Department of Exploration Geophysics

Integration of Geochemical and Geophysical Data for Downhole Rock Mass Characterisation

Maria Cornelia Kitzig

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

May 2018

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: ....

Date: 10-02-18

Abstract

Exploration for mineral deposits is more challenging today as the target depth for new discoveries is increasing. Deep exploration often requires drilling through several hundreds metres of barren cover, which is expensive and time consuming. As deeper holes are drilled it is imperative that each hole delivers as much information as possible—preferably in real-time—to allow changes to be made such as geo-steering and altering or updating of the drill plan whilst the drill rig is still on site. Typically, most drilling campaigns are conducted to obtain rock samples for geochemical analysis to measure the metal content. Thus, various technologies that enable real-time elemental analysis are being developed; for example: DET CRC's Lab-at-Rig® or Boart Longyear's TruScan®. Geophysical wireline logging could assist the real-time decision making, but is not widely used in base metals and gold exploration despite many obvious advantages to those familiar with such data. One issue is the cost of standby for a dedicated and skilled operator waiting for the drilling to finish; another is that the hole can collapse before logging is performed. Thus, DET CRC has developed driller-deployable geophysical borehole tools that provide petrophysical data as part of the drilling process (e.g., AutoSonde™and AutoShuttle). Both of these new measuring and sensing technologies offer a wealth of real-time data that will require timely analysis and interpretation to complement an effective exploration strategy.

This study demonstrates several benefits of joint analysis of interdisciplinary data from three different ore deposits and offers methods and workflows that enable real-time data interpretation. The core objective was to identify the necessary data combinations to classify the rock mass in terms of lithological, mineralogical and textural features. It is not the single link between one measurement and a specific rock property that is the focus, but the importance of combining the 'right' data for useful classification. What is 'useful' and necessary depends on the purpose or reason of the classification; for example, lithological distinction, ore-waste delineation or to highlight textural differences or rock quality properties. In addition, an emphasis is placed upon the use of petrophysical data as it offers high resolution real-time data but is less understood amongst geoscientists. Examples are given where petrophysical downhole logs can substitute for elemental assay data, or where petrophysical measurements are indicative of rock mass properties that are not detectable with elemental or mineralogical analysis.

Furthermore, this thesis shows that statistical analysis prior to classifying a data set can: firstly, give valuable insight into the interrelation of the data variables; secondly, provide a means for reducing dimensions of multivariate data sets; and thirdly, guide the optimal choice of variables for rock mass classification. This study also highlights the importance of necessary pre-processing steps for multivariate, cross-disciplinary data analysis and evaluates different clustering and prediction methods. The results indicate that a combination of two to three petrophysical downhole logs are sufficient for a range of possible classification purposes in various geological environments. For example, the natural gamma, resistivity and density logs sufficiently distinguish major lithology and identify the units of economic value at a sulphide deposit, whilst a combination that includes p-wave velocity data can separate quartz-rich, coarse grained from mica-rich, fine grained barren strata.

A second example from an iron ore deposit demonstrates the limitations of clustering algorithms when a classification according to strict, industry standard cut-off grades is desired and how this can be overcome by applying adaptive learning algorithms for variable prediction. This second example also indicates that real-time iron ore classification is possible with data extracted from Lithodensity logging, which enables cost effective exploration and resource evaluation campaigns through near real-time interpretation and decision making. The final case study highlights the possibility of predicting and interpolating rock quality properties from petrophysical downhole measurements on data from a nickel-copper deposit.

The results and findings of this research should encourage the wider application of geophysical downhole methods in future exploration campaigns and trigger a demand-pull from the industry for real-time measurement technologies and complementing data analytical workflows. These technologies and data analysis procedures are essential to ensure cost and time effective exploration for mineral deposits under deep cover in the future.

Acknowledgements

I would like to thank my supervisor Anton Kepic for providing me the opportunity of this research project and his continued support.

I am grateful to my panel members and the staff and students at the department of exploration geophysics for their support, guidance and friendship.

This study would not have been possible without the data sets, kindly provided by BHP, First Quantum Minerals, the Deep Exploration Technologies CRC and Boart Longyear.

I would like to acknowledge Curtin University and the Australian Federal Government for financial support through an Australian Postgraduate Award Scholarship (APA), and a Curtin Research Scholarship (CRS). Furthermore, the Cooperative Research Centre for Landscape Environments and Mineral Exploration (CRC LEME), and the Deep Exploration Technologies Cooperative Research Centre (DET CRC) for financial support through top-up scholarships and the Australian Society of Exploration Geophysicists (ASEG) Research Foundation for providing a three year research grant to perform field measurements, purchase the licenses for necessary software and to attend workshops and conferences.

The work has been supported by the Deep Exploration Technologies Cooperative Research Centre, whose activities are funded by the Australian Government's Cooperative Research Centre Programme. This is DET CRC Document 20../... .

The greatest gift in life is family and knowing that they will support you wherever you go and whatever you do—I love you.

List of Publications

Journal articles

Kitzig, M. C., Kepic, A., & Kieu, D. T. (2016). Testing Cluster Analysis on Combined Petrophysical and Geochemical Data for Rock Mass Classification. *Exploration Geophysics.*

Journal articles in review or accepted for publication

Kieu D. T., **Kitzig M. C.**, & Kepic A. () Prediction of Sonic Velocities from Other Borehole Data: An Example from the Kevitsa Mine Site, Northern Finland. *Geophysical Prospecting.*

Expanded abstracts published in conference proceedings

Kitzig, M. C., & Kepic, A. (2016, September). Automatic Classification of Iron Ore Lithologies Using Petrophysical and Geochemical Data. In Near Surface Geoscience 2016 - *First Conference on Geophysics for Mineral Exploration and Mining.*

Kieu, D. T., **Kitzig, M. C.**, & Kepic, A. (2016, September). Estimation of P-wave Velocity from Other Borehole Data. In Near Surface Geoscience 2016 - *First Conference on Geophysics for Mineral Exploration and Mining.*

Kieu, D. T., Kepic, A., & **Kitzig, M. C**. (2015). Classification of Geochemical and Petrophysical Data by Using Fuzzy Clustering. *ASEG Extended Abstracts*, 2015(1), 1-4.

Contents

List of Figures

List of Tables

List of Abbreviations

Measurements and units

AC, Ap.Charge	apparent chargeability
De, DEN	density
IP	induced polarisation
MS, MagSus	magnetic susceptibility
NG, Nat.Gamma	natural gamma radiation
Re, RES/16Res, 16N	resistivity\16-inch resistivity
SP	spontaneous potential
P-wave, Vp	pressure wave, primary wave velocity
S-wave, Vs	shear wave, secondary wave velocity
API	American Petroleum Institute units
g/cc	gram per cubic centimetre
keV	kilo electron volt
kg	kilogram
km/s	kilometre per second
ms	milli-second
mV	milli-volt
SI	base unit, International System of Units
ppm	parts per million
wt%	weight percent
Ω·m	ohm-metre
ICP-AES	inductively coupled plasma - atomic emission spectroscopy
ICP-MS	inductively coupled plasma - mass spectrometry
XRF	X-ray fluorescence
LOI	loss on ignition, measure of volatiles in assay samples

Statistics and Data Analysis

$\bar{x}$	median, centre of ordered sample set
$\tilde{x}$	mean, arithmetic average of sample set
StD, s	sample standard deviation

MaD	median absolute deviation
log	logarithmic transformed variable
inv	inverse of the variable
AHC	agglomerative hierarchical clustering
FCM	fuzzy c-means clustering
KM	k-medoids clustering
c	number of clusters of FCM
m	weighting exponent of FCM
α	cut-off value of FCM
PAM	partitioning around medoid
PCA	principal components analysis
FIS	fuzzy inference system
PCC	Pearson's correlation coefficient
RMSE	root mean squared error

Various

BIF	banded iron formation
SGG	spectral gamma-gamma ratio
RQD	rock quality designation
UCS	unconfined compressive strength
QC	quality control

Chapter 1

INTRODUCTION

Total exploration expenditure in Australia has fallen by 71% over the last five years, from around \$4,600 million in June 2012 to around \$1,300 million in March 2017 (Schodde, 2017). No new, significant discoveries were made for most commodities over the last two years, except for a few moderate sized gold finds and the rate of discovery is more closely related to the metres drilled than the total expenditure (Schodde, 2017). Thus, exploration efficiency is not increasing. A major challenge of making new discoveries is the increased depth of the cover, which masks the surface expression of mineral deposits traditionally used to find deposits. Since 2000, only 12 out of 469 gold discoveries worldwide were deeper than 500m and all of those were brownfield targets (Schodde, 2017). The exploration methodology used to make a discovery also tends to change with depth. The importance of geochemical methods declines rapidly from 0 to 50m with geophysical methods taking a greater role thereafter (Schodde, 2017). Regarding exploration below depths of 200m, geophysical methods, drilling (as the sole method), and extrapolating information from nearby deposits, are commonly used with about equal importance (Schodde, 2017).

Hence, to improve exploration success for deep targets, the industry needs to advance current tools and methods. Increasing the effectiveness of drilling campaigns is the main focus in this thesis, since drilling enables subsurface sampling and measurement of physical, chemical or mineralogical properties. Obtaining this information in a timely manner is crucial to characterising the target mineral systems in detail, which is a necessary step to make the next big discovery. At present, usually only a few holes or even just one ‘wild-cat’ hole are drilled into a geophysical ‘target’ and much of the obtained lithological, mineralogical and geochemical data is insufficient to fully understand the mineralising system.

Despite the many holes drilled upon geophysical targets most holes are not geophysically logged afterwards to better understand the geophysical response of the target. Often only core measurements are used to determine the target's properties, or the core may not be studied at all if the hole did not intersect anything 'useful'. Also, much of the more sophisticated elemental and mineralogical analysis relies upon core or rock chip testing done in calibrated laboratories off-site. Thus, information useful to understanding the mineral system is received after the drilling campaign is largely finished or the hole locations and trajectories have already been set. This lag of information creates a conundrum for the explorer: if the holes do not intersect encouraging mineralisation the next drilling campaign will not happen, but the limited window of time and site access plus rig costs associated with waiting can use the allocated exploration budget anyway. An exploration revolution needs useful information delivered in a timely manner to make decisions during the drilling campaign. Therefore, this thesis takes a forward view of current and possible future data streams that will be amenable to automation in data collection and to investigate how to process that information to aid geological interpretation in real-time—and possibly aid engineering and geo-metallurgy for future mining.

Measuring subsurface petrophysical properties and understanding their relation to lithology, mineralogy, alteration and structural/textural features are key for future successful greenfield exploration. Petrophysical logging is a process that collects high spatial resolution data that may be integrated with the drilling process. New drilling and sensing technologies, which aim to increase efficiency and reduce the cost of exploration drilling, are being developed and commercialised by the Deep Exploration Technologies Cooperative Research Centre (DET CRC). Besides a coiled tubing rig, adapted from the oil and gas sector for mineral exploration, new measuring and sensing technologies include autonomous downhole logging-while-drilling tools (AutoSonde$^{\text{TM}}$and AutoShuttle) and top-of-hole analytical instruments (Lab-at-Rig®). Hillis et al. (2014) demonstrate how implementing these technologies in conjunction with 'prospective drilling' may overcome many of the current challenges of exploring under deep, barren cover. These new technologies may provide a wealth of subsurface information at relatively low cost, adding valuable knowledge to existing exploration models. Furthermore, logging-while-drilling and top-of-hole sensing technologies provide a wide range of near real-time data, whose timely analysis and interpretation is critical for real-time decision-making. This also allows for cloud-based storage of petrophysical and geochemical data, available for near real-time interpretation and classification from almost anywhere in the world.

Successful exploration will require a better understanding of the relationships between physical and chemical, mineralogical or textural properties of rocks. Although petrophysics generally cannot directly map lithology in many hard-rock environments, the measured physical properties may be related to different rock-mass features of interest such as alteration style, rock quality/strength or texture. Several factors hinder the advancement of knowledge in this regard.

Firstly, barriers may exist between geologists, geophysicists and geochemists, inhibiting cross-disciplinary research and the exchange of knowledge across these geoscientific fields. While respective groups might work on related problems, this can often happen without communication or collaboration from within the same company or research initiative. Automated analysis that looks at data across the notional geological disciplines could circumvent some of the ‘silo’ problems in exploration, especially with smaller exploration companies that do not have a collection of experts, or even know whom to call.

Secondly, even with conventional drilling for mineral deposits, not all of the possible information (measurements) is collected. Often only sections of the drill holes are assayed, usually the mineralised portion; thus, missing the opportunity to gain useful insights into the mineralising system, possible halos and other spatially-related characteristics. Furthermore, although geophysical downhole logging is routinely implemented for some commodities such as iron ore, coal or uranium, a comprehensive suite of petrophysical measurements complemented by elemental assay data is rarely available. The unfamiliarity of such data to the geologists planning drilling campaigns tends to mean that such logging is seen as not sufficiently useful to make the effort. Someone is required to analyse the petrophysical data afterwards and if there is no plan to do this then there is no reason to collect this type of data. The lack of these comprehensive data sets further inhibits studying the links between petrophysical properties and related rock features in mineralising systems.

Geophysical logging has proven useful in several applications throughout the minerals industry, such as hole-to-hole correlation, ore body delineation, grade estimation or geotechnical characterisation. For example, natural gamma logging is used for stratigraphic correlation of iron ore deposits hosted in banded iron formations (BIF). Other methods may be equally useful in correlating lithology and rock mass characteristics in other stratiform mineral deposits. Wanstedt (1992) used density, magnetic susceptibility and natural gamma logs to delineate ore from waste at a sulphide-hosted base-metal deposit in Sweden and King et al. (1994) used conductivity logging at a mine in the Sudbury complex to discriminate ore from waste in blast holes. Later, McDowell et al. (2004) used conductivity logging in conjunction with magnetic susceptibility for nickel grade estimation in the Sudbury base metal deposits. The relationship of petrophysical measurements to grade is well established for natural gamma logging and uranium grade (Conaway and Killeen, 1978), as well as for magnetic susceptibility and iron grade in magnetite ore bodies (Virkkunen and

Hattula, 2000). Predicting rock mechanical and geotechnical properties from wireline logs was demonstrated by McNally (1990) and Elkington et al. (1982).

Other studies use a range of multivariate statistical methods to automatically determine rock properties from a combination of different measurements. Pechnig et al. (1997) used linear and multi-linear regression and factor analysis on wireline log data from the German Continental Deep Drilling Program to determine lithology, porosity and fracture intensity. Maiti et al. (2007) applied neural network modelling to the same data set for lithofacies classification. Lithofacies interpretation using principal components analysis and multi-level hierarchical clustering of wireline log data was demonstrated by Ma et al. (2014), while Urbancic and Bailey (1988) applied factor analysis to geophysical well logs to delineate favourable zones for gold mineralisation. Urbancic and Bailey (1988) also showed how geophysical log signatures relate to specific halos resulting from sericitisation and pyritisation and thus demonstrated how these measurements might help to vector towards a deposit. Fullagar et al. (1999) developed their own algorithm for automated rock-mass classification from geophysical borehole logs. Their algorithm uses centroids and distance measures to automatically group data, similar to conventional clustering algorithms. Templ et al. (2008) tested different cluster methods on regional geochemical data and offers a comprehensive discussion about the problems and possibilities of multivariate data analysis. The application of fuzzy c-means clustering to well-log data from the Ocean Drilling Program (ODP) to classify the rocks with respect to their magnetic properties was demonstrated by Dekkers et al. (2014). Imamura (1994) also used fuzzy c-means clustering to determine engineering properties from borehole data. Mahmoodi and Smith (2015) and Mahmoodi et al. (2016) applied fuzzy k-means and neural network modelling to borehole density, natural gamma and magnetic susceptibility log data to classify and predict lithology at the Victoria property of the southern Sudbury basin.

This thesis follows some of this work but builds a more generic framework for integrating various data sets and provides several novel examples of using the mixed data sets to classify and predict or interpolate missing properties of the rock.

1.1 Thesis Objectives

The aim of this research is to add to the existing knowledge about the relationship between petrophysical rock properties and their chemical and mineralogical composition and textural features. Petrophysical measurements are analysed in conjunction with geochemical and mineralogical data where possible and linked to specific rock mass characteristics. One objective is to show how to identify the best combination of measurements to classify the rocks for a specific purpose in different geological environments. For example, different classification objectives might be:

Distinguishing general lithology/alteration;

Separating ore from waste lithology;

Identifying different textural features of the rock mass;

Recognising a specific ore mineral;

Highlighting differences in rock quality properties.

A second objective is to identify important and necessary pre-processing steps for multivariate data analysis and assess the performance of different data clustering algorithms. For example, how important is prior knowledge of the statistical properties of the data; which tools may be useful in examining these properties and how can this knowledge add to a successful classification? What impact does the choice of clustering algorithm and parameters have on the results? How can the classification process be automated to complement real-time data acquisition and interpretation?

Another goal is to ascertain if any combinations of petrophysical downhole measurements can be substituted for elemental assay or geotechnical data for certain rock mass classification problems. As mentioned earlier, the implementation of downhole logging entails a range of advantages regarding data collection and quality. Data resolution is usually much higher than for elemental assay data and turn-around times for data availability are significantly reduced. This study demonstrates the advantages of collecting petrophysical downhole measurements from every drill hole with three selected case studies. The results may ultimately encourage some ‘demand pull’ from the industry towards utilising and optimising existing technologies for novel and widespread application.

Finally, the learnings from this research may be used to establish possible workflows to improve data utilisation for future exploration strategies. The workflow provides an example of a possible real-time data acquisition and interpretation scenario for time and cost effective greenfields exploration campaigns for targets under thick cover.

1.2 Thesis Outline

The thesis is organised in a way that reflects the structure of a possible workflow. Firstly, some summarised background information about commonly applied geophysical methods is given in Chapter 2. This chapter represents the minimum necessary prior knowledge to ensure meaningful data analysis and interpretation.

The methodology is outlined in Chapter 3 and first looks at the general structure of geoscientific data and how knowledge about their statistical properties and interrelationships can guide the subsequent analytical process. Important pre-processing steps are identified and compared using different data analysis and illustrative tools. All steps described in this chapter are performed on a comprehensive test data set collected at a historic sulphide mine and the same data is used for clustering and prediction processes in later chapters. A similar procedure may be applied to new exploration campaigns where historic data can be revisited to add prior knowledge and guide the choice of measurements and clustering method. For this study, three cluster methods are compared; of the three the fuzzy c-means method is chosen for subsequent cluster analysis. The impact of optimising the clustering parameters and three different methods to reduce dimensionality of data sets is also demonstrated in the methodology.

Chapter 4 concentrates on applying fuzzy c-means clustering to different combinations of input data with the aim to reproduce predefined classes. In the first example from the aforementioned sulphide deposit, cluster analysis is used to group the data in terms of major lithology as well as to distinguish sulphide minerals and textural features. The distinct combinations of input variables that successfully group the data through clustering are chosen to predict these rock mass characteristics on new data from the same deposit in Chapter 5. When adapted for exploration for new targets this should encourage the collection of as much information as possible from at least one or two initial drill holes to help determine what measurements are needed for certain classification objectives. This information can then be used to set-up and train an algorithm that automatically classifies the data acquired during subsequent drilling operations.

The second example of an iron ore deposit shows that cluster analysis may not successfully group data if the desired classes are not representative of the natural data structure. Iron ore classification is based on distinct cut-off values for certain grades of sellable ore that cannot be reproduced by clustering algorithms. Chapter 5 shows how neuro-adaptive learning algorithms can be used to overcome this issue. In addition, an experiment to substitute the elemental assay data with petrophysical logs was conducted to explore the possibility of automatic and real-time iron ore classification with either future wireline logging or a suitable measurement while drilling technology. The final example in Chapter 5 aims to test if neuro-adaptive learning algorithms may be used to predict one petrophysical property (sonic velocity), that is generally not measured, from a combination of others that are easier and quicker to collect; sonic logging is slower and requires a separate logging run as well as significant quality control (QC) processing afterwards. In this case, the predicted p-wave velocity is subsequently used to predict the rock quality designation, an important engineering property for mine and pit planning that is usually collected in the final phase of a drilling campaign.

Chapter 2

BACKGROUND OF COMMON GEOPHYSICAL AND GEOCHEMICAL MEASUREMENTS

The information presented in the following sections regarding the various petrophysical measurements is based on Telford et al. (1990) unless otherwise cited.

2.1 Electrical methods

2.1.1 Resistivity

Resistivity logging records a formation's resistance to the flow of electrical current passed into it from electrodes on the borehole tool. The depth of investigation is determined by the spacing between the measuring electrodes and the current electrodes, typically 16 inches for the short normal and 64 inches for the long normal device. Single-point resistance logs measure the resistance of a point in the borehole to an electrical ground on the surface. Resistivity logs can help delineate fracture zones and permeable formations since rocks are resistive and the fluid (water, brine) that saturates them is conductive. Saturated, porous and permeable sedimentary rocks are therefore highly conductive and show low resistivity readings. Exceptions are hydrocarbon fluids, which are insulators. High resistivity readings in oil exploration well logging may thus indicate hydrocarbon bearing formations. Metals, on the other hand, are good conductors of electricity and their log response to resistivity measurements is usually low. Moreover, the various major rock-forming minerals and thus the rocks themselves also have distinct electrical properties that may help distinguishing lithology by means of resistivity logging. For example, quartz has a higher resistivity than micas and quartz-rich lithology may be separated from mica rich rocks using resistivity logging (see Section 4.1.3).

2.1.2 Self-potential / spontaneous potential

The self-potential log (SP) is a passive measurement that records a naturally occurring spontaneous potential difference between a point in the borehole and a ground electrode on the surface. There are several mechanisms responsible for the build-up of a potential difference within a formation, such as electrochemical or kinetic potentials. An example of a kinetic potential is the so-called streaming potential, which is the result of pressure differences between the formation and borehole fluids. Electrochemical potentials arise when electrons or ions move through the rock due to chemical potentials; for example, a difference in salinity of the formation water and borehole fluid (shale potential and diffusion potential). A mineralisation potential occurs around sulphide rich formations, possibly due to a difference in oxidation potential between the rock and borehole fluid; however, the exact mechanisms are not well understood.

