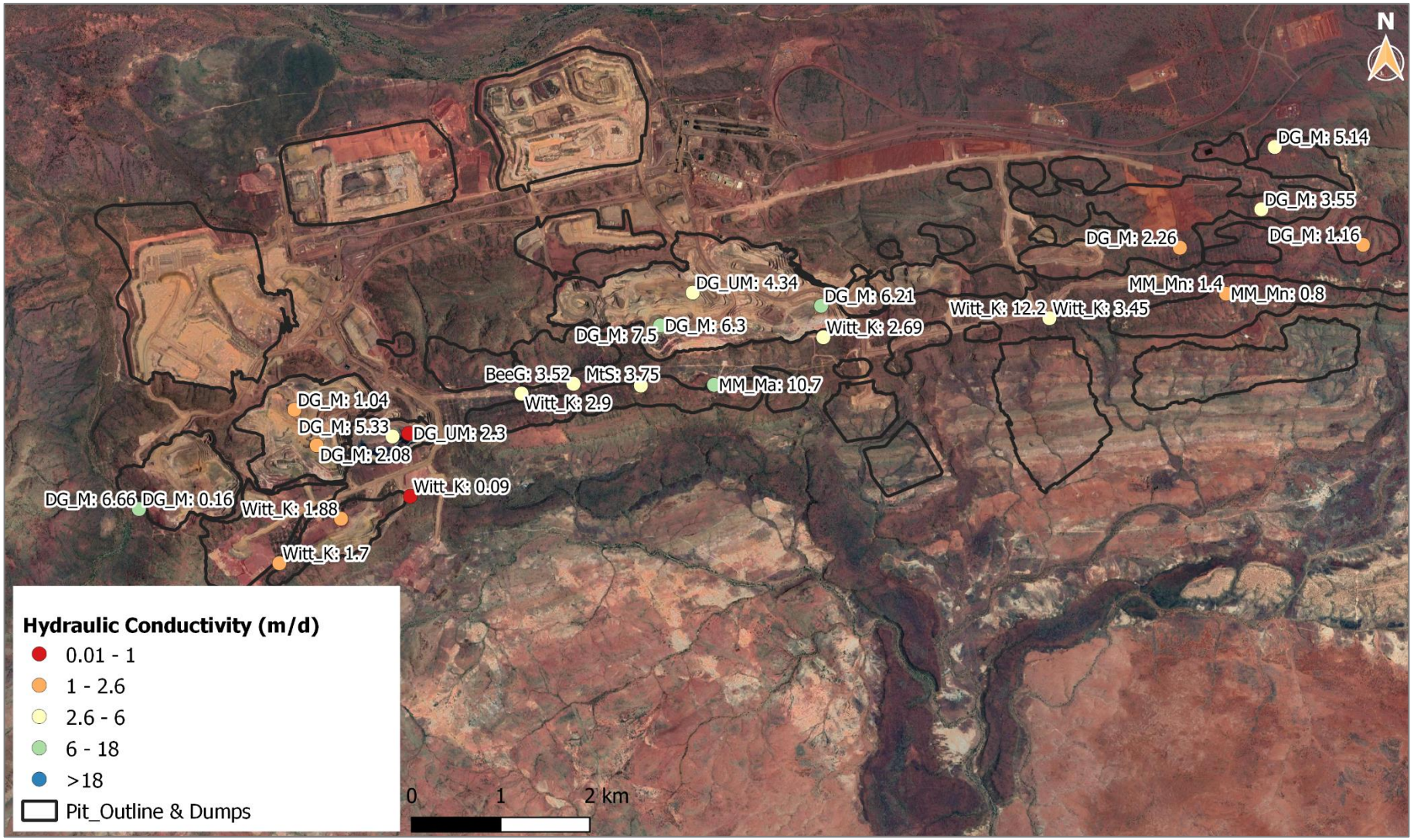

Figure 4.41 Brockman 4 (BIF Deposit) – Aquifer Code: K (m/d) on each tested well