2.1.3 Induced polarisation

Induced polarisation (IP) measures the time it takes for a voltage to decline or increase after the flow of an electrical current is stopped or started. In the presence of metallic minerals an electrolytic reaction occurs at the interface of the mineral and the pore fluid. Due to the higher current velocities through the metal compared to the fluid, the build-up of charge at the interface results in an effect known as over-voltage. After the current flow stops the ions diffuse back to their equilibrium state, which results in the decay of this residual voltage. The typical time-domain measurement of the IP effect is the chargeability in milliseconds, which compares the integrated voltage decay to the steady voltage from the current-flow interval. The commonly used standard integration time after the current was cut off is the Newmont standard (t1 = 150ms, t2 = 1100ms). Although the IP effect is directly influenced by mineral concentrations it is further complicated by several other features such as the fluid content of the rock mass and the texture of the metallic minerals. Disseminated textures theoretically result in a higher IP signal compared to massive ones but fracture and vein fills can show higher readings as well. Nonetheless, IP surveys can add valuable information when exploring for disseminated sulphide deposits where other electrical methods give ambiguous or negative results.

2.2 Magnetic methods

2.2.1 Magnetic susceptibility

The magnetic susceptibility describes the ease with which a material can be magnetised, expressed by the ratio of the intensity of the magnetisation of the material to the intensity of the applied magnetic field. Ferrimagnetic minerals like magnetite, titano-magnetite, maghemite and pyrrhotite have the strongest influence on a rock's magnetic susceptibility, which is why mafic igneous rocks have a higher magnetic susceptibility (where magnetite is present as a minor or accessory constituent) than felsic igneous rocks or sediments (except for maghemite bearing detrital units). Pyrrhotite bearing sulphide deposits exhibit high magnetic signatures as well and it may be possible to infer the abundance of pyrrhotite from the magnetic susceptibility log.

2.3 Nuclear methods

2.3.1 Natural gamma

The natural gamma log records the natural radioactivity of the surrounding formation caused by the radioactive decay of mainly potassium ^{40}K, thorium ^{232}Th and uranium ^{238}U and their daughter products. The total count log, as the name implies, measures the total amount of gamma rays reaching the detector. The depth from which the gamma rays emanate—the volume of investigation—depends on the electron density of the rock mass and the initial energy of the gamma ray, since they interact and lose energy through scattering processes. A spectral gamma log counts gamma rays as a function of their energy and sorts them into bins or channels of certain energy ranges. They are useful in determining the concentration of K, Th and U in the rock with accuracies matching those of an elemental assay analysis. The natural gamma ray log is a good lithology indicator often used to delineate shale units in sedimentary environments for hydrocarbon exploration. Shales and their metamorphic equivalents contain potassium-bearing minerals such as muscovite and biotite micas as well as the clay mineral illite. The most abundant potassium bearing minerals are potassium feldspars such as orthoclase and microcline, which are important constituents in granites and pegmatites. Thorium and Uranium also occur in acidic igneous rocks, as primary minerals such as uraninite, thorite or monazite and as impurities in accessory minerals like zircon, apatite, titanite (sphene) and monazite. Pegmatites and granitic rocks thus exhibit a high gamma signature whereas basic igneous rocks show negligible radioactivity. Uranium is a highly soluble, mobile element easily carried through the rock in hydrothermal solutions and often deposited at considerable distances from the source. It is also readily oxidised and transported through groundwater and precipitated in some porous, reducing environments such as carbonate- or sulphide-bearing sandstones. Thorium and uranium accumulations in sediments also occur through weathering processes and concentrations of the heavy mineral fraction (monazite, zircon, etc.) as placer deposits in high-energy fluvial and coastal environments (Johnson, 1991).

2.3.2 Density

As previously mentioned, gamma rays interacting with a material may lose part or all of their energy through different processes, depending on their initial energy. For natural gamma logging, this initial energy is unknown. However in density logging a gamma source with a known energy is applied (gamma-gamma logging). The most commonly used source is the ^{137}Cs (caesium) isotope, which emits gamma rays at an energy of 662keV. The source and detector of the borehole tool are shielded by a thick lead section such that only gamma rays that interact with the surrounding formation can reach the detector. The gamma rays are recorded in the form of a spectrum relating to their energy and abundance. Most gamma rays with a specific source energy interact with the rock by Compton scattering, a process whereby the gamma ray loses part of its energy depending on the material's electron density. The density log, usually calibrated against a sandstone, limestone or dolomite matrix gives accurate bulk densities for these lithologies; however for most other rock types (igneous, metamorphic, metal bearing) it gives only qualitative results. Some of the lower-energy gamma rays interact with the formation by photoelectric absorption where all of the energy of the photon is used up to eject an electron from an atom's inner shell. This process depends on the element's atomic number as well as its electron density. Photoelectric absorption is therefore indicative of the formation's composition and thus lithology. More information on the use and operation of these processes are given in Section 5.3.

2.3.3 Neutron activation

The neutron activation methods apply a neutron source in the borehole, usually the californium 252 isotope ^{252}Cf, to irradiate the surrounding rock mass with fast neutrons. The neutrons interact with the nuclei of the surrounding medium via particle reactions, elastic and inelastic scattering, losing energy and slowing down to thermal energies in the process. Thermal neutrons captured by a nucleus are terminated and the target nucleus elevated to an excited state. During its decay to ground state the nucleus emits a gamma ray with an energy characteristic of that particular nucleus (Czubek, 1979). Therefore, neutron activation methods can detect the presence and abundance of different elements in the formation. Characteristic gamma ray spikes at distinct energies indicate the presence of chemical elements and the count rate (pulse height) indicate their abundance. Neutron activation logging can be used for downhole assaying but the technology is rarely implemented due to the risk of losing a radioactive source in the drill hole. However, the rewards of collecting high-resolution chemistry in-situ to complement petrophysical measurements greatly outweigh the risks. One of the problems of interpreting laboratory assay data and petrophysical downhole measurements is the different resolution and volume of investigation. Small scale variations of petrophysical properties can be lost when this data is resampled to match the assayed intervals. Thus, downhole assaying may be invaluable for a meaningful joint interpretation of geochemical and petrophysical data.

2.4 Acoustic methods

2.4.1 Sonic wave velocity, amplitude and full waveform

The borehole tool for seismic attribute logging comprises a transmitter that generates an acoustic signal, which is subsequently detected by differently spaced receivers to measure the interval travel-times of acoustic waves through the adjacent formation (and the borehole fluid). A sonic log records only the transit times of the waves whilst full waveform logging also records the amplitudes and the full waveform signals at a number of wideband receivers. The travel-time of acoustic waves through the rock depends on a range of factors, most notably density, but also texture, composition, abundance and orientation of planar features such as faults and fractures. While acoustic logs may distinguish these characteristics on a borehole scale, their main application is to characterise the rock mass for processing and interpretation of surface seismic data. However, downhole sonic data can provide valuable information about the engineering properties of the rock mass when the relationship is well defined for the particular geological environment. Such data is usually only collected from a fraction of available drill holes of a future mining prospect and the inferred rock property interpolated to fill the 3D volume for pit planning. If sonic logs were collected more regularly the mining risks associated with poor rock quality, fracture zones and faults could be greatly reduced.

2.4.2 Acoustic Televiewer

Ultrasonic pulses emitted by a rotating transmitter are reflected at the borehole wall and their amplitude and travel-time recorded. The resulting image provides information about fractures, breakouts, thin beds and their orientation. The data can also be used as a high-resolution calliper. As for sonic logging, acoustic televiewer data may be invaluable for pit planning, where knowledge about the 3D orientation of weak zones is of great importance. Drill core samples are often used to gather this information; however this data is prone to errors or may not be available when core recovery is poor. Some problems include: the core slipping in the core barrel during drilling, resulting in erroneous orientation; manual handling of the core leading to incorrect orientation lines drawn or erroneous depth fidelity.

2.5 Other measurements

Temperature and calliper measurements are routinely gathered with the other logs and used mainly for calibration, correction and processing of these logs. The calliper tool consists of one to four spring-loaded arms, pressed against the borehole wall when the tool is raised up, to measure changes in hole diameter.

2.6 Laboratory elemental assay data

Most quantitative laboratory assay methods determine a sample's chemical composition from spectral analysis. Commonly applied emission spectroscopy methods include X-ray fluorescence (XRF), inductively coupled plasma atomic emission spectroscopy (ICP-AES) and inductively coupled plasma mass spectrometry (ICP-MS). From the submitted sample (e.g., 1m half core from diamond drilling or ~3kg split of rock cuttings from reverse circulation (RC) drilling) only about 0.5-1g undergoes the analytical procedure. Elemental assay data is therefore only representative of a small volume of rock compared to most of the aforementioned petrophysical measurements. In addition, in contrast to the previously described petrophysical methods, which are in-situ measurements, assay data relies on the accurate recovery of rock material from a drill hole. The concentrations of the assayed elements are reported in weight percent (wt%) relative to the total sample weight. Thus, the respective abundances always sum to 100%, making assay data a form of closed data where the decrease in concentration of one element necessitates the increase in one or more of the others. This could introduce certain problems for statistical data analysis as outlined by Templ et al. (2008). Although it is possible to ‘open’ such data through different log-transformations (Templ et al., 2008), these methods were not applied in the present study because the resulting variables are not easily interpretable in terms of geochemical composition.

Chapter 3

METHODOLOGY

The general methodology of the data analysis process is explained using a real data set collected at the DET CRC (Deep Exploration Technologies Cooperative Research Centre) Brukunga Drilling Research and Training Facility test site. The drill site is located at the historic Brukunga sulphide mine in the Adelaide Hills in South Australia. The two major lithologies are unmineralised meta-sedimentary successions and pyrite- and pyrrhotite-bearing mineralised sections. The geology and mineralogy of the area is described in more detail in Section 4.1.1. where this data is used for cluster analysis. A comprehensive collection of geochemical and petrophysical data was acquired from drill hole DDH01 and the available measurements are summarised in Table 3.1. The data variables (petrophysical and geochemical) are pre-processed (filtered or interpolated where necessary) in WELLCAD (see Appendix I) and exported in matching depth intervals. Statistical analysis, data clustering and variable predictions are performed using MATLAB built-in functions or new code, which was written when existing functions were insufficient. The software packages and MATLAB functions and code used throughout this study are summarised in Appendix I.

3.1 Structure of geoscientific data

Examining the structure of a data set is a necessary and important step for any form of data analysis. Statistical properties and their visual representation provide important information about the data structure. For example, geoscientific data are often left or right skewed, can show bi- or multi-modal distributions or span several orders of magnitude. Logarithmic data transformations may be applied to some variables to obtain normally distributed data if the subsequent data analysis relies

Table 3.1: Geochemical assay data and geophysical downhole measurements available from drill hole DDH01 at Brukunga. This data is used to describe the methodology and is also used for cluster analysis in the later chapters.

Geochemical assay data	SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, Na_2O, K_2O, MnO, TiO_2, P_2O_5, Cr_2O_3, LOI (loss on ignition), total C and total S in wt% and Ni, Cu, Zn and Pb in ppm
Petrophysical downhole measurements	Density (De) [g/cc], s-wave (Vs) and p-wave (Vp) [km/s or m/s], Natural Gamma (NG) [API], Spontaneous Potential (SP) [mV], 16-inch\Resistivity (16 Res\Re) [Ω·m], Apparent Chargeability (AC) [ms], Magnetic Susceptibility (MS) [SI]

on such distribution. Some of the basic statistical properties are summarised and explained below:

minimum/maximum (x_{min}/x_{max}) The minimum and maximum are the extreme values of a set of sample values. The **range** is defined as $max - min$ and is a measure of the spread of the data. The range is an example of a non-robust statistical property because it is strongly influenced by outliers.

median ($\tilde{x}$) The median is the centre of an ordered set of sample values (i.e., $x_1 \leq x_2 \leq \cdots \leq x_n$), where 50% of the sample values are smaller than (or equal to), and the other 50% of the sample values are larger than (or equal to) the median. The first quartile (Q1 or 25th percentile) is the median of the first half of the data; the third quartile (Q3 or 75th percentile) is the median of the second half of the data. The values between Q1 and Q3 make up the interquartile range (IQR). The **median absolute deviation** (MaD) is defined as the median of the absolute deviations of the samples from the median. $MaD = \widetilde{|x_i - \tilde{x}|}$. Statistics based on median calculations are considered robust, because these measures are little influenced by outliers.

mean ($\bar{x}$) The mean is the arithmetic average of a set of sample values. It is a measure of the centre of gravity of the data and is strongly influenced by extreme outliers. The mean also represents the first of the mathematical moments of the data. The moments describe the distribution of the data and thus the shape and symmetry of the probability density function. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, where x_i is the i^{th} sample and n is the total number of samples.

standard deviation (s) The standard deviation of a set of sample values indicates how dispersed the data is around the mean. The standard deviation is the square root of the **variance**, which in turn represents the second central moment around the mean. $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

skewness (γ) The skewness is the measure of asymmetry of the data distribution. Negative skewness indicates a tail to the left and positive skewness indicates a tail to the right of the mean. However, it is more a measure of balance because the value for the skewness can be low (perfectly normal distributed data has a skewness of zero) if the tails balance each other out. This may be the case when the distribution shows a long thin tail to one side and a short fat tail to the other. Skewness for multi-modal distributions may also give insignificant values. The skewness of a sample is the third standardised moment about the mean. $\gamma = (\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3)/s^3$.

kurtosis (κ) The kurtosis is the measure of the shape and thickness of the tail of skewed data. It is indicative of infrequent but extreme outliers in the sample distribution. The value for the kurtosis of a univariate, normal distribution is 3. The kurtosis is the fourth standardised moment about the mean. $\kappa = (\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4)/s^4$.

Table 3.2 and Table 3.3 show the statistical properties for the petrophysical measurements and elemental assay data of the test data set from drill hole DDH1 respectively. These statistical properties can be represented visually in the form of box plots or histogram plots for a more intuitive interpretation. The box plots in Figure 3.1 show the data distribution around the median, represented by the centre line in each box. The bounds of the box are the first and third quartiles (Q1 and Q3, 25th and 75th percentiles) of the data variable. Asymmetry of the centre line with regards to the bounds of the box and its whiskers indicates the direction and amount of skew of the data distribution (e.g., total carbon and sulphur, resistivity, magnetic susceptibility etc.). The red crosses represent outliers, resulting in long thin tails of the distribution function for the respective variable. This is reflected in high values for the kurtosis. In this example, the MnO box plot shows a long tail of outliers relative to the range of the main group of data and its value for kurtosis (Table 3.3) is the highest of the data set.

The box plots in Figure 3.1 indicate that many of the data variables are skewed and thus not normally distributed. Using the box plots as a guide, a log transformation may be a reasonable pre-processing step for CaO, MgO, K_2O, MnO, TiO_2, Cr_2O_3, total carbon, total sulphur and the magnetic susceptibility of the dataset. It is also possible to use the values of the calculated statistical properties as a guide. A log transformation is recommended where the standard deviation exceeds the mean or where the skewness is greater than one. The respective values are highlighted in bold in Table 3.2 and Table 3.3.

Although box plots provide an idea about the geometry of the data distribution,

Table 3.2: Statistical properties for the petrophysical data. The magnetic susceptibility shows the highest value for the kurtosis and skewness, indicating an asymmetric, log-normal distribution which is the common distribution for this measurement (Lapointe et al., 1986; Latham et al., 1989). Asymmetric distribution for the resistivity and apparent chargeability data is indicated by standard deviations exceeding the mean. The indicators for asymmetric distribution are highlighted in bold.

Property	De	Vs	Vp	NG	SP	16Res	AC	MS
Min	2.2	2.5	3.8	0	–190	5.26	0.0002	0.0001
Max	3.6	3.7	7.2	311	373	1036	0.0276	0.20
Mean	2.8	3.1	5.4	113	133	**285.8**	**0.0055**	**0.007**
StD	0.2	0.2	0.7	58	106	**342.5**	**0.0061**	**0.016**
Median	2.8	3.1	5.3	104	113	41.9	0.0022	0.003
MaD	0.1	0.2	0.6	38	70	33.0	0.002	0.003
Skewness	0.6	–0.1	0.1	0.6	0.03	0.8	0.9	**7.3**
Kurtosis	4.3	2.7	1.9	3.3	2.8	2	3.2	**77**

StD = Standard deviation; MaD = Median absolute deviation

Table 3.3: Statistical properties for the elemental assay data of the test data set. High values for kurtosis and skewness indicate asymmetric data distribution. A standard deviation exceeding the mean also indicates asymmetry or log-normal distribution. Examples of asymmetric distribution are highlighted in bold.

Property	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	Na_2O	K_2O
Min	42.2	9.1	3.0	0.6	1.2	1.0	0.2
Max	78.9	18.3	26.1	11.9	7.9	3.3	4.9
Median	58.5	14.3	8.1	1.7	2.6	2.1	3.0
MaD	7.8	1.6	3.3	0.4	0.4	0.4	0.4
Mean	60.5	14.4	10.0	2.3	2.8	2.1	3.0
StD	9.3	2.1	5.8	1.9	1.1	0.5	0.7
Skewness	0.2	0.0	0.8	**3.1**	**2.2**	0.1	−0.8
Kurtosis	1.9	2.3	2.4	13.2	9.2	2.1	4.8

Property	MnO	TiO_2	P_2O_5	Cr_2O_3	LOI	C_{tot}	S_{tot}
Min	0.03	0.4	0.07	0.00	0.7	0.02	0.0
Max	1.59	1.4	0.24	0.06	9.6	0.48	15.2
Median	0.08	0.6	0.13	0.01	2.6	0.07	2.0
MaD	0.02	0.6	0.02	0.002	1.4	0.03	2.0
Mean	**0.11**	0.6	0.14	0.02	3.3	0.10	**4.0**
StD	**0.14**	0.1	0.02	0.01	2.5	0.09	**4.5**
Skewness	**7.98**	**2.5**	0.57	**2.98**	1.0	**1.69**	0.9
Kurtosis	**78.81**	**11.8**	4.61	**12.84**	2.9	5.80	2.5

StD = Standard deviation; MaD = Median absolute deviation

they do not show the mode. The mode of a data variable indicates the number of peaks in the distribution. For example, if a set of geological samples contains an equal number of granite and basalt samples then silica content (amongst others)

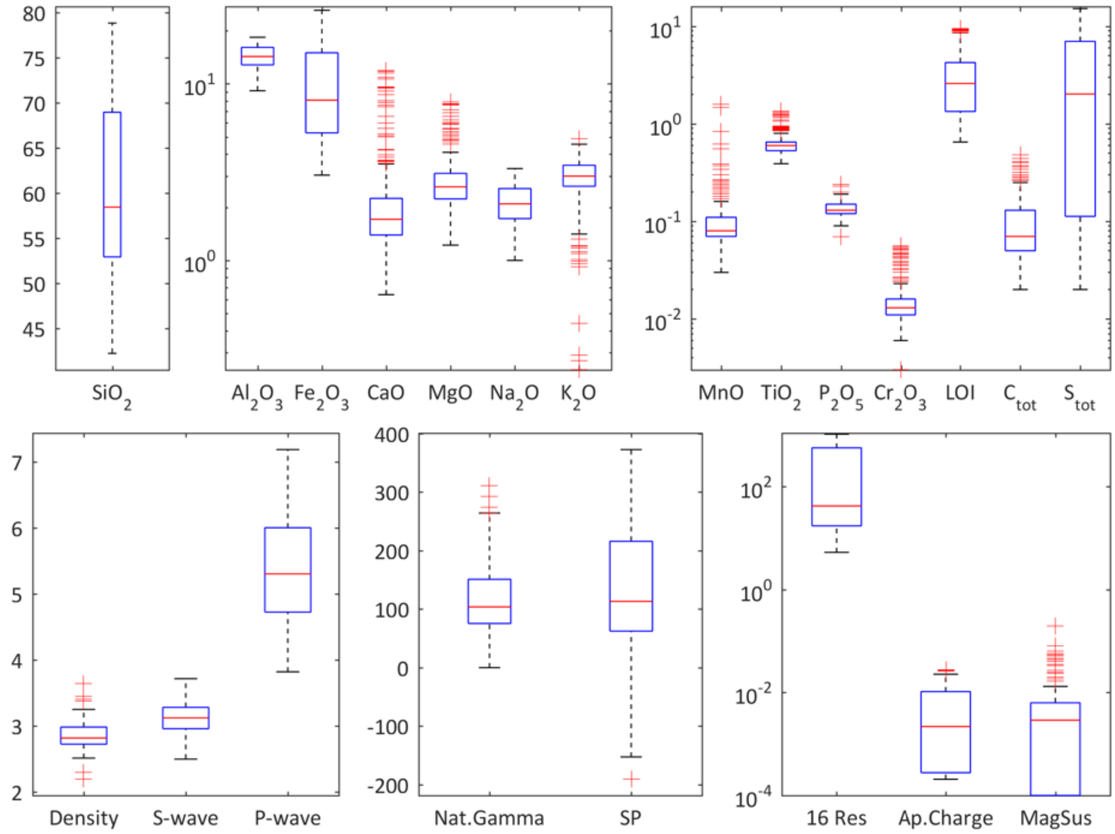


Figure 3.1: Box plots show basic statistics for the data variables. Note the different scales of the grouped plots. The centre line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme values that are not considered outliers and the red crosses are outliers. Long tails of outliers (e.g. CaO, MgO MnO and K_2O) indicate the direction and amount of skew in the data.

shows a bimodal distribution, one peak of higher values for the granites and one peak of lower values for the basalt. Histogram plots are a better tool to visualise these multi-modal distributions. Figure 3.2 shows histogram plots of the same data used to create the box plots plus additional plots of the log-transformed variables. The histograms clearly indicate a bimodal distribution for SiO_2, Fe_2O_3, LOI and p-wave velocity. After log transformation, this is also evident for the total sulphur assay data and the resistivity, magnetic susceptibility and apparent chargeability measurements.

Although a log transform might help to see the data structure more clearly in histogram plots, it can also negatively influence the result of subsequent cluster analysis because it distorts the data or disperses it over a larger absolute range (statistical range; maximum minus minimum value, see previous description). Clustering algorithms use distance measures to find clusters and group data and changing the relative position of the sample points to each other may lead to different cluster solutions. Therefore, it seems prudent to decide on a case-by-case basis as to whether a log transform is necessary and if in doubt, test both data sets and compare.

The information obtained from statistical data analysis is also useful to decide on subsequent data pre-processing steps that may be necessary to ensure successful data clustering. Examples of possible pre-processing procedures are described in the following section.

3.2 Data pre-processing: standardisation versus normalisation

Geoscientific data sets usually include variables with different units of measurement and scales and the values of some of the measurements can span several orders of magnitude (see min/max values in Tables 3.2 and 3.3). As mentioned previously, most clustering algorithms use some form of distance measure to group data. It is important to transfer the different variables to a common (comparable) space before analysis so that no unwanted weighting is introduced due to differences in absolute values of the respective variables. For example, the range of resistivity values are frequently in the order of $10^4\Omega\cdot$m, whereas the density rarely exceeds 4g/cc for common rocks. Or, depending on the detection limits of the applied tools, the magnetic susceptibility can take on values below 10^{-5} and span several orders of magnitude. It is also recommended to ‘clean’ the data of any extreme outliers that may represent unreasonable values for the respective measurement such as negative or extremely high values.

Several methods exist to transfer data of different scales to a comparable space. The most common method to standardise data is by calculating the z-score (standard score) for each variable by subtracting the mean from each sample and dividing by the standard deviation (see formula below). A different method uses the median instead of the mean, and the median absolute deviation instead of the standard deviation to calculate the standardised values. This method is called m-standardised, or m-score throughout this study. Normalisation is a type of feature scaling where the normalised variable is scaled to a certain range, usually to between 0 and 1 or -1 and 1. The three data transformations are calculated according to the formulas below:

z-standard score: $$x_{new} = \frac{x_i - \bar{x}}{s}$$

m-score: $$x_{new} = \frac{(x_i - \tilde{x})}{MaD}, \qquad \text{where:} \quad MaD = (\widetilde{|x_i - \tilde{x}|})$$

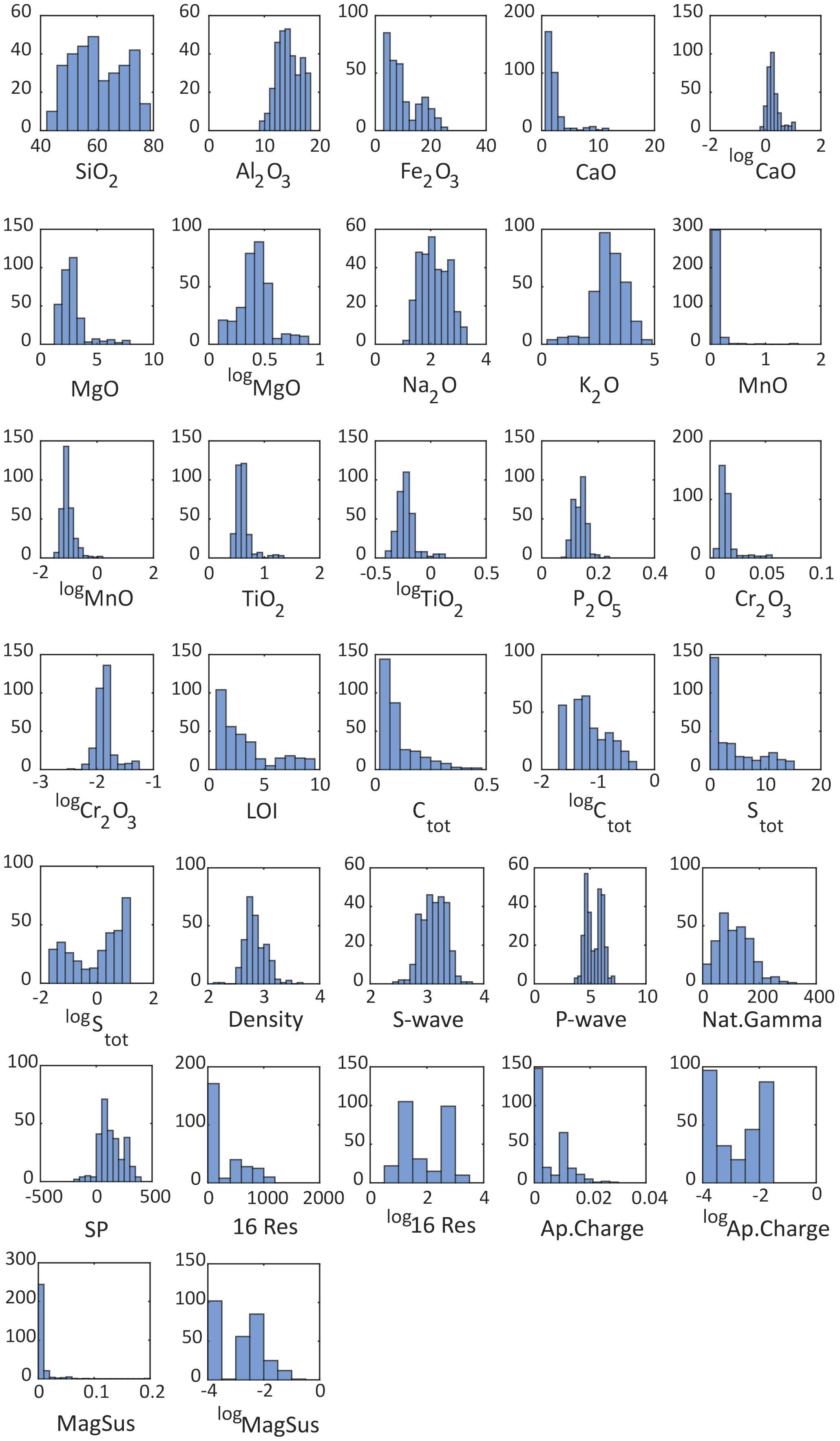


Figure 3.2: (caption on next page)

normalised: $$x_{new} = \frac{(x_i - x_{min})}{(x_{max} - x_{min})}$$

Where: x_{new} is the new sample point, x_i is the original sample, $\bar{x}$ is the mean, s is the standard deviation, $\tilde{x}$ is the median and MaD is the median absolute deviation.

The respectively transformed petrophysical variables are compared in box plots in Figure 3.3. The z-score centres the variable on its mean and expresses the distance of a sample from the mean in terms of the standard deviation. The standardised variable has a mean of zero and a standard deviation of one (Figure 3.3, top). Similarly, a variable standardised via the m-score has a median of zero and a median absolute deviation of one (Figure 3.3, middle). Normalisation scales the data such that the minimum value is set to zero and the maximum to one (Figure 3.3, bottom). The shape properties of the variables (e.g., skewness, kurtosis) are retained in all methods; however, due to the bounded nature of normalisation the resulting smaller standard deviations can suppress the effect of outliers. If extreme outliers are present then the main group of the data will be squeezed into a rather small range, which could introduce bias and have an undesired effect on the performance of distance-based clustering algorithms. The same may be true for m-standardised data because its calculation uses the median and median absolute deviation, both of which are robust statistics that are not influenced by outliers (see previous section). However, the information contained in outliers may be of vital importance when analysing geoscientific data for minerals exploration. For example, an exploration drill hole that has a total length of 100m may have intersected 2m of mineralised rock. If sampled in 1m intervals, these two metres only represent 2% of the total data, and would be classified as outliers that may be ignored in subsequent analysis if an inappropriate standardisation method was applied.

The different data transformations also affect the apparent interrelationship of the variables differently. One way of illustrating the relationship between variables are dendrograms, another is to calculate correlation coefficients. Dendrograms, or tree-diagrams, illustrate the grouping of data from hierarchical clustering where

Figure 3.2: Histogram plots of the original and log-transformed variables. The plots provide an easy means of examining the basic data structure. SiO_2, Fe_2O_3, Vp and others show a bimodal distribution while CaO, MnO and the magnetic susceptibility are strongly left skewed. Potassium shows a somewhat right-skewed distribution. After log transformation resistivity, apparent chargeability and magnetic susceptibility data also show a bimodal distribution.

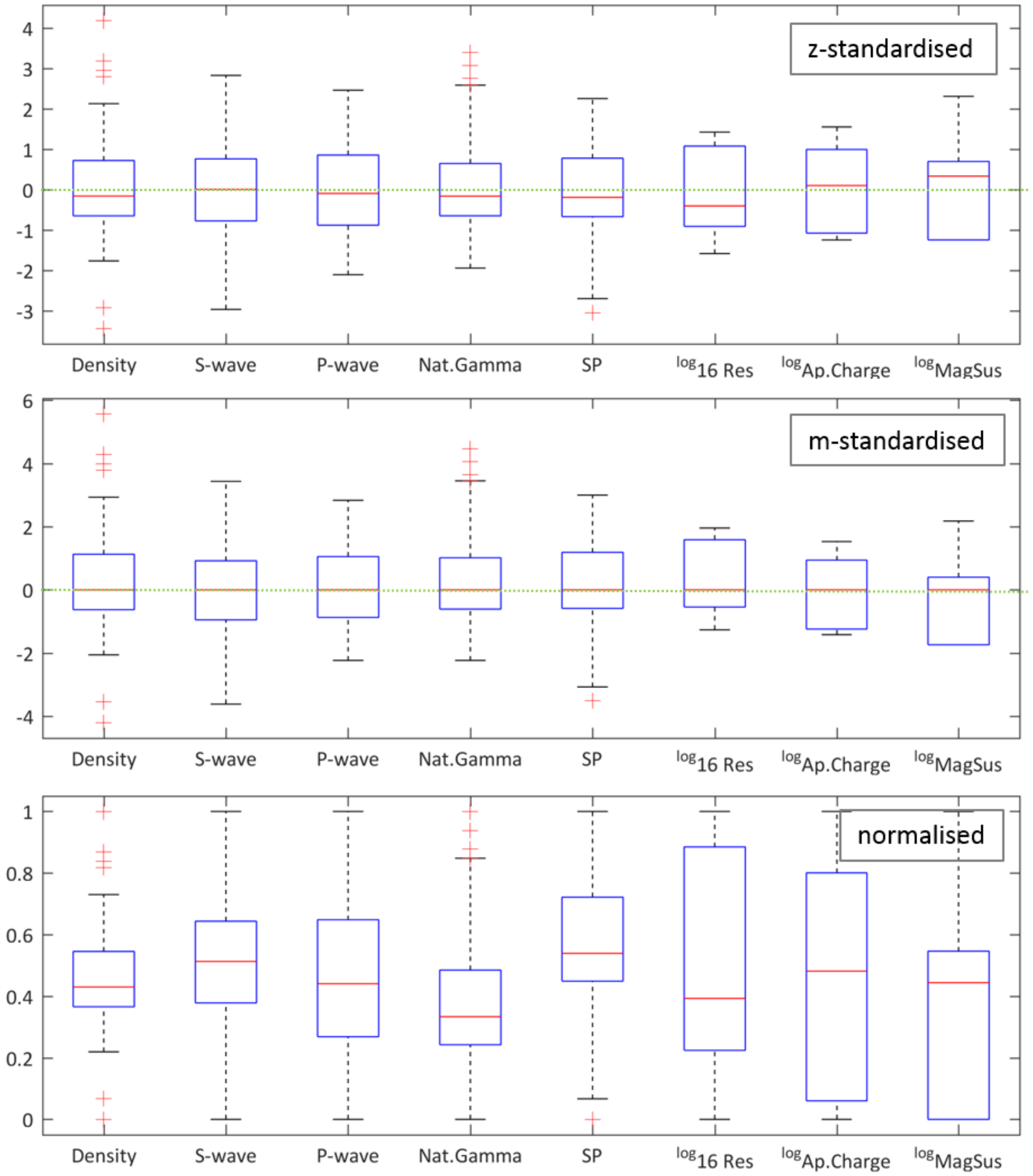


Figure 3.3: Box plots of the petrophysical variables after applying the different data transformations. Top: z-standardisation centres the data around the mean (green dotted line), middle: m-standardisation centres the data around the median (green dotted line coincides with median line in each box), bottom: normalisation scales the data between 0 and 1. The shape properties of the data distribution is retained, but the data is squeezed into different absolute ranges (compare y-axis).

similar or correlated variables are linked early in the tree. The height of the links of the individual branches represents the dissimilarity (or distance) between the variables or groups of variables. Shallow links indicate closer relationships between the respective variables while high links signify poor or inverse correlations.

Figure 3.4 and Figure 3.5 show dendrograms from hierarchical clustering of the raw data *variables*, compared to dendrograms from clustering of the log-transformed, standardised and normalised variables (dendrogram for m-standardised variables are only shown for the assay data; dendrogram of log-transformed variables are only shown for the petrophysical data). The scale of the y-axis provides a direct measure of similarity between the variables, with larger numbers indicating lesser correlation in general. The absolute range of values on the y-axis is largest for the raw data ($0–10^4$) and smallest for the normalised data (0–10) because normalisation scales the values to a comparatively similar range of 0 to 1 compared to about 10^{-4} to 10^2 for the raw data (Figure 3.1 compared to Figure 3.3).

The absolute values of the data variables seem to have more influence on how they are grouped in dendrograms than their actual relationship in terms of rock properties. For example, the raw petrophysical data variables are ordered according to their absolute values with smallest values for the apparent chargeability and highest values for the resistivity (compare to Table 3.2). After log transformation the order of the variables has changed, but they are still arranged from smallest to largest values. This arrangement, or linkage, does not reflect the relationship of the variables and subsequent cluster analysis of raw data can result in insignificant classifications. The linkage of the z-standardised and normalised variables (Figure 3.4) better reflect their relationship. For example, Vp and Vs are always linked early as would be expected, since these variables are highly correlated. Density and spontaneous potential indicate sulphide mineralisation as do the magnetic susceptibility and apparent chargeability, but the latter highlight a pyrrhotite-rich, disseminated mineralisation style. In unmineralised rocks, the natural gamma and resistivity signals are higher than in the mineralised parts and both variables are linked together when z-standardisation is applied.

For comparison to the linear correlations between pairs of variables, the Pearson's correlation coefficients were calculated for the raw and log-transformed variables.

Table 3.4 shows the correlation coefficients for the petrophysical data on the lower triangle and the corresponding p-value on the upper triangle. The p-value is the result of testing the null hypothesis, stating that there is no meaningful correlation between the pair of variables. A value below 0.05 considers the relation as significant. Table 3.5 shows the correlation coefficients and corresponding p-values after log transformation of the resistivity, apparent chargeability and magnetic susceptibility data. The strong linear correlation between magnetic susceptibility and apparent chargeability is only evident after log transformation and their relationship would not be identified from investigating correlation coefficients of the raw data. However, these two variables are linked early in all dendrograms, regardless of prior transformations, which demonstrates the advantage of using dendrogram analysis to identify the relationship between not normally distributed variables.

Table 3.4: Correlation coefficients (bottom triangle) and p-test values (top triangle, grey) of the raw petrophysical data. A p-value below 0.05 considers the corresponding correlation as significant. P-values of insignificant correlations shown in italic, the strongest correlations are highlighted in bold. The natural gamma data does not show a significant correlation to any other variable; resistivity shows inverse correlations.

	De	Vs	Vp	NG	SP	Re	AC	MS
De		0.00	0.00	*0.75*	0.00	0.00	0.00	0.00
Vs	0.41		0.00	0.04	0.00	0.00	0.00	*0.24*
Vp	0.46	**0.73**		0.03	0.00	0.00	0.00	0.00
Ga	−0.02	−0.12	−0.13		*0.14*	*0.58*	*0.82*	*0.86*
SP	**0.53**	0.30	0.43	−0.09		0.00	0.00	0.00
Re	−0.50	−0.20	−0.21	−0.06	−0.48		0.00	0.00
AC	0.46	0.22	0.25	0.01	**0.52**	−0.65		0.00
MS	**0.34**	0.07	0.21	−0.01	0.31	−0.30	0.28	

Table 3.5: Correlation coefficients (bottom triangle) and p-test values (top triangle, grey) of the log-transformed petrophysical data. P-values of insignificant correlations shown in italic, the strongest correlations are highlighted in bold. In contrast to Table 3.4, the magnetic susceptibility is now strongly correlated to apparent chargeability. This relationship is also evident in all dendrograms in Figure 3.4.

	De	Vs	Vp	Ga	SP	Re	AC	MS
De		0.00	0.00	*0.70*	0.00	0.00	0.00	0.00
Vs	0.41		0.00	0.04	0.00	0.00	0.00	*0.24*
Vp	0.46	**0.73**		0.03	0.00	0.00	0.00	0.00
Ga	−0.02	−0.12	−0.13		*0.12*	*0.71*	*0.85*	*0.93*
SP	0.53	0.30	0.43	−0.09		0.00	0.00	0.00
Re	−0.55	−0.26	−0.27	−0.02	−0.49		0.00	0.00
AC	**0.55**	0.25	0.29	0.01	**0.55**	−0.94		0.00
MS	0.50	0.24	0.37	−0.01	0.48	−0.79	**0.79**	

Comparison of the dendrograms and correlation coefficients also shows that the z-standardised variables match best with the correlation coefficients of the log-transformed data (Figure 3.4 and Table 3.4). The shallowest link in the dendrogram is that between apparent chargeability and magnetic susceptibility whose correlation coefficient is the highest (0.79); Vp and Vs are linked next and their correlation coefficient is the second highest (0.73). The natural gamma and resistivity are linked very late to the rest of the data, which is reflected in the low or negative correlation coefficients respectively.

Figure 3.5 shows category dendrograms for the raw assay data compared to the normalised and standardised data. The absolute values of the y-axis are again largest for the raw data and smallest for the normalised data. Most chemical elements are grouped similarly regardless of which pre-processing step was applied, except for manganese (MnO) and carbon to some extent. Manganese shows a log-normal distribution (Figure 3.2) but was not log transformed prior to standardisation and normalisation. Grouping of the normalised data links manganese with carbon and then with elements indicative of basic (mafic) rocks (the chemistry and petrology of the rocks at Brukunga is described in Section 4.1.1); z-standardisation leads to a similar grouping, but carbon is now linked with sulphide mineralisation; m-standardisation highlights manganese as an outlier as it is linked to the rest of the data very late in the tree. The significance of this is discussed in detail in Section 3.5.

3.3 Comparison of different cluster methods

Many different strategies and algorithms exist to cluster and group data. The following section compares three examples of cluster algorithms: agglomerative hierarchical clustering (AHC), fuzzy c-means (soft, mean-based) clustering (FCM) and k-medoids (hard, median-based) clustering (KM). All three methods use the Euclidean distance metric but their respective partitioning (grouping) method differs.

An example of hierarchical clustering of data *variables* was shown in the previous section. The method is now used to cluster the *samples*. **Agglomerative hierarchical clustering** starts by assigning each data point its own cluster and then links these clusters in a step-wise (hierarchical) manner based on a specific distance calculation until all points belong to the same cluster. From the various methods to calculate the linkage (distance) between the data, the *group average linkage* is the most widely used method in hierarchical clustering problems. There, the linkage is calculated based on the average distances between samples and clusters. Another method of linking data is the so-called *Ward's linkage*. Ward's linkage combines clusters and points to them by minimising the within group sum of squares of the distances to the respective centroid, thereby minimising the within-cluster variance. Ward's linkage seems to be a sensible choice for rock-mass classification purposes, since the variance of different lithologies or rock classes (within group variance) should be quite small compared to the overall variance of the data. Ward's distance measure is calculated as follows:

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r + n_s)}} \left\| \bar{x}_r - \bar{x}_s \right\|_2 ,$$

> where $\| \, \|_2$ is the Euclidean norm, $\bar{x}_r$ and $\bar{x}_s$ are the centroids of clusters r and s, n_r and n_s are the number of elements in clusters r and s.