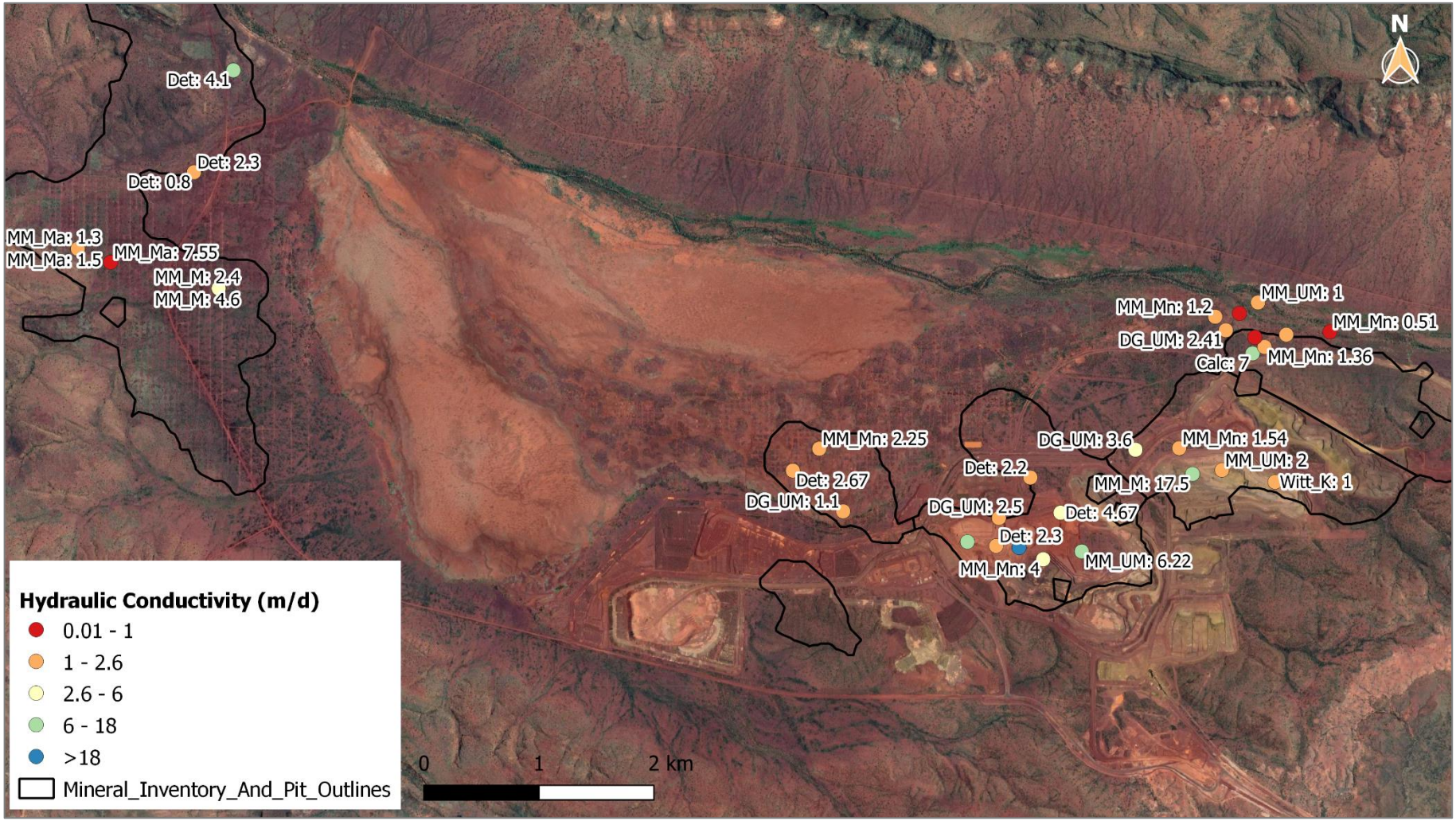

Figure 4.42 Silvergrass (MMIF Deposit) – Aquifer Code: K (m/d) on each tested well