Both methods were tested but the group average linkage was dismissed immediately

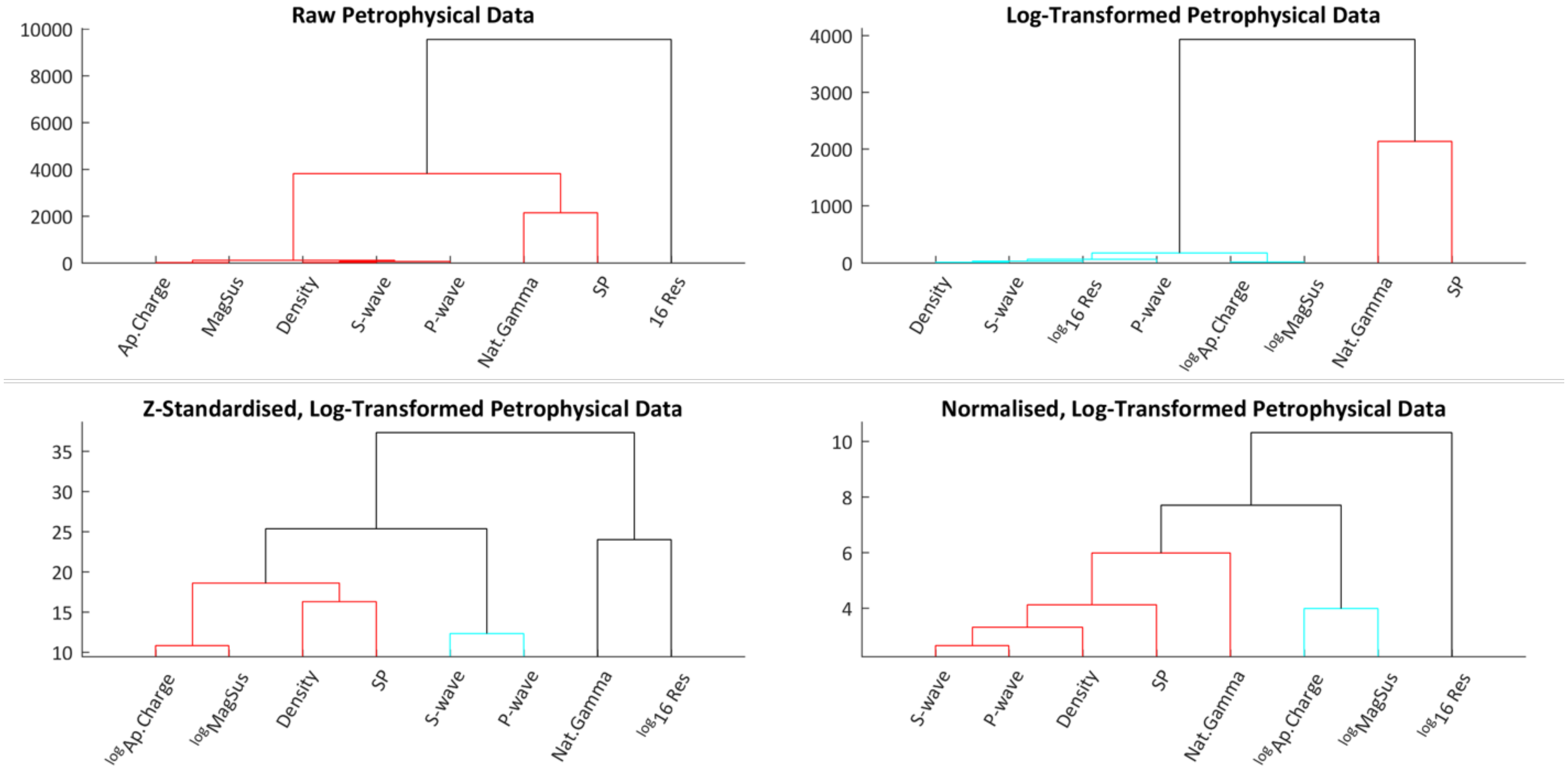


Figure 3.4: Category dendrograms for the different transformations of the petrophysical data. The y-axis (height of the links) represents the dissimilarity or distance between the variables (categories) or variable groups (clusters). The linkage of variables differs depending on which pre-processing strategy is applied. The raw and log-transformed variables are ordered based on their absolute values, which may not indicate their actual relationship. After standardisation or normalisation, the linkage of variables better reflects their relationship with respect to lithological changes.

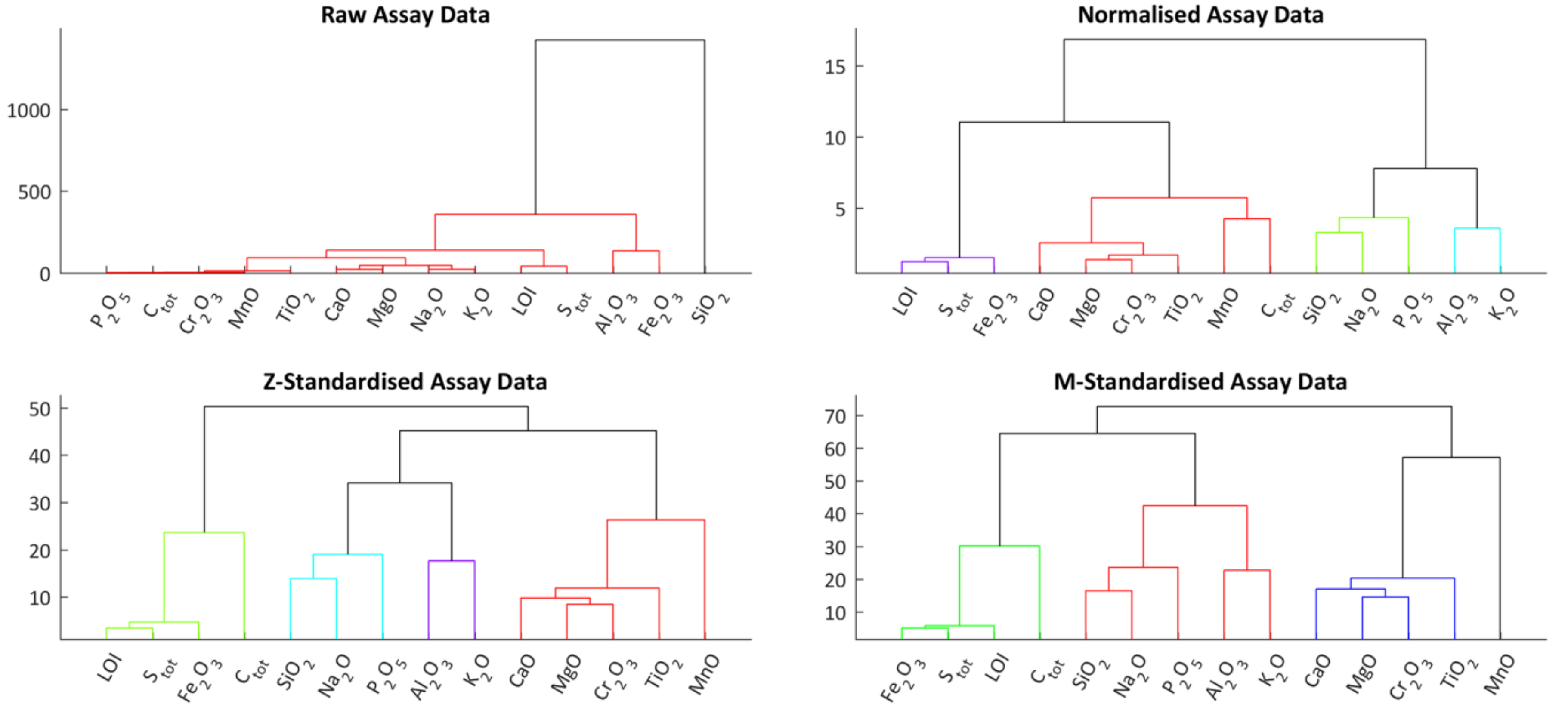


Figure 3.5: Category dendrograms for the different transformations of the elemental assay data. The dendrogram of the raw data (top left) orders the variables with respect to their absolute values (smallest for P_2O_5 and largest for SiO_2). The variable grouping is similar for all three data transformations but the height of the links (y-axis) is smallest for the normalisation because values are in the range between 0 and 1 and thus very similar. The main groups of variables (different colours) are indicative of the different major lithologies.

due to underwhelming results in all tests (Figure 3.6 A, B).

Fuzzy c-means (Bezdek et al., 1984) clustering is an example of soft clustering where samples can be part of more than one cluster simultaneously. The initial cluster centre calculations are based on the mean of the variables. The algorithm groups the data into clusters based on the distances of the samples to the centroids by minimising the following least squares objective function:

$$J_{fcm}(U,V) = \sum_{j=1}^{n}\sum_{k=1}^{c} u_{jk}^{m} \|z_j - v_k\|_2^2 \ ,$$

$$\text{subject to} \sum_{k=1}^{c} u_{jk} = 1 \ ,$$

> where n is the total number of sample points $z = z_1, z_2, \ldots, z_n$, c is the number of clusters, m is the weighting exponent ($m \geq 1$), and $V = v_1, v_2, \ldots, v_c$ are the centre values. $U = u_{jk} \in [0,1]$ is the membership matrix whose elements u_{jk} represent the membership degree of the jth data point to the kth cluster. $\| \ \|_2$ is the Euclidean norm.

The number of clusters c and the weighting exponent m need to be defined prior to clustering. A cut-off value α (Yang et al., 2008) for the membership degree is defined, such that a sample with a membership value above α is assigned to one unique cluster and membership below the cut-off results in the sample belonging to more than one cluster at the same time (non-uniquely classified). Higher values for m and α set fuzzier cluster boundaries and low values specify crisper boundaries. The effect of changing these parameters is shown in the Section 3.4. The FCM MATLAB code used for clustering throughout this study was modified to choose the initial centre values based on the data distribution, which makes the algorithm more stable—the solutions are identical for subsequent runs, which is not always the case when the initial centre values are randomly chosen (algorithm written and made available to me by my colleague Duy Thong Kieu, shown in Appendix I).

The **k-medoids** clustering method is similar to fuzzy c-means (or k-means), but instead of minimising the sum of the weighted squared Euclidean distances, the sum of the pairwise dissimilarities (distances) between a sample point and its centre is minimised. The centre of a cluster is not based on the mean as it is for k-means clustering, but on the median of the variables. This centre (in multidimensional space) is called the medoid and represents the most central point of the respective cluster. The algorithm partitions the data around the medoids (partitioning around medoids (PAM), Kaufman and Rousseeuw (2009)). The objective function tries to minimise the absolute distances (dissimilarities) to the medoid and is therefore more robust to noise and outliers.

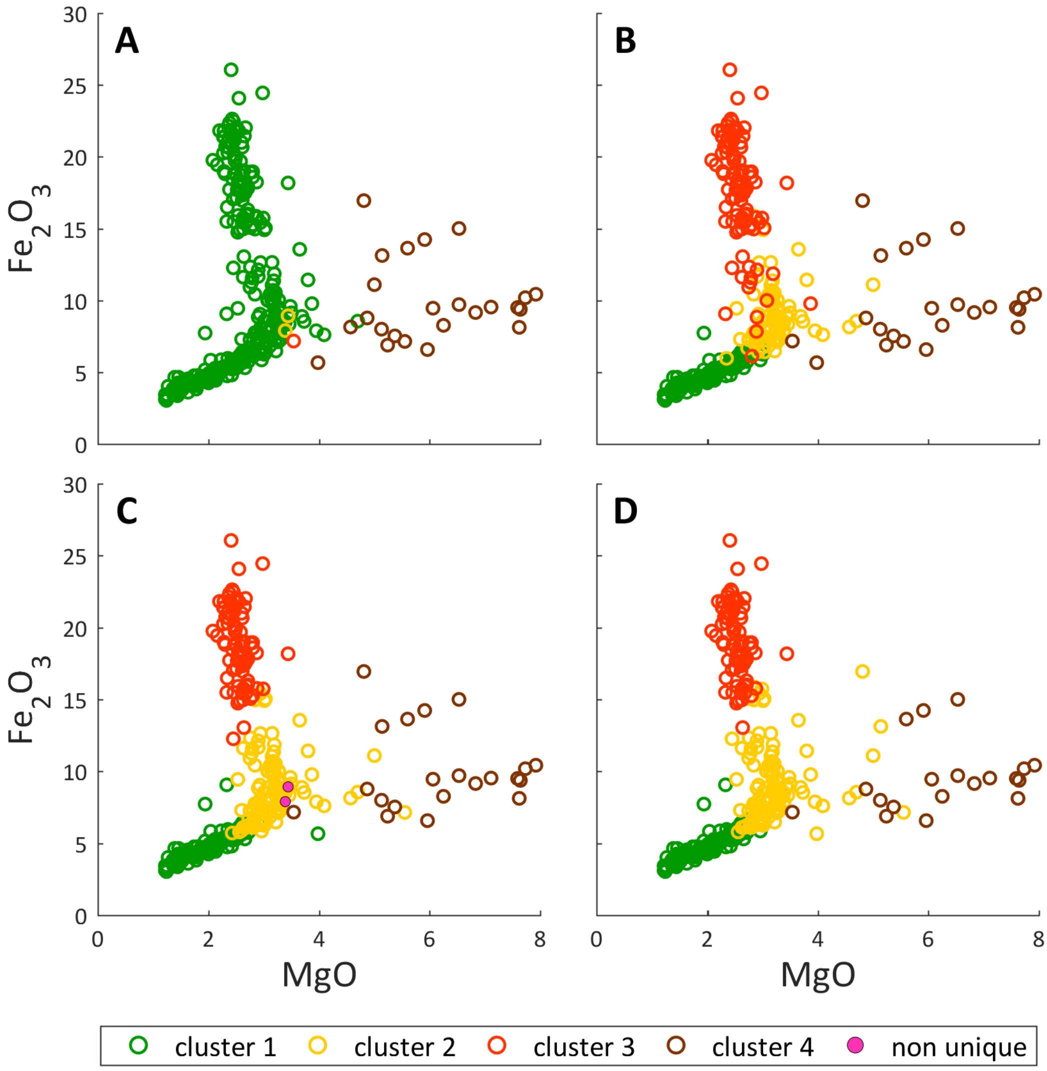


Figure 3.6: Results for clustering the z-standardised assay data plotted on MgO versus Fe_2O_3 scatter plots. A) Agglomerative hierarchical clustering (AHC) method with group average linkage, B) AHC method with Ward's linkage, C) fuzzy c-means (FCM) clustering and D) k-medoids clustering. Hierarchical clustering with the group average linkage did not find a satisfactory solution (most samples assigned to one cluster), whilst the other three methods found similar solutions. The FCM method (C) shows some non-uniquely classified samples where the membership degree is below the cut-off value.

All three methods were tested on the complete set of assay data (14 variables), which was first z-standardised. Figure 3.6 shows the results plotted on MgO versus Fe_2O_3 scatter plots. The results from hierarchical clustering using the group average linkage method (Figure 3.6 A) did not find a satisfactory solution and AHC using this distance calculation was not tested again. The three other methods yield similar and satisfactory results (Figure 3.6 B, C, D) and are valid choices for litho-geochemical classification. The measure of what is a 'good' solution and the difficulties of defining this will be discussed in later sections and chapters.

3.4 Optimising fuzzy c-means clustering parameters

The FCM clustering method allows the user to specify some input parameters that affect the cluster solution mainly in regard to the ‘fuzziness’ of the cluster boundaries. The weighting exponent m can take values between 1 (hard boundaries) and 30 (very soft, blurred boundaries), with values around 2 being the most widely used setting (Bezdek et al., 1984). The cut-off value α defines above what membership value ($\alpha < 1$) a sample is uniquely assigned to one particular cluster.

The main feature of FCM clustering is that a sample can belong to more than one cluster, defined by its membership value. Depending on the application, it may be preferred to highlight samples that represent mixtures or possibly altered states of the main lithology (setting soft boundaries) or, on the other hand, it may be more desirable to suppress these small-scale variations by setting harder boundaries. This would lead to a cleaner solution and a more ‘blocky’ classification of the rock mass where odd intervals are not separated but included in one of the major rock groups.

The effect of changing these parameters is demonstrated on the geochemical assay data (as before, 14 variables) and shown in Figure 3.7: (A) $m = 2$, $\alpha = 0.6$ (softer boundaries, high membership cut-off), (B) $m = 1.4$, $\alpha = 0.6$ (hard boundaries, high membership cut-off), (C) $m = 2$, $\alpha = 0.4$ (softer boundaries, low membership cut-off) and (D) $m = 1.4$, $\alpha = 0.4$ (hard boundaries, low membership cut-off). The rather fuzzy settings in (A) and (C) result in a high number of non-uniquely classified samples, whilst reducing m (B and D) lowers the number of non-uniquely classified samples and, setting a low cut-off value in addition leads to very few ‘unclassified’ samples. The solution in (D) seems preferable for grouping main lithologies or major changes in rock-mass features because most samples are assigned to a specific cluster and small-scale variations and mixed samples are not highlighted as non-uniquely classified samples. The settings for (D) were used in the classification procedures presented in the following chapters, unless otherwise specified.

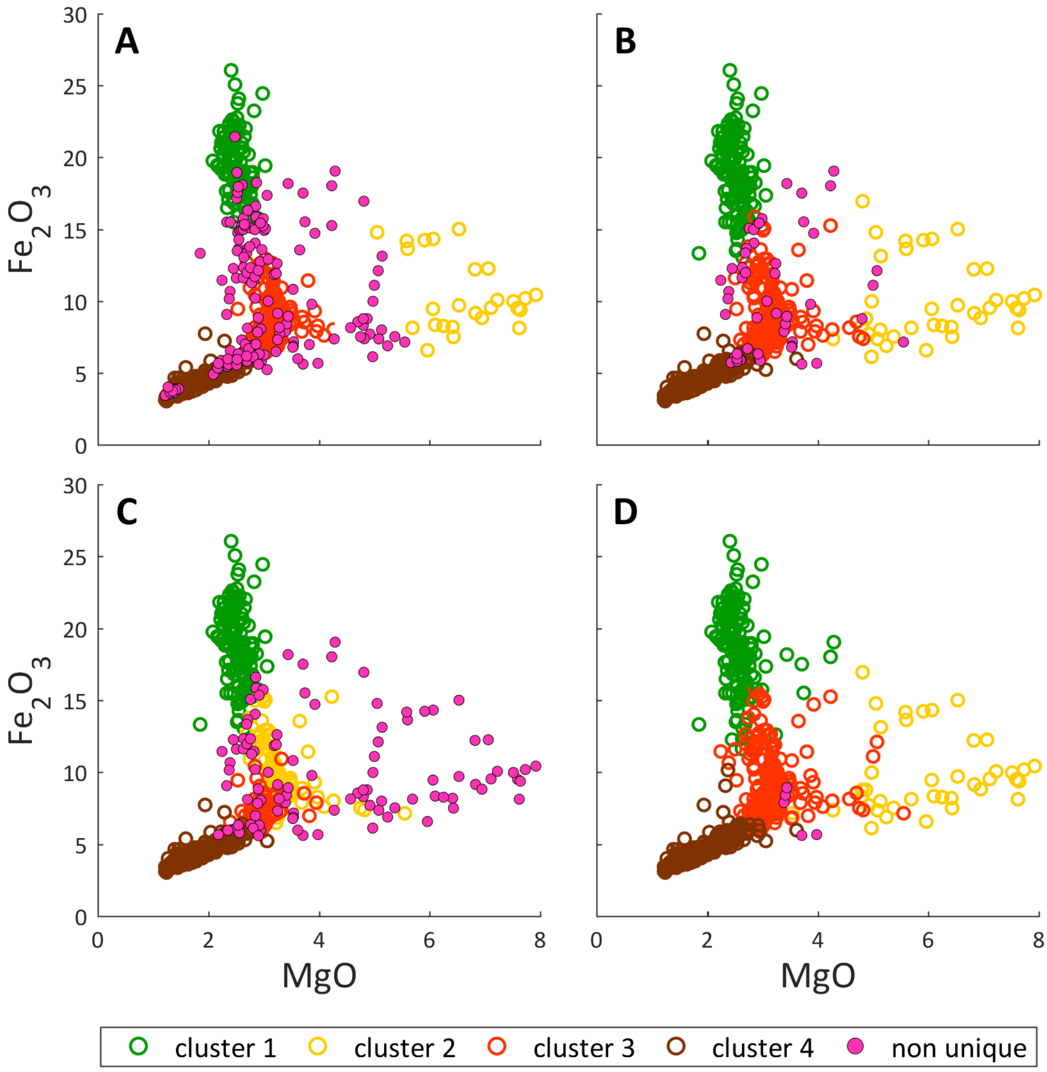


Figure 3.7: The effect of changing FCM clustering parameters is demonstrated on clustering the elemental assay data with different settings: (A) $m = 2$, $\alpha = 0.6$ (softer boundaries, high membership cut-off) shows a high number of non-uniquely classified samples, (B) $m = 1.4$, $\alpha = 0.6$ (hard boundaries, high membership cut-off) and (C) $m = 2$, $\alpha = 0.4$ (softer boundaries, low membership cut-off) show less non-uniquely classified samples and (D) $m = 1.4$, $\alpha = 0.4$ (hard boundaries, low membership cut-off) is the preferred solution for classifying main lithologies.

3.5 Reduction of high dimensional data sets

A large number of input variables can negatively affect the performance of clustering algorithms both in terms of the classification result and computational time because this excess dimensionality results in too many degrees of freedom (Steinbach et al., 2004). There are several possible methods that can be used to reduce the number of input variables (dimensions) without compromising the integrity of the data.

As demonstrated earlier, hierarchical clustering and dendrogram analysis can help identify closely related variables and can also be used to reduce the dimensionality of a data set. Variables that are correlated or similar are linked early during agglomerative clustering, indicated by shallow links in the dendrogram plot (e.g., Figure 3.4, Figure 3.5). Choosing just one variable from a predefined number of groups; for example, one element from each of the different coloured groups in the dendrograms in Figure 3.5, reduces the dimension from 14 to 4. For the present example, the SiO_2, Al_2O_3, Fe_2O_3 and CaO assay data were selected. Those four variables are in different groups for the normalised and z-standardised data but not for the m-standardised data (Figure 3.5). As seen from the previous statistical analysis (Section 3.1), the MnO data is highly skewed and represents an outlier in regard to the rest of the variables when no log-transform is performed prior to standardisation (the data in Figure 3.5 was not log-transformed). In the case of the m-standardised data, MnO would always be chosen as one of the four clusters (groups) since it is linked to the rest of the data at the very top of the tree. It is very unlikely though that manganese plays a major role in general lithological classification of this data and its inclusion in further steps would very likely lead to insignificant classifications. This shows how important the application of the 'right' pre-processing method can be. Here, choosing one variable from the four groups of the z-standardised data seems most reasonable to reduce data dimensions.

Another way of reducing data dimensionality is to perform principal components analysis (PCA). PCA projects the original data onto a new, orthogonal coordinate system where the axes represent the variance of the data in descending order. The individual components are constructed such that the first principal component accounts for the largest variance of the data, the second component accounts for the second largest variance of the data but orthogonal to the first and so on. In most instances, the first few principal components account for most of the variance of the original data, thus retaining important information of the data in fewer variables. The first four components, which account for 85.5% of the variance, were chosen for this example. PCA was performed on z-standardised data, which is the common procedure.

A third approach to reducing the number of variable dimensions, especially when analysing elemental assay data, is the expression as ratios. The classification of certain rock types is often based on ratios instead of the oxides. The following four ratios were chosen for clustering: TiO_2/Al_2O_3, Al_2O_3/SiO_2, $S/(S+Fe_2O_3+Al_2O_3)$ and $CaO/(Fe_2O_3+SiO_2)$, some of which were also used to define a baseline classification for comparison (see below). The three reduced subsets of data ((B) chosen from dendrogram, (C) chosen from PCA, (D) chosen as ratios) were first z-standardised and then clustered into four groups using the (1) agglomerative hierarchical clustering method, (2) fuzzy c-means and (3) k-medoids clustering. The results are compared to a pre-established baseline classification based on cut-off values of some elemental ratios ((A) predefined groups). These groups are defined by manually dividing the data into groups that represent the different lithologies (a more detailed reasoning for choosing these groups is given in Section 4.1.2):

Class 1 – $S/(S+Fe_2O_3+Al_2O_3)$ <0.07, $CaO/(Fe_2O_3+SiO_2)$< 0.07;

Class 2 – $S/(S+Fe_2O_3+Al_2O_3)$ >0.07 and < 0.21, $CaO/(Fe_2O_3+SiO_2)$< 0.07, Fe_2O_3/SiO_2 >0.11;

Class 3 – $S/(S+Fe_2O_3+Al_2O_3)$ >0.21, $CaO/(Fe_2O3+SiO_2)$ <0.07;

Class 4 – $CaO/(Fe_2O_3+SiO_2)$ >0.07.

Table 3.6 and Figures 3.8 and 3.9 summarise the clustering results. In Figures 3.8 and 3.9, the four clusters for each data set (A = predefined groups, B = dendrogram-based, C = ratio-based, D = PCA-based) and method are plotted on MgO versus Fe_2O_3 scatter plots. The plots show similar results independent of the cluster and reduction methods. Table 3.6 presents the results in terms of two numerical measures of success: (1) as the percentage of intervals matching those of the baseline classification (number of matching intervals divided by total number of intervals multiplied by 100) and (2) in terms of percentage of class changes along the data dimension (down the borehole). This second measure gives an indication of the level of noise of the classification where less changes yield a ‘cleaner’, less noisy result.

Table 3.6: Results of hierarchical- (AHC), fuzzy c-means (FCM) and k-means (KM) clustering for the three reduced data subsets: (B) dendrogram-based, (C) ratio-based and (D) PCA-based. (A) is from the predefined classes representing the baseline for comparison.

Cluster-method:	Pre-defined	AHC			FCM			KM		
Data	A	B	C	D	B	C	D	B	C	D
Class matching %	N/A	83	79	65	81	84	78	82	86	80
Class changes %	15	24	19	23	25	15	24	24	14	26

The FCM and k-medoids methods performed best with the ratio-based data set (C) with 84% and 86% matching intervals and 15% and 14% interval changes compared to the 15% changes of the assay-based classification. Hierarchical clustering worked best on the data set chosen by hierarchical grouping of variables, but the ratio-based set performed well also. The PCA-based subset is the least well classified for all three methods. It is possible that the chosen first four components that account for ~85% of variance in the data do not contain enough information for the data to be clustered satisfactory. Principal component analysis also obscures the original data because it is not clear which variables are part of which component and to what degree. Reducing data dimension via PCA was therefore not applied again in this study. In cases where it is difficult to decide on a sensible choice of ratios, for example, because the lithological or litho-geochemical characteristics are not well known, dendrogram analysis seems best suited for reduction of the data dimension. Further examples of using dendrogram analysis are shown throughout this study.

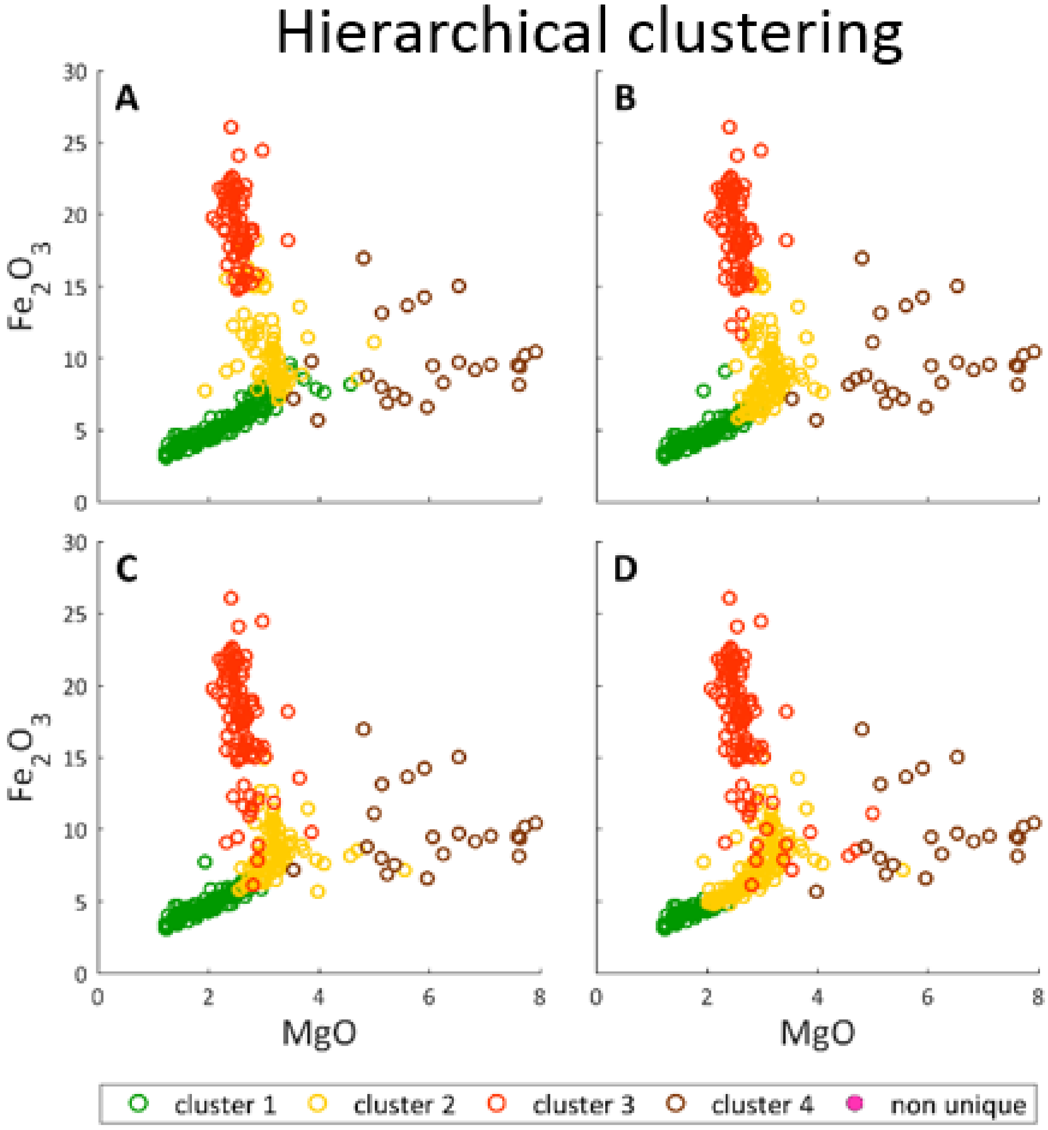


Figure 3.8: Results of Hierarchical clustering of the different reduced data sets, plotted on MgO versus Fe_2O_3 scatter plots. Data set: (A) predefined groups, (B) selected from dendrogram groups, (C) ratio-reduced, (D) PCA-reduced. The plots show overall similar results, but some differences are apparent on closer examination. Compare with Figure 3.9 and Table 3.6.

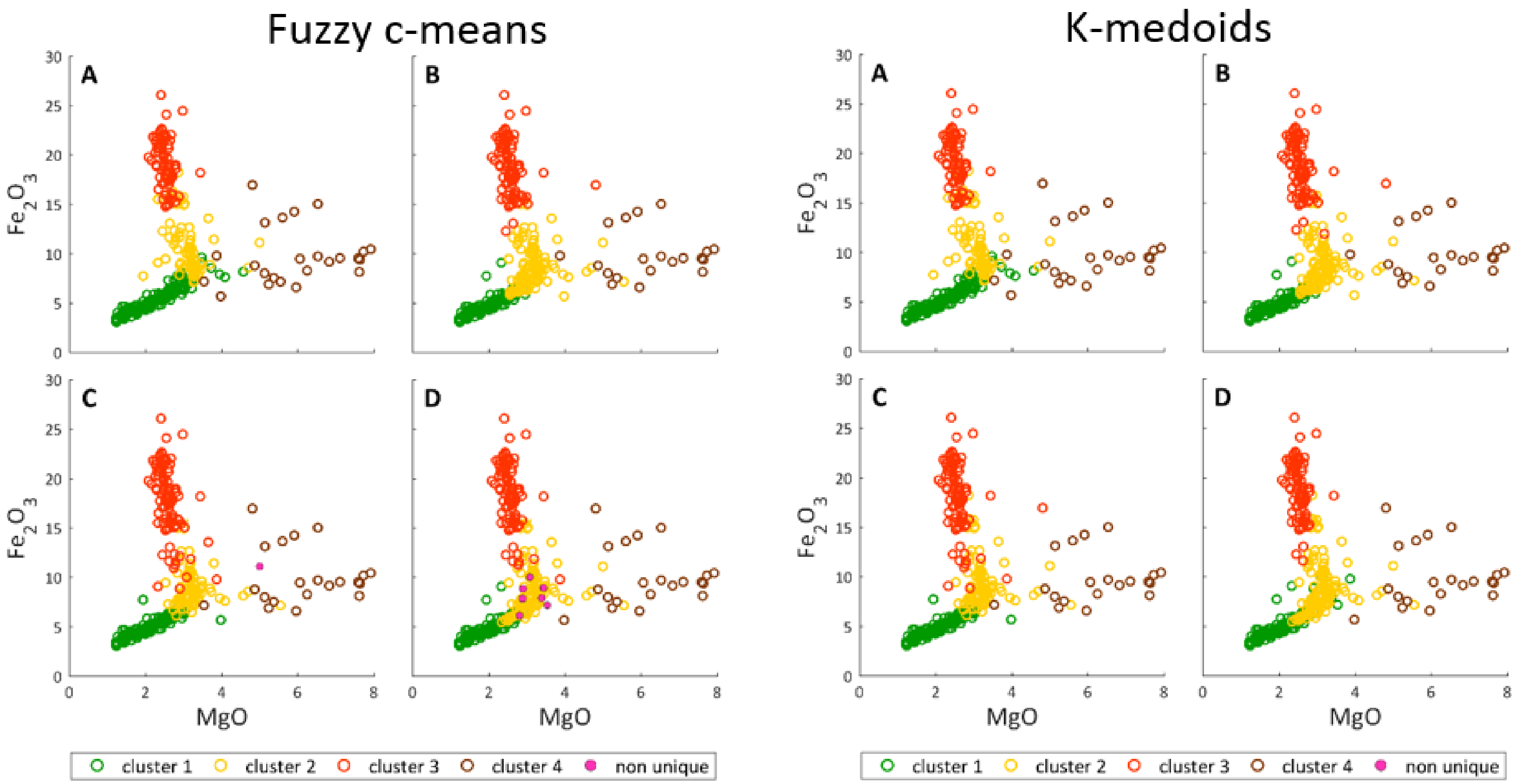


Figure 3.9: Results of FCM (left) and KM (right) clustering of the different reduced data sets plotted on MgO versus Fe_2O_3 scatter plots. Data set: (A) predefined groups, (B) selected from dendrogram groups, (C) ratio-reduced, (D) PCA-reduced. The plots show overall similar results but some differences are apparent on closer examination. Compare with Figure 3.8 and Table 3.6.

Summary

The analytical procedures shown in the previous sections outline important and necessary steps with regards to geoscientific data analysis. Investigating the structure of the data variables is important to decide which data transformations may need to be applied. Log transformations may be necessary to obtain normally distributed data and standardisation is essential if the data ranges differ considerably. Which standardisation method is applied depends on the data and the purpose of the subsequent classification. Median-based standardisation is robust against outliers but may not be the preferred choice when the information contained in outliers is important to distinguish certain lithologies, such as thin mineralised sections within barren rocks.

Dendrogram analysis of clustered variables provides important information about their relationships, that might not be obvious from calculating correlation coefficients. Visual representation of data statistics, in the form of box plots, histogram plots or dendrograms, generally provide more intuitive means of studying their structure and correlation. Reducing the amount of input variables is important for optimal cluster performance and dendrograms can again be useful in determining which variables may be included or omitted.

Furthermore, the choice of clustering method should take the purpose of the classification and the structure of the data into account. For geoscientific data, the fuzzy c-means method is best suited because of the manner in which it deals with uncertainty. Instead of being forced into a specific class (cluster), fuzzy c-means allows samples to belong to more than one cluster and these samples can be easily highlighted and identified as minor or undetermined lithology.

Fuzzy c-means clustering of prior z-standardised data was applied throughout the subsequent technical part of this study. The importance of choosing the ‘right’ input data for clustering is investigated in the following chapter.

Chapter 4

DATA CLUSTERING

One of the objectives of this research is to identify suitable combinations of measurements to classify rocks for a specific purpose. Another aim is to investigate how merging petrophysical data to more traditional elemental measurements can improve classification, or act as substitutes to enable real-time feedback. For example, the rock can be grouped according to lithological differences, economical value or textural features. Not every measurement or analysis is necessary for a particular distinction; on the contrary, some variables when included in the clustering process might even lead to misclassification if the information contained in those variables is not representative of the desired distinction. So, it is important to understand the underlying relationship between a specific measurement and the rock's properties to choose the right combination (or at least to start with a sensible combination) of input variables for a particular intended purpose. For example, when classifying main lithologies, a few major elements from the chemical analysis may be sufficient since many rock types are classified by their chemical composition (e.g., silica content decreases from felsic to intermediate to mafic to ultramafic rocks). Modal mineralogy, of course, is often more important for lithological classification and characterisation of alteration style but mineralogical analyses are more difficult and expensive to obtain and are rarely acquired in a minerals exploration context. However, petrophysical measurements can be used as indicators for certain minerals or differences in mineralogical compositions and two examples are shown in this chapter.

Another classification objective is the task of distinguishing between economic grade and waste, which can be achieved by choosing assay data of the element or elements of value (e.g., Fe for iron deposits, Fe and S for sulphide deposits) or the measurement of another rock property that is indicative of the valuable formation (e.g., density or conductivity for many base metal deposits). Textural features on the other hand, are unlikely to be detected reliably by chemical analysis, but petrophysical measurements may be invaluable here. An example is shown in Section 4.1.4.

A principal aim of the following data clustering procedures is to determine what data variables are important for certain aspects of rock mass classification and the best combination of petrophysical data to substitute for elemental assay data. The different cluster methods described previously are compared on the initial lithological classification of the first data set (Brukunga sulphide deposit); however, in the other case examples only fuzzy c-means clustering is used because it proved to be the most appropriate method to cluster geoscientific data (Section 4.1).

4.1 Testing cluster analysis on data from the historic Brukunga pyrite deposit

The three previously described cluster methods are tested on different combinations of input variables from the data set acquired at the historic Brukunga sulphide mine. The objectives are to study the relationship between different data variables (petrophysical and geochemical), identify those that contain meaningful information for rock mass classification, identify where variance in the data corresponds to lithological changes or changes in rock properties and ultimately find the minimum required, or best choice of, variables necessary to satisfactorily classify the rock mass. The data variables and their structure were described earlier in Chapter 3 (Methodology) and the available measurements summarised in Table 3.1.

First, different combinations of data are tested for general lithological classification. This procedure demonstrates how a comprehensive set of data can be utilised to learn about the relationships between the measurements and rock properties. The process also exemplifies how important it is to obtain as many measurements as possible from the first drill holes of an exploration target. The learnings from this initial analysis not only determines which of the measurements are most important to classify the rock mass, but can also be adapted to new targets of similar geology for future exploration campaigns. It should be noted however, that consistent calibration of downhole tools is critical for the successful classification of subsequent data of a given prospect. Some issues arising from using uncalibrated data are discussed throughout the next chapters.

4.1.1 Geology and mineralisation of the Brukunga Pyrite mine area

The most recent stratigraphic classification by Jago et al. (2003) places the pyrite-rich units at Brukunga within the Early Cambrian Talisker Formation of the Early to Middle Cambrian Kanmantoo Group (Figure 4.1). The Talisker Formation in the Brukunga area is composed of dark pyrite rich siltstones with sandstones and minor calc-silicate interbeds and a massive sandstone unit (Cooalinga Member, not present everywhere) at its base. It is bound by sandstones, greywackes and conglomerates of the Backstairs Passage Formation in the footwall, and by graded greywackes and laminated siltstones of the Tapanappa Formation in the hanging wall. The mine site is located at the western flank of a regional scale syncline, which is tectonically

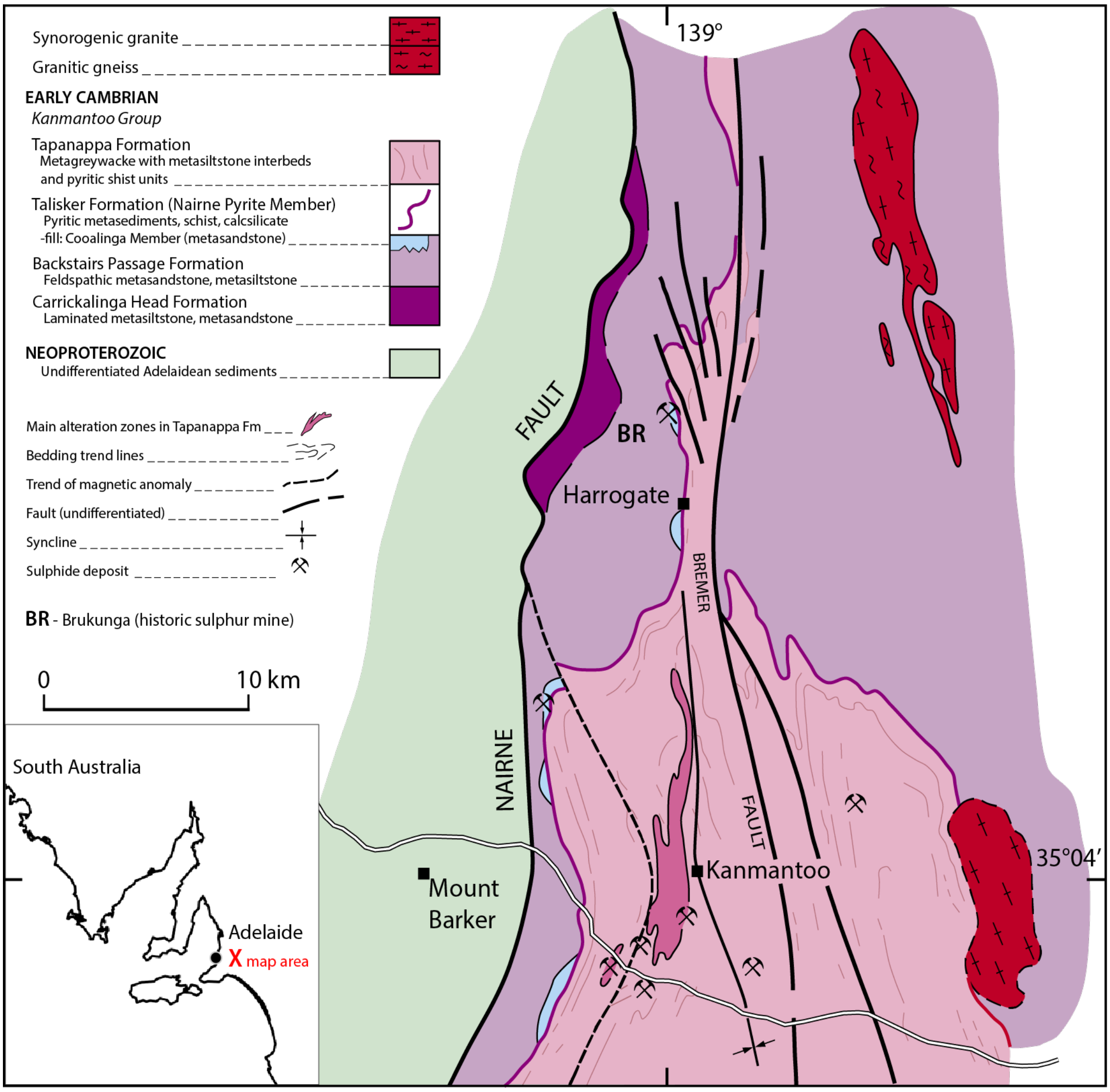


Figure 4.1: Geological overview of the Nairne - Mount Barker area showing the location of the historic Brukunga pyrite mine ((BR) (after Toteff (1999)).

complicated by north-south trending major and minor faults that developed during the Delamerian orogeny (Figure 4.1).