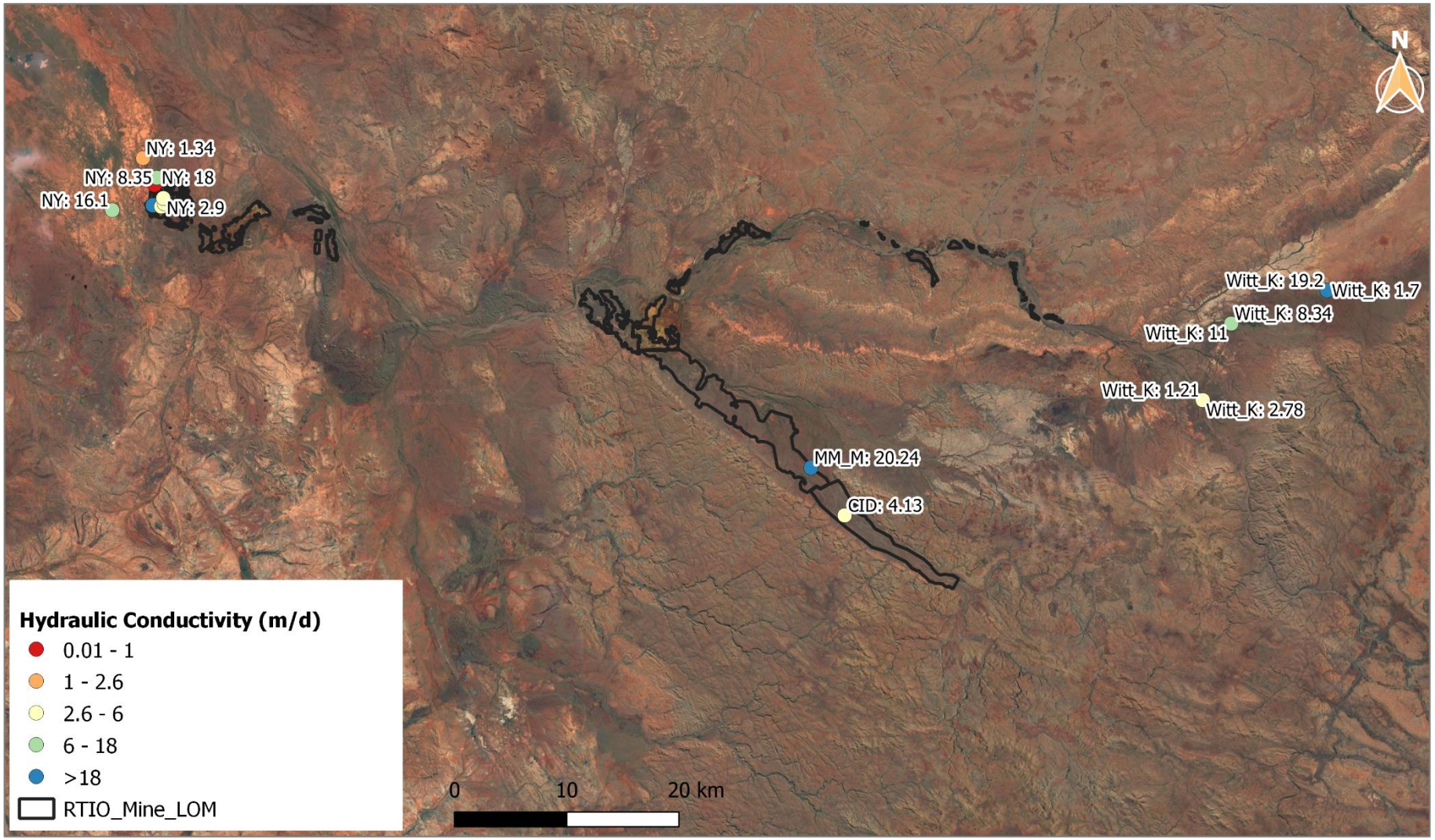

Figure 4.43 Robe Valley (CID Deposit) - Aquifer Code: K (m/d) on each tested well