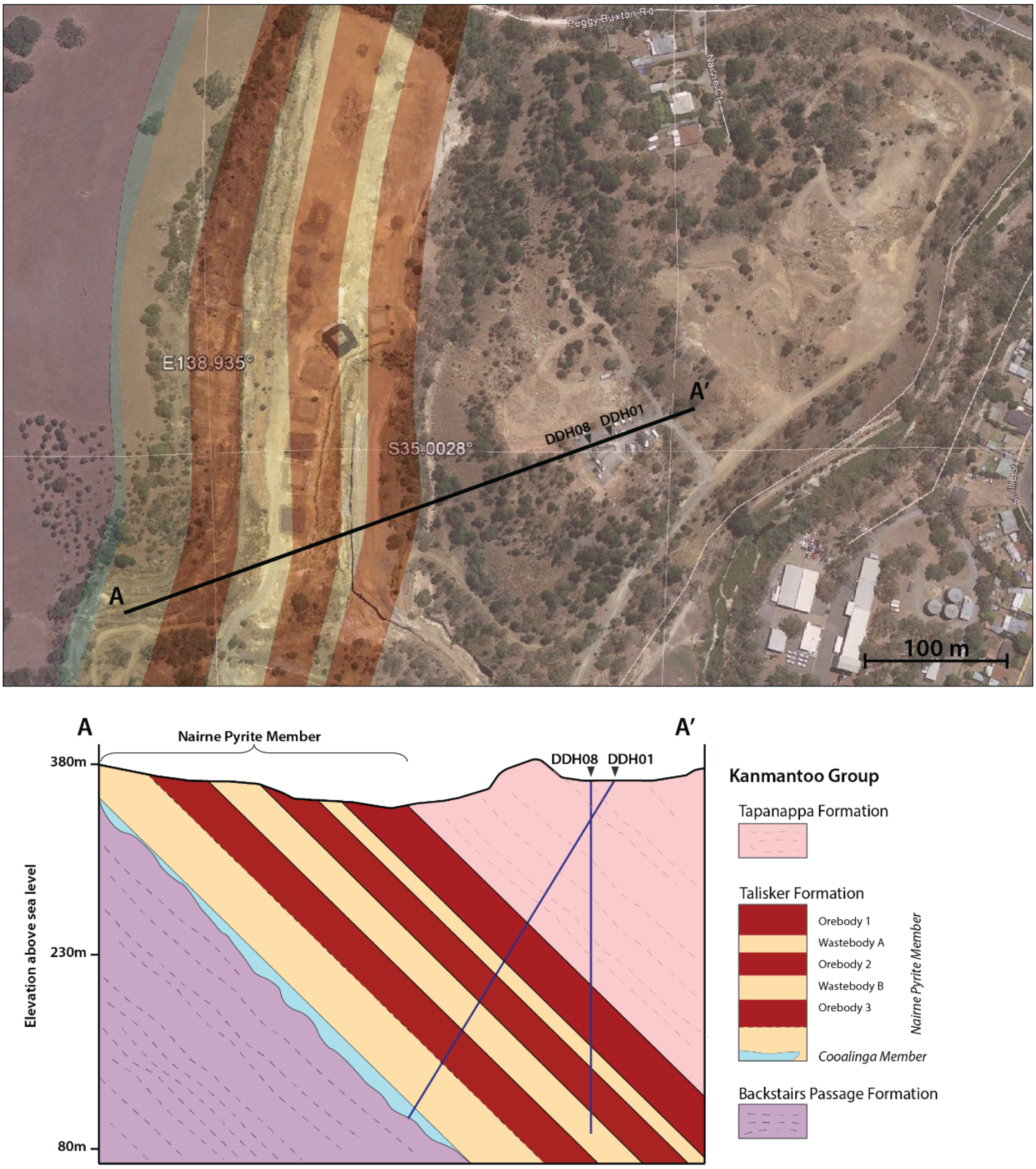


Figure 4.2: Geological overview and cross section of the Brukunga area. Data from drill hole DDH01 is used for cluster analysis in this chapter and data from DDH08 is used for predicting lithology and rock mass characteristics in Chapter 5. The geology is interpreted from drilling; ore-waste classification after George (1967).

Sulphide mineralisation is believed to be of primary sedimentary origin with some sulphide remobilisation and metamorphic segregation during amphibolite facies synorogenic metamorphism (LaGanza, 1959; Skinner, 1958). George (1969b) described these textures and associated mineralogy in great detail, differentiating between two fractions of sulphide-silicate rock and various vein types. He divided the pyrite-rich units into five sub-units based on their sulphide content; these are from oldest to youngest: Orebody 3, Wastebody B, Orebody 2, Wastebody A and Orebody 1 (Figure 4.2). The ore bodies contain about 12 volume percent pyrite and 6 volume percent pyrrhotite and the waste bodies contain about 10 volume percent pyrrhotite and only little pyrite. The silicate mineralogy also changes with respect to the grade of mineralisation. In pyrite and pyrrhotite poor units, quartz, plagioclase (60% albite and 40% anorthite component) and muscovite are accompanied by biotite whilst sulphide rich units contain iron deficient phlogopite instead of biotite and some K-feldspar. George (1969b) links this to the local sulphur activity within the different beds during metamorphism. Where primary sedimentary sulphur content was low, all iron was incorporated into biotite during metamorphism; however, with increasing sulphur activity, pyrrhotite was formed first and at the highest sulphur activities pyrite developed in addition to pyrrhotite. Both sulphide minerals were remobilised during deformation; pyrite is often present in bedding plane parallel segregations and axial plane veins, while pyrrhotite is present in up to 3cm thick 'tension gash' veins or disseminated throughout the matrix (George, 1969a). These 'tension gashes' cross-cut the older pyrite segregations and veins and adjacent areas are depleted in pyrite (George, 1969a).

The overlying Tapanappa formation is composed of massive to laminated, coarse to fine grained, sharp-based greywacke that grades upward into finer and darker laminated siltstones with minor sulphide and calc-silicate interbeds (Jago et al., 2003). None of these authors mention occurrences of dolerite in the area but Gum (1998) describes two samples of dolerite in the historic drill hole DDH14 at Brukunga. In addition, Lawrence (1980) found dolerite in the Talisker Formation in the Eastern Mt. Lofty Ranges (Nairne Pyrite equivalent) and Lyons (2012) mentions similar occurrences at the Kanmantoo Copper deposit about 10km south-east of Brukunga.

4.1.2 Lithological classification using different combinations of input variables

The general geological description of the previous section provides some idea about how the rock mass at Brukunga may be classified. Barren rocks may be separated from mineralised parts and the two main types of sulphide mineralisation (high-grade pyrite and pyrrhotite, low-grade pyrrhotite) should be identified. It may also be possible to separate the dolerites from the main rock mass. In addition, quartz-rich meta-sandstones may be separated from more mica-rich meta-siltstones and meta-mudstones.

The different units have certain characteristics that are reflected in some of the petrophysical measurements and chemical analysis but not others. It is of great benefit to establish these relationships beforehand to increase the chances of success of a classification by choosing the right input variables.

The general structure of this data set is examined and described in Chapter 3. Some variables (e.g., SiO_2, Fe_2O_3, total sulphur, $_{\log}$resistivity, Vp) show a bimodal distribution pointing to the presence of at least two different main units which are readily separable. Some other variables are highly skewed suggesting the presence of a minor unit with quite different properties. The aim of the following classification procedure is to reproduce the classes defined from assay data:

Class 1 – $S/(S+Fe_2O_3+Al_2O_3) < 0.07$, $CaO/(Fe_2O_3+SiO_2) < 0.07$;

Class 2 – $S/(S+Fe_2O_3+Al_2O_3) > 0.07$ and < 0.21, $CaO/(Fe_2O_3+SiO_2) < 0.07$, $Fe_2O_3/SiO_2 > 0.11$;

Class 3 – $S/(S+Fe_2O_3+Al_2O_3) > 0.21$, $CaO/(Fe_2O_3+SiO_2) < 0.07$;

Class 4 – $CaO/(Fe_2O_3+SiO_2) > 0.07$.

Class 1 represents barren rocks from mainly the Tapanappa formation, class 2 are low-grade sulphide bearing rocks (mainly pyrrhotite), class 3 is the group of economic value (high-grade pyrite rich units) and class 4 represents dolerite whose chemical and petrophysical properties are markedly different from the other lithologies. To demonstrate how these classes can be reproduced by means of fuzzy c-means clustering, the above-defined ratios as well as the individual elements/oxides are used as input variables and the resulting clusters compared. The applied clustering parameters are: weighting value $m = 1.6$, cut-off value $\alpha = 0.4$ and a cluster number of 4.

Figure 4.3 shows the results on different scatter plots where the clusters (1 to 4) represent the individual classes (1 to 4): (A) are the predefined classes as per the above definition, (B) shows the results of fuzzy c-means clustering of the ratios ($S/(S+Fe_2O_3+Al_2O_3)$, $CaO/(Fe_2O_3+SiO_2)$, Fe_2O_3/SiO_2) and (C) shows the results of clustering variables (elements/oxides) that were used to define these ratios (S, Fe_2O_3, Al_2O_3, CaO, SiO_2). The results for clustering the ratios match the predefined classes with 91% accuracy and the results based on clustering the respective elements/oxides show an 86% match. As a consequence of the hard boundary settings and low cut-off value, most intervals are uniquely classified. Since FCM clustering divides the data into natural groups by a measure of similarity, these clusters rarely exactly match a classification based on arbitrary cut-off values. In Figure 4.3 for example, cluster 2 from FCM clustering (B and C) includes some samples of class 1 and 4 of the predefined classes (A). Similarly, cluster 3 contains some samples that have S/(S+Fe_2O_3+Al_2O_3) ratios below the 0.21 cut-off that divides classes 2 and 3. In the next step, the three previously described cluster methods are tested on different combinations of input data chosen from the pool of assay and petrophysical data. Since geochemical assay data is of considerably lower resolution (1m intervals) than petrophysical measurements (usually between 10 and 25cm intervals) and to increase the number of samples, the assay data was interpolated and resampled to 0.5m intervals, as were the petrophysical logs to match these intervals. This example illustrates the problems of combining usually high resolution petrophysical data with low resolution geochemical data. To analyse these data streams together a trade-off has to be made where spacial petrophysical information is lost in order to accommodate the assay data.

For cluster analysis, all major element assay data and most petrophysical logs are chosen separately first and then combined into one large set of data. Subsequently, combinations of three variables from the pool of petrophysical measurements are chosen, to ascertain which combination can substitute for assay data and still yield the desired classifications. The respective data combinations are z-standardised before clustering (Section 3.2). The results are reported in terms of intervals matching the predefined classes and in terms of class changes down the drill hole and summarised in Table 4.1. The percentage of class matches is calculated by dividing the number of matching intervals by the total number of intervals, multiplied by 100. The percentage of class changes is the total count of intervals that are not identical to the preceding interval, divided by the total number of intervals multiplied by 100. Less class changes are favourable because small-scale variations are ignored and the resulting classification represents the major lithological differences in an easy to interpret manner.

All three cluster methods give overall good to very good results, with some discrepancies depending on the individual input data sets. Clustering of the assay data (14 variables) yields very good results while clustering of petrophysical data (9 variables) does not reproduce the predefined classes equally well, although classification success is still above 70%. However, combining both petrophysical and geochemical data into a larger data set, yields slightly better results than clustering the separate data sets.

As mentioned in Section 3.5 (reduction of high-dimensional data sets), hierarchical clustering of the input variables and their representation as dendrograms can give information about their interrelationship. Figure 4.4 shows a dendrogram for the combined petrophysical and assay data including the ratios used for predefined

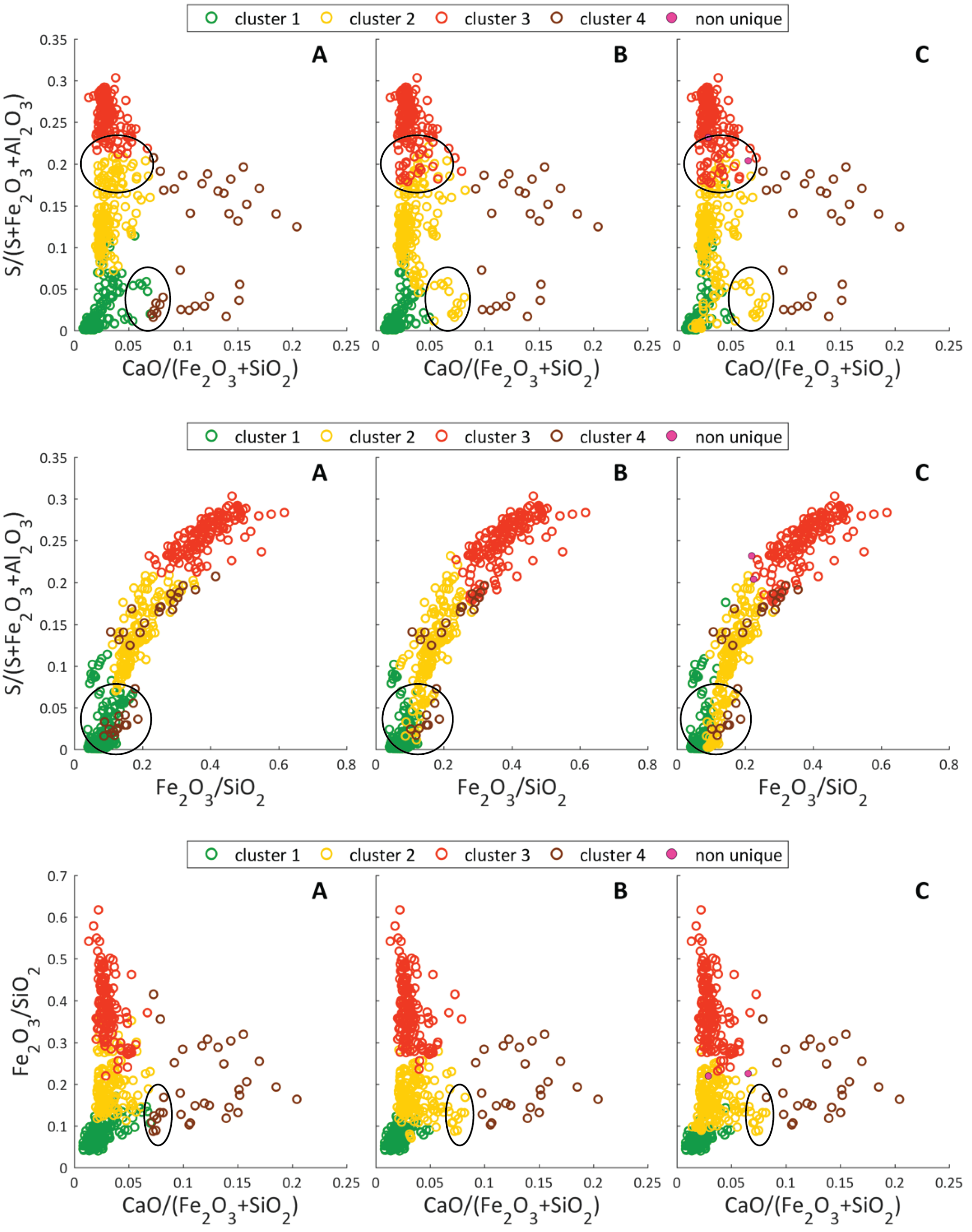


Figure 4.3: Comparison of the predefined classes in (A) and results of fuzzy c-means clustering of the three ratios that define these classes (S/(S+Fe_2O_3+Al_2O_3), CaO/(Fe_2O_3+SiO_2), Fe_2O_3/SiO_2) in (B) and fuzzy c-means clustering of the five elements/oxides involved (S, Al_2O_3, CaO, Fe_2O_3, SiO_2) in (C). The axes represent the three different ratios of the predefined classes (see text). Although the results of clustering (B and C) appear to be similar to the predefined classes (A), there are some apparent differences. For example, some samples of class 1 and 4 are grouped with cluster 2 and some samples of class 2 are grouped with cluster 3 by FCM clustering (black ellipses).

classification. The tree is colour-coded to show four clusters or groups that include similar or correlated variables. The four groups are defined at a particular hierarchical level, where a horizontal line (dashed line in Figure 4.4) intersects exactly four of the vertical connecting lines. Based on this tree, different input subsets from the pool of petrophysical variables are chosen, such that each of the three variables are selected from one of the different main groups. For example, density correlates with iron and sulphur since these elements are the main constituents of pyrrhotite and pyrite, which have a much higher density than the remaining rock-forming minerals at Brukunga. The dendrogram also shows that the magnetic susceptibility is closely linked to apparent chargeability and then aluminium, suggesting that pyrrhotite may be more abundant in the laminated shale-richer units. The latter variables are then linked with the density-iron group which confirms the presence of a pyrite-pyrrhotite system. Resistivity on the other hand correlates with the amount of silicates present, especially the quartz content of the rock mass, or inversely with the percentage of interconnected sulphides. The natural gamma log is indicative of the dolerite dykes such that the gamma reading is lower where dolerite is present due to the absence of potassium in this rock type. The inverse of the natural gamma log is included in the dendrogram to highlight its relationship to those elements that are more abundant in dolerite, namely calcium, titanium, magnesium and chromium.

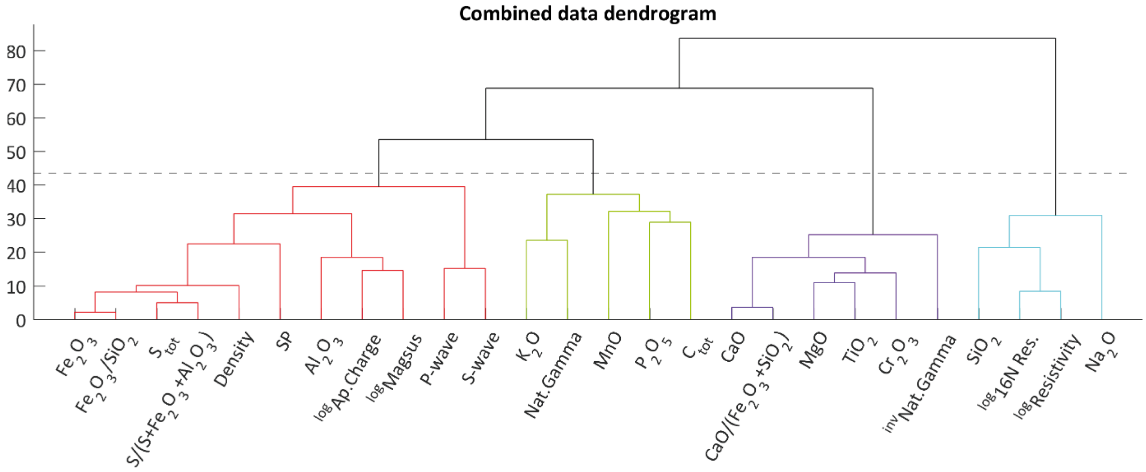


Figure 4.4: The category dendrogram can be used to determine which variables can be substituted for each other. If petrophysical measurements are substituted for assay variables, the measurement with the closest link to the respective assay variable is likely the best choice. For example, resistivity may be substituted for silica; natural gamma for potassium or density for iron or sulphur. The dashed line indicates the hierarchical level that defines four clusters.

Based on the dendrogram analysis the following subsets are chosen for clustering: (a) natural gamma, 16-inch resistivity, density (NG/16N/De); (b) natural gamma, 16-inch resistivity, spontaneous potential (NG/16N/SP); (c) natural gamma, 16-inch resistivity, apparent chargeability (NG/16N/AC); (d) natural gamma, 16-inch resistivity, magnetic susceptibility (NG/16N/MS). The results summarised in Table 4.1 indicate that all subsets are valid choices for a successful lithological classification at the Brukunga site, with the combination of density, natural gamma and 16-inch resistivity performing best. These results illustrate how prior knowledge about the general geology when paired with analytical tools such as dendrograms, provide the necessary information to choose the right combination of input data for successful lithological rock mass classification. The outcomes also indicate that the choice of cluster method is subordinate when the ‘right’ data combination is selected. In fact, the results from the different clustering algorithms are similar to each other for each of the respective data combinations. Clustering a sensibly-chosen subset—the one that makes most geological sense—of only three variables gives superior results compared to using all available data. The inferior results from clustering all available data may be in part a result of the previously mentioned excess dimensionality, but also because some of the variables do not add any useful information or their inclusion may even distract from the desired solution. Applied to exploration campaigns, this directly translates to improved cost and time efficiency by knowing exactly which logging data to collect from subsequent drill holes.

Table 4.1: Results of clustering the different subsets of data using the three different cluster methods show that all methods and subsets are viable choices for lithological classification. The data set including all petrophysical variables (Pdat) does not recover the predefined groups as well as the data set including all assay data. Fuzzy c-means clustering of natural gamma, 16-inch resistivity and density data (NG/16N/De (a)) was found to perform best.

Cluster Method ⇓	*Dataset* ⟹	Cdat	Pdat	CPdat	NG/ 16N/ De/ (a)	NG/ 16N/ SP/ (b)	NG/ 16N/ AC/ (c)	NG/ 16N/ MS/ (d)	NG/ SP/ AC/ (e)
AHC	Class matching %	83	73	84	87	85	85	87	85
	Class changes %	7	5	7	11	7	14	14	7
FCM	Class matching %	84	77	87	92	85	87	43	86
	Class changes %	12	7	8	11	12	15	13	10
KM	Class matching %	87	71	89	91	87	87	87	86
	Class changes %	10	9	7	11	10	14	14	9

Cdat - SiO_2,Al_2O_3, Fe_2O_3, CaO, MgO, Na_2O, K_2O, MnO, TiO_2, P_2O_5, Cr_2O_3, LOI, total C and total S in wt%; **Pdat** - density (De), s-wave (Vs) and p-wave (Vp) velocities, natural gamma (NG), spontaneous potential (SP), resistivity (Re), $_{\log}$16-inch resistivity (16N), $_{\log}$apparent chargeability (AC), $_{\log}$magnetic susceptibility (MS); **CPdat** - Cdat and Pdat combined.

4.1.3 Petrophysical measurements for identifying sulphide minerals

Similar to grouping major lithology, petrophysical data may help to differentiate between different mineral constituents or textural rock features. The present example aims to separate (1) pyrite-pyrrhotite rich units at the Brukunga deposit from (2) units that contain mainly pyrrhotite, (3) generally poorly mineralised units and (4) dolerite intervals if possible. This can be achieved because the physical properties of pyrite and pyrrhotite are markedly different (Table 4.2) and their different response to petrophysical measurements can be used to estimate their relative abundance in a rock. Unfortunately, no professional mineralogical analysis (XRD etc.) was undertaken on samples from drill hole DDH1 and thus defining the exact relationships between mineralogy and petrophysics is not possible for this example. However, in order to correlate petrophysics and mineralogy and to compare the cluster results, a rough estimation of the relative abundances of the major minerals was calculated from the assay data by using a constrained non-linear optimisation function (MATLAB ‘fmincon’ function, Appendix I). The function can be used to estimate the relative abundances of a predefined number of minerals based on their chemical composition. Pyrite and pyrrhotite are calculated from the weight percent of elemental sulphur and iron assay data (only for the mineralised part of the hole from 134m onwards) and the estimates are then compared to the visual logging performed on the drill core by a geologist (Figure 4.5). The relative abundance of logged and calculated pyrite matches each other well, but the visual logging seems to have underestimated the absolute amount in some areas. The same is true for pyrrhotite, which is in parts greatly underestimated by the logging geologist. This is likely because pyrrhotite is mainly disseminated throughout the matrix and obvious patches only occur in places of structurally induced remobilisation. The magnetic susceptibility log (log-transformed, $_{\log}$MS) on the other hand shows a strong correlation to the calculated pyrrhotite abundance and may therefore be used to distinguish these

Table 4.2: The physical properties of pyrite and pyrrhotite indicate which petrophysical measurements are useful to distinguish lithologies containing different amounts of these minerals.

Mineral	Density [g/cc]	Hardness [Moh’s]	Vp [km/s]	Vs [km/s]	Magnetic property
Pyrite - FeS_2	5.0	6.0–6.5	7.8	4.85	paramagnetic
Pyrrhotite - $Fe_{1-x}S$ ($x = 0$ to 0.17)	4.6	3.5–4.5	4.6	2.7	ferrimagnetic, varying inversely with iron content

minerals through cluster analysis. Since the other physical properties of the two sulphides are quite different as well, the Vp (P-wave) or density logs may be used as a second input variable. The natural gamma log (Nat. Gamma) is included to separate the dolerite.

Two sets of input data are clustered: (1) Vp, magnetic susceptibility and natural gamma (Vp/MS/NG) and (2) density, magnetic susceptibility and natural gamma (De/MS/NG). Results of clustering these two data sets into four clusters (FCM parameters: $m = 1.6$, $\alpha = 0.4$) are shown on scatter plots in Figure 4.6 and Figure 4.7 respectively. The plots labelled (A) show the clusters on Vp versus magnetic susceptibility (B) on density versus natural gamma and (C) on pyrrhotite versus pyrite scatter plots. The samples are grouped slightly differently depending on whether Vp (Figure 4.6) or density (Figure 4.7) was clustered. Cluster 1 are poorly mineralised samples containing only low amounts of pyrrhotite, cluster 2 samples contain mainly pyrrhotite and are characterised by high magnetic susceptibility, cluster 3 represents the main mineralisation (pyrite-rich) and cluster 4 are dolerites. The main difference is that when clustering Vp instead of density, some pyrite free samples are grouped with samples of economic value, which is incorrect (red circles on x-axis in Figure 4.6 C). This indicates that other rock properties, such as textural

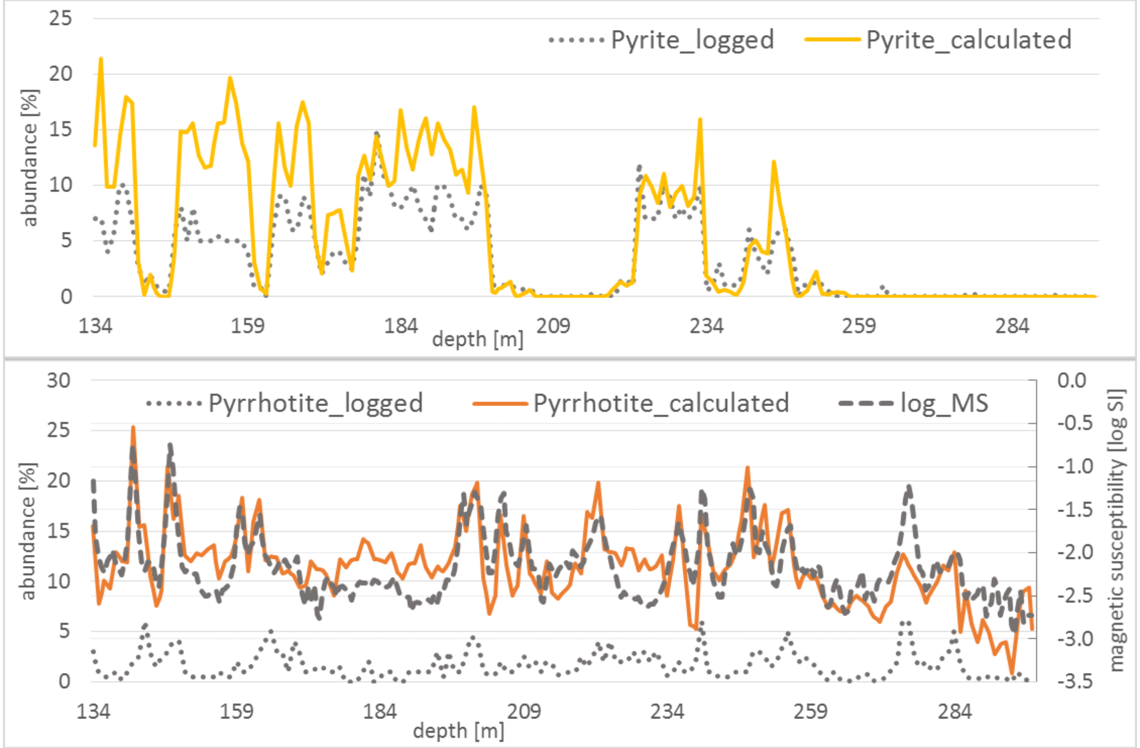


Figure 4.5: Top: comparison between logged (from visual inspection of drill core) and calculated pyrite shows a good match. Bottom: logged versus calculated pyrrhotite suggests pyrrhotite is underestimated from visual logging. Comparison of the calculated amount of pyrrhotite with magnetic susceptibility shows a very good match.

features (that cause anisotropy), may be responsible for the higher p-wave velocities and that pyrite content is not the major factor influencing sonic wave velocities through these rocks. Vp is therefore not the best candidate to distinguish between different mineralogical compositions in this example.

In contrast, the difference in density between pyrite and pyrrhotite is quite small; however, due to the overall increase in sulphide mineral content of the pyrite-rich units, clustering density can successfully differentiate between the desired units. It has to be noted however, that some misclassification occurred for both data sets where a number of the pyrite-rich samples were grouped with cluster 2, that should only include pyrrhotite-rich but pyrite-poor samples. Figure 4.8 illustrates the cluster results in the form of pseudo-lithology well-log plots compared to the input petrophysical logs and sulphide mineralogy. For the case of clustering density, magnetic susceptibility and natural gamma (De/MS/Ga), cluster 1 (green) coincides with pyrite-free units, cluster 2 (yellow) with peaks in the magnetic susceptibility log, cluster 3 (brown) coincides with high pyrite abundance and cluster 4 is distinctive of low natural gamma areas and thus dolerite.

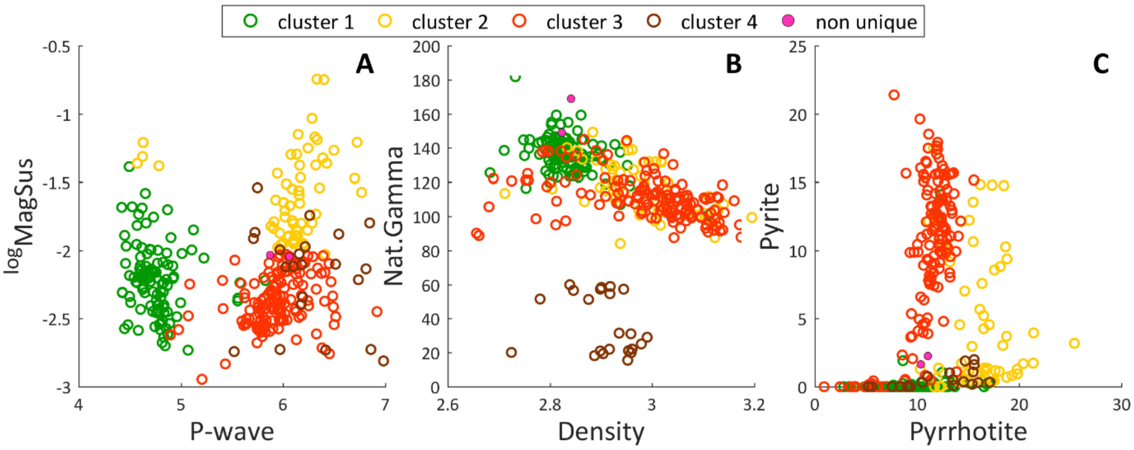


Figure 4.6: Scatter plots of the cluster results for Vp, magnetic susceptibility and natural gamma; cluster 1 are poorly mineralised samples containing only low amounts of pyrrhotite, cluster 2 samples contain mainly pyrrhotite (some pyrite-rich samples are misclassified into cluster 2), cluster 3 represents the main mineralisation (pyrite-rich, some pyrite-free sample are wrongly grouped here) and cluster 4 are dolerites. (A) illustrates how the data is clustered in relation to magnetic susceptibility ($_{\log}$MagSus) and Vp (p-wave). (B) shows the clusters on a natural gamma versus density plot for comparison with Figure 4.7 (B), where the clusters are better defined by use of density. (C) are the clusters in relation to pyrite and pyrrhotite content whose distinction was the aim of this cluster analysis. Some samples without mentionable amounts of pyrite (red circles on x-axis) are included in cluster 3, which should only contain intervals of economic value and some samples with significant pyrite content are incorrectly grouped with cluster 2. The plots indicate that Vp might not be the best choice for distinguishing pyrite-rich units despite the markedly different p-wave velocities for pyrite and pyrrhotite.

These results indicate that the p-wave velocity of the mineralised rocks in drill

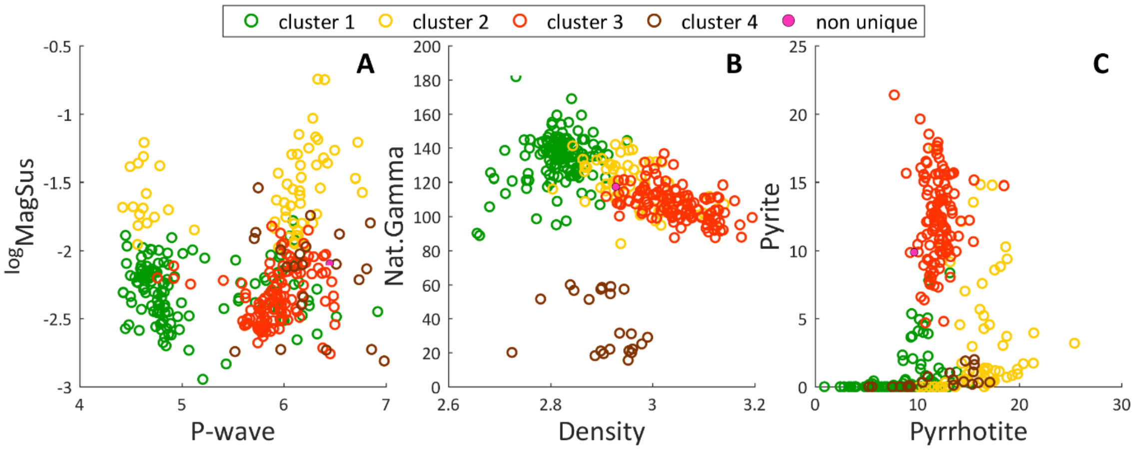


Figure 4.7: Cluster results for density, magnetic susceptibility and natural gamma plotted on different scatter plots; clusters as previous. (A) illustrates how the data is clustered in relation to magnetic susceptibility ($_{\log}$MagSus) and Vp; in contrast to Figure 4.6, more low-velocity samples are included in cluster 2 and high-velocity samples are grouped with poorly mineralised samples of cluster 1. (B) shows the clusters on a natural gamma versus density plot; compare to Figure 4.6 B. (C) are the clusters in relation to pyrite and pyrrhotite content; in contrast to Figure 4.6, cluster 3 only contains pyrite-rich samples, which are of economic value, but some pyrite-rich samples are still grouped with cluster 2.

hole DDH01 is not only influenced by the abundance of sulphide minerals, but also by other possibly textural features of the rock. There is no objective measure of texture available for these rocks, but description of textural characteristics from core photos indicate that the high velocity areas are rather massive rocks with little laminar features such as bedding planes and foliation, whereas the low-velocity intervals coincide with strongly laminated typically finer-grained rocks that would also exhibit a different overall mineralogy. If Vp is more correlated to texture than mineralisation, then this may be exploited to distinguish different rock textural features via clustering.

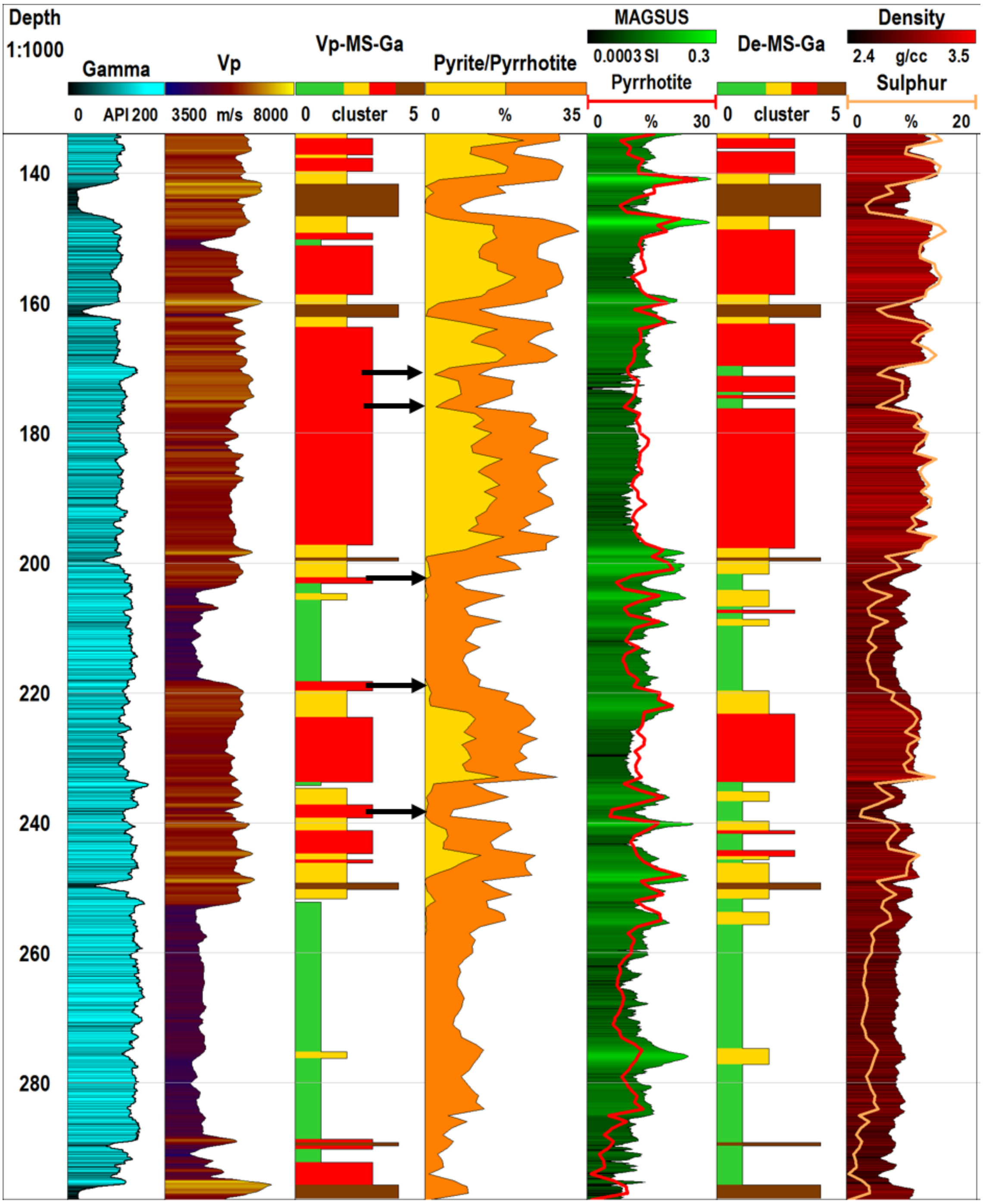


Figure 4.8: Well-log plots comparing the cluster results to the input petrophysical logs and sulphide mineralogy. The log-plot build from the clusters of Vp, magnetic susceptibility and natural gamma (Vp-MS-Ga) shows some intervals with high Vp but containing negligible amounts of pyrite that were wrongly grouped with cluster 3 (red, black arrows indicate misclassification). The log-plot build from clustering density instead of Vp (De-MS-Ga) on the other hand groups these intervals with cluster 1 (green) that represents barren rocks. Cluster 2 (yellow) coincides with areas of high magnetic susceptibility indicating high pyrrhotite content. Cluster 4 corresponds to the low gamma units indicative of dolerites in both cluster outcomes (Note that the scale of the cluster results (0–5) is set for readability, the actual number of clusters is 4).

4.1.4 Petrophysical measurements for distinguishing mineralogy and rock texture

This example aims to separate rocks of different mineralogical compositions that also exhibit different textural features. Data of the unmineralised section (no sulphide minerals present) of drill hole DDH01 was chosen for this purpose (Tapanappa formation, upper 134m). The Tapanappa formation constitutes quarzitic to pelitic successions that are characterised by gradations ranging from coarse grained, quartz-rich basal units to laminated, more mica-rich units at the top of each cycle. These successions represent sedimentation during marine transgressions and the original bedding is preserved. Two of these transgression cycles are apparent at the base of the Tapanappa formation in drill hole DDH1 (Figure 4.11, gradually decreasing silica content in the green log on the left from 134m to 125m and from 120m to 105m) whereas the remainder of the formation is constituted mainly of the finer grained meta-pelitic rocks with some quartz-rich, coarser grained sections, cross-cut by occasional dolerite dykes. The distinct rock types have different physical and chemical/mineralogical properties which are reflected in some of the petrophysical logs.

Table 4.3: Physical properties of common rock-forming minerals

Mineral	Density [g/cc]	Hardness [Moh's]	Vp [km/s]	Vs [km/s]
Quartz	2.65	7.0	5.9–6.0	3.9–4.0
Plagioclase	2.6–2.7	6.0–6.5	6.2–6.6	3.2–3.7
Muscovite	2.7–2.8	2.5–3.0	5.1–5.8	2.8–3.3
Biotite	2.7–3.3	2.5–3.0	4.8–5.2	2.0–2.8

Quartz and plagioclase possess higher hardness and therefore, on average, a higher p-wave velocity than the micas (Table 4.3). This may be slightly offset by the higher density of the micas compared to quartz and feldspars due to the increase of sonic velocity with density. In addition, the quartz-rich metasediments are typically coarser grained and massive, whilst pelitic units are fine-grained and anisotropic due to preferred oriented micas. Coarse-grained quartz-rich textures can be positively correlated to the unconfined compressive strength of a rock, which in turn may be correlated to p-wave velocity (Tandon and Gupta, 2013). It seems therefore reasonable to choose at least the Vp log to cluster the data in terms of texture and/or mineralogy. Some of the important data variables are shown in the form of a dendrogram in Figure 4.9. Contrary to the dendrogram in Figure 4.4, which was created from all intervals from this drill hole, this tree is based on only the unmineralised samples and shows some interesting differences: the sonic velocities

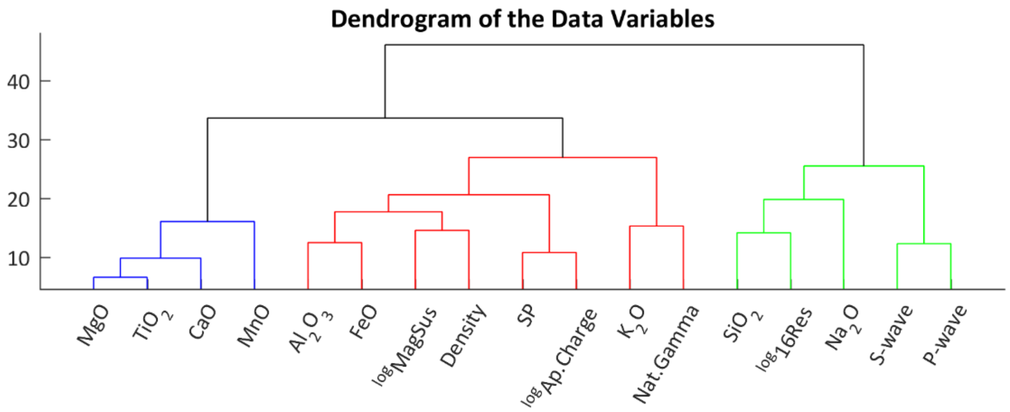


Figure 4.9: Dendrogram of some of the important variables from the unmineralised section of drill hole DDH01. In comparison to dendrograms of the entire data set, some variables are now grouped differently, markedly the sonic velocities are now grouped with silica content and resistivity instead of with iron, sulphur and other variables indicative of mineralisation.

are now grouped with silica content and resistivity, whereas aluminium, potassium and natural gamma are grouped with iron and the petrophysical variables that were previously indicative of the mineralisation.

To distinguish quarzitic from pelitic units, a combination of either Vp and resistivity or Vp and natural gamma might be suitable. The two subsets were clustered into three clusters; the Vp and natural gamma subset was also tested on four clusters. The results are shown as scatter plots in Figure 4.10 and as log-plots in Figure 4.11. The clusters are colour coded and labelled in the same way in both plots: (A) three clusters of Vp and 16-inch resistivity, (B) three clusters and (C) four clusters of Vp and natural gamma.

All three cluster solutions could separate the quarzitic and pelitic units quite successfully. Grouping the data into 3 clusters does give a ‘cleaner’, less busy looking classification in the pseudo-lithology logs in Figure 4.11, but separating into 4 clusters highlights the strongly pelitic units from intermediate units (clusters 1 and 2 in Figure 4.10 and Figure 4.11 C). It should be mentioned that the clusters in Figure 4.10 (bottom) show considerable overlap in terms of quartz and mica content of the respective clusters. This is not necessarily due to misclassification—comparison to core photos shows a good match—but could be attributed to the fact that the mineralogy was calculated from assay data and also because assay data has a different area of investigation compared to petrophysical logs. Nonetheless, the main trends are visible despite this overlap but cannot be verified in the absence of detailed and objective textural and mineralogical data that would have been beneficial in unravelling some of these relationships.

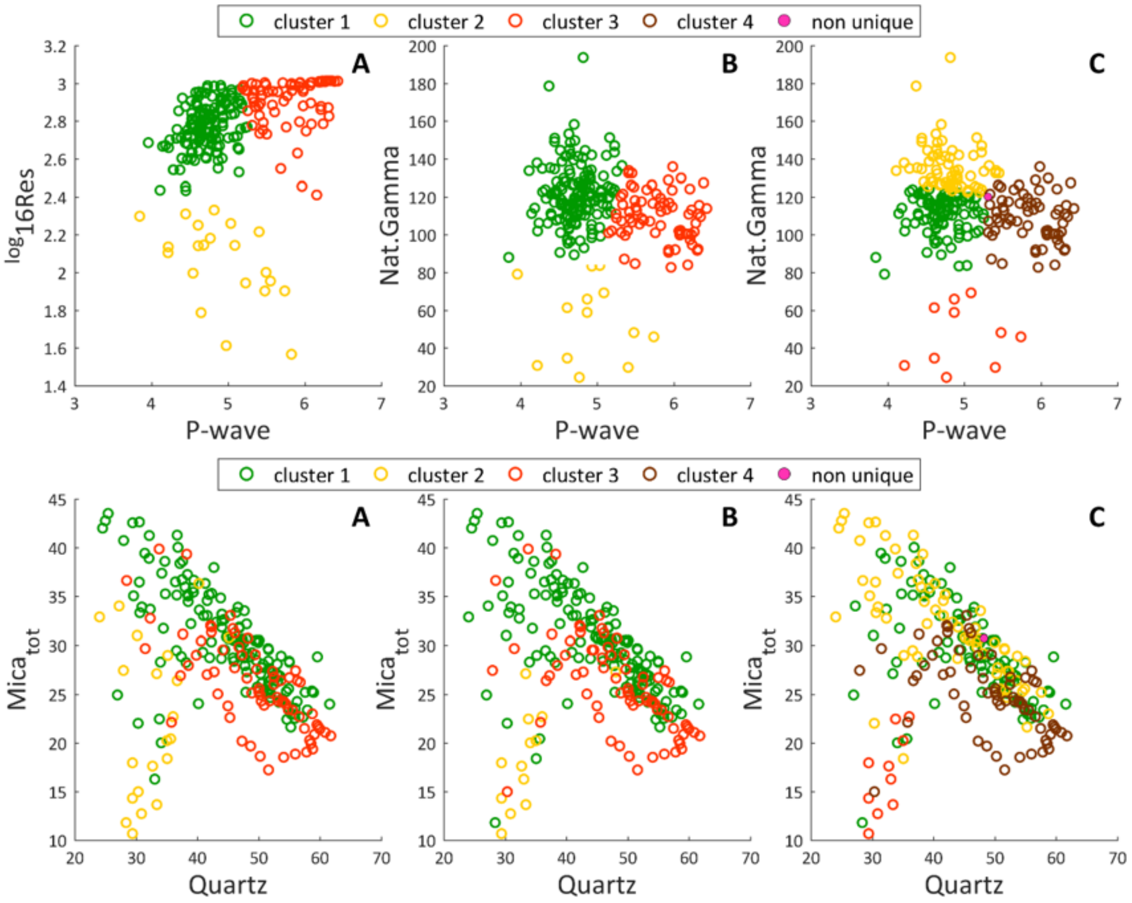


Figure 4.10: Results of clustering - top: axis represent the clustered variables; bottom: plotted on calculated quartz and mica content. (A) 16-inch resistivity and Vp, cluster 2 are primarily dolerite but also two slightly mineralised intervals that could not be separated (Figure 4.11), cluster 3 are mainly high velocity and highly resistive, quarzitic rocks and cluster 1 are more or less pelitic rocks; (B) Vp and natural gamma (three clusters), the clusters are quite similar to those in (A) separating high from low-velocity samples and low gamma samples, which represent the dolerite but this time not the slightly mineralised intervals since their gamma signature is higher (Figure 4.11); (C) Vp and natural gamma (four clusters), the main difference is that now a high gamma cluster is separated (cluster 2), which represents the highly pelitic units.

Summary

The learnings from the presented data analysis are manifold. First and foremost, it is vital to understand the relation between the data (measurements, analysis) and certain rock mass characteristics to increase the chances of successful classification. Some tools like dendrogram analysis or correlation coefficients between variables can help gain a better understanding of these aspects. Secondly, clustering a large number of input variables does not necessarily provide better results; a small set of carefully chosen variables out-performed the larger data sets in most classification attempts. Lastly, petrophysical measurements are of great importance when it comes to distinguishing rock textural features or even mineralogical differences.

Another example of where petrophysical data can be substituted for assay data to classify iron ore lithologies according to their economic value is shown in the following section.

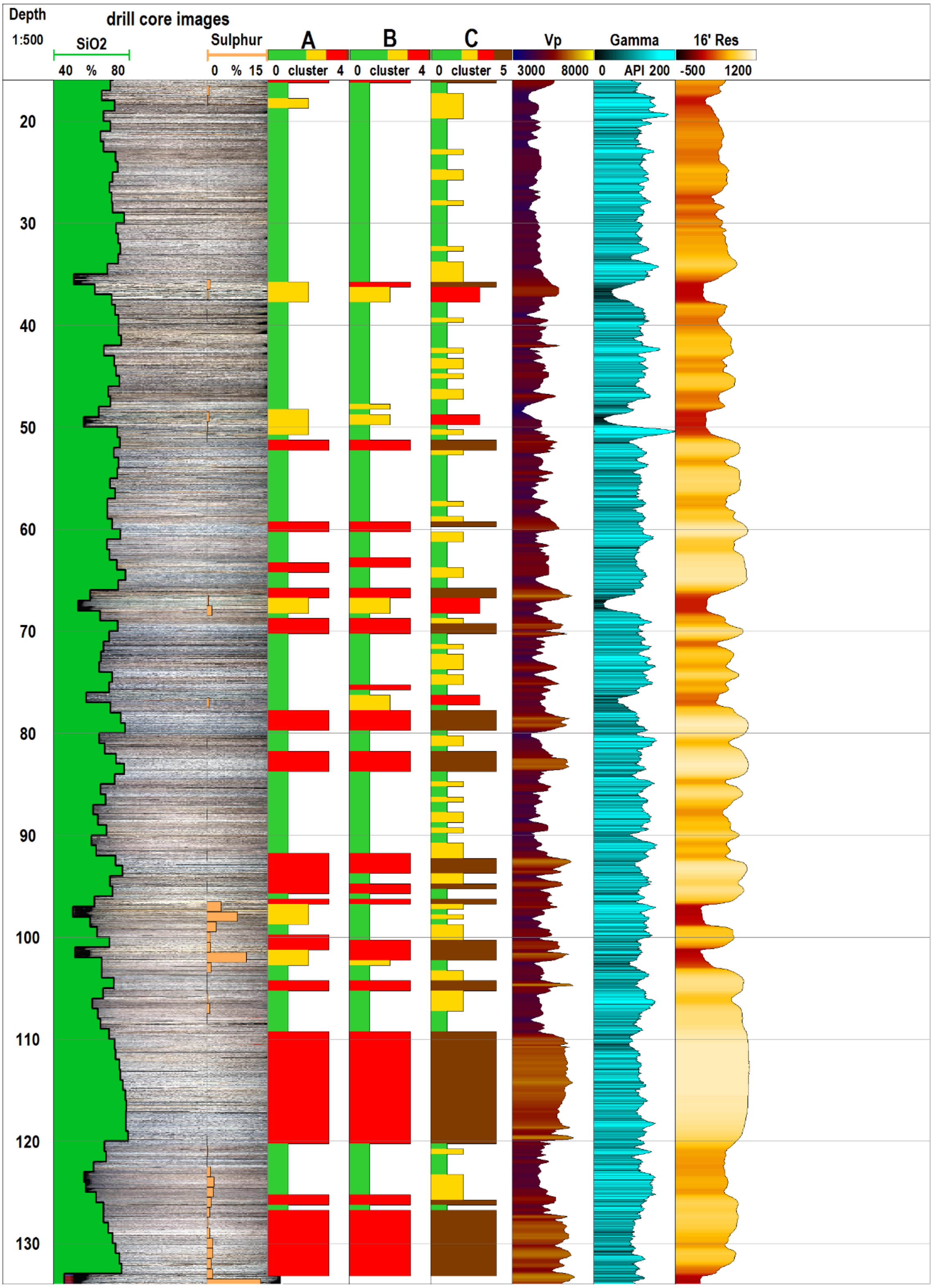


Figure 4.11: Well-log plots of the top ~130m of the drill hole DDH01. (caption continued on next page)

4.2 Classification of Pilbara iron ore lithologies

The classification of BIF-hosted iron ore deposits of the Pilbara region of Western Australia is usually based on industry standard cut-off values (personal communication and work experience with Rio Tinto Iron Ore) and aims to distinguish between high-grade and low-grade ore as well as shale and BIF waste units. The common method of classification is a workflow that relies heavily on manual input of specific codes into databases based on elemental assay results. Even this could be easily automated, by implementing nested ‘if-then’ statements (in Microsoft Excel for example) that group the data automatically. The following section evaluates the potential for automated iron ore classification using petrophysical measurements and FCM clustering. As before, different input variable combinations are tested and compared to a pre-established classification scheme based on the aforementioned industry cut-off grades.

4.2.1 Stratigraphy of Pilbara BIF-hosted iron ore deposits

Most economically valuable iron ore deposits of the Pilbara region in Western Australia are hosted by the Pre-Cambrian Hamersley Group that, together with the underlying Fortescue Group and overlying Wyloo Group forms the Mount Bruce Supergroup (Figure 4.11). The Fortescue Group, with a maximum thickness of ~4km, consists of mainly basic lava, pyroclastic rocks, shale and sandstone whereas the Wyloo Group can reach double the thickness and is comprised mainly of mixed clastic sediments with locally thick dolomites and basalts (Trendall, 1970). The stratigraphy of the Hamersley group, shown in Figure 4.12, is comprised of banded

---

...(continued) SiO_2 and sulphur assay data are plotted to the left, overlying a mosaic image of core photos. A, B, C represent the cluster solutions from Vp and resistivity, Vp and gamma (three clusters) and Vp and gamma (four clusters) in the form of pseudo-lithology logs. Geophysical downhole logs of Vp, resistivity and natural gamma are shown to the right. The core image log shows brighter and darker sections, which coincide with quarzitic (high SiO_2) and pelitic intervals (lower SiO_2 and high gamma). The quarzitic intervals are identified as cluster 3 (red) in A and B. Cluster 2 (yellow) in A and B are the dolerites except for the sulphide rich intervals in A at around the 100m mark. These could not be separated by clustering Vp and resistivity since both units are characterised by low resistivity and variable p-wave velocity. Solution C shows the dolerites in red (cluster 3) and the quarzitic units in brown (cluster 4), cluster 2 are the high gamma, pelitic units and cluster 1 are intermediate meta-siltstones. Note the two transgressive cycles mentioned in the text, apparent by the gradual decrease in SiO_2 content (green bars on the left) from 134m to 125m and from 120m to 105m.

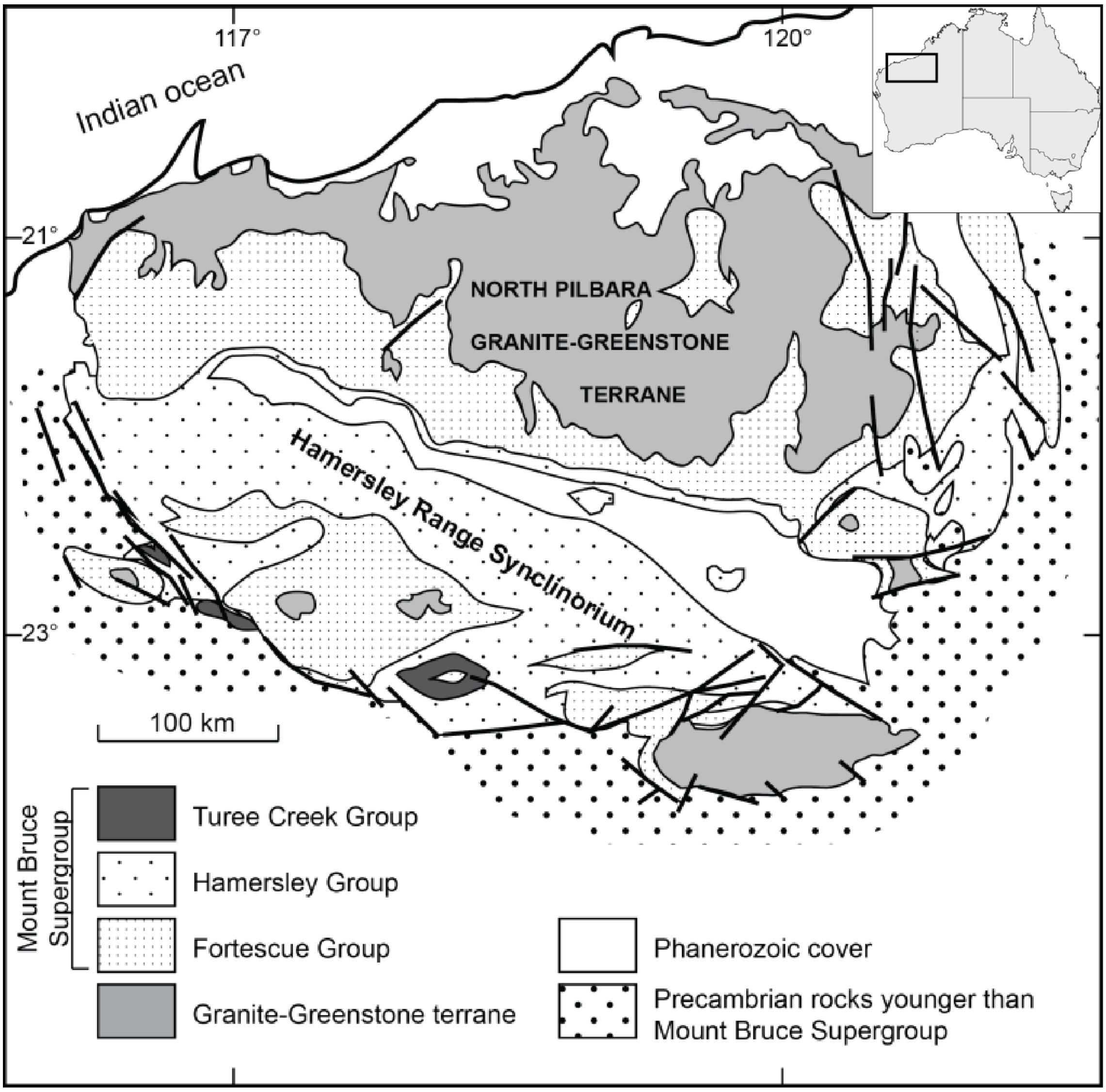


Figure 4.11: Geological overview of the Pilbara region in Western Australia. The economically valuable banded-iron formations are hosted mainly by the Hamersley Group of the Mount Bruce Supergroup (after Trendall (1970)).

iron formations separated by major shale, carbonate or volcanic units cut by locally abundant dolerite dykes. The iron formations from oldest to youngest are: The Marra Mamba formation, the Brockman formation (with the important Dales Gorge and Joffre members) and the Boolgeeda formation. The iron formations exhibit internal macrobanding characterised by alternating BIF and shale units, and are used for stratigraphic correlation due to their remarkable lateral continuity. The Dales Gorge member is comprised of 17 BIF macrobands separated by 16 shalebands that are numbered 1 to 16 from oldest to youngest (Trendall, 1970). The data clustering and prediction shown in the following aims to identify and separate these different units according to their economic value. The banded iron formations, in their primary sedimentary (and metamorphosed) state, consist of chert-magnetite mesobands representing an uneconomic waste unit due to their high silica content.

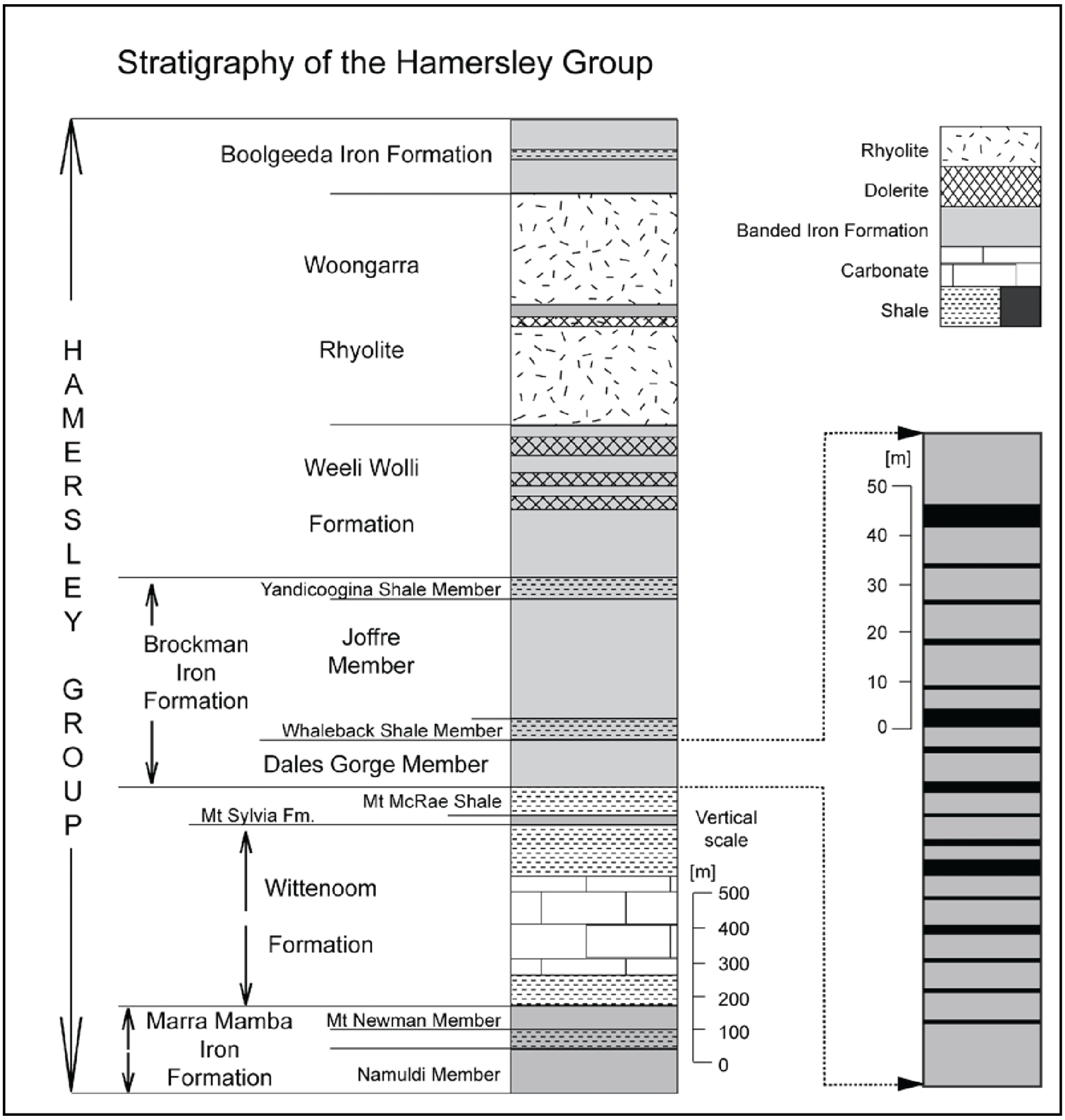


Figure 4.12: Stratigraphy of the Hamersley Group illustrating the position of the important iron formations in relation to major uneconomic shale, carbonate or volcanic units. Shale interbeds, illustrated for the Dales Gorge member to the right, internally subdivide the iron formations. The 17 BIF macrobands (grey) are separated by the 16 shale bands (black) that are used for stratigraphic correlation (after Trendall (1970)).

The economic iron ore unit consist of haematite-rich concentrations where most silica has been leached from the rock mass. Analysing data variables that reflect these differences may result in a successful classification and prediction of iron ore lithology.

4.2.2 Clustering of different combinations of input variables

The data presented in the following sections was kindly provided by BHP and comprises major and some minor element chemistry as well as borehole geophysical density, magnetic susceptibility and natural gamma logs in 3m intervals (863 samples in total). The classification is based on a series of industry standard cut-off values and is defined as follows:

Class 1 – waste BIF – Fe% <50%, Al_2O_3% <3%;

Class 2 – waste shale – Fe% <50%, Al_2O_3% >3%;

Class 3 – low-grade ore – 50% < Fe% <58%, Al_2O_3% <3%;

Class 4 – high-grade ore – Fe% >58%, Al_2O_3% <3%.

Fuzzy c-means clustering, setting hard cluster boundary parameters of $m = 1.4$ and $\alpha = 0.4$, is tested on three different combinations of assay data first. These are (a) Fe, P, SiO_2, Al_2O_3 LOI, CaO, Mn, MgO, TiO_2, K_2O, S, Na_2O; (b) Fe, SiO_2, Al_2O_3 and (c) Fe, Al_2O_3. Figure 4.13 shows the results of clustering these input sets on the Fe versus Al_2O_3 scatter plots (B), (C) and (D) respectively; (A) shows the predefined classes for comparison. The clusters represent a more natural division of the data based on their similarities and differences and cannot reproduce the arbitrarily defined classes very well. As might be expected, the best results are obtained when clustering only those variables that were used to define the desired classes (Fe and Al_2O_3 in (D)). Compared to this desired classification, clusters 1, 3 and 4 from fuzzy c-means clustering contain samples with higher shale content. The application of such classification to mining and ore processing would lead to a dilution of grade and can cause problems with ore processing.

A dendrogram of the combined chemical and petrophysical variables (Figure 4.14) is used to determine which variables contain similar information and may be substituted for each other. There are only three petrophysical variables available for this data set: density, natural gamma and magnetic susceptibility. The natural gamma log, gathered routinely from every prospective iron ore drill hole, is indicative of the presence or absence of shale units. It might be expected that density correlates well with iron content, but the distinct characteristics of the BIF host rocks and the mineralised iron ore formations result in Fe% showing a bimodal distribution while the density is normally distributed (Figure 4.15). BIF is a usually dense, non-porous rock composed of micro-crystalline silica (chert) and iron oxides (mainly magnetite).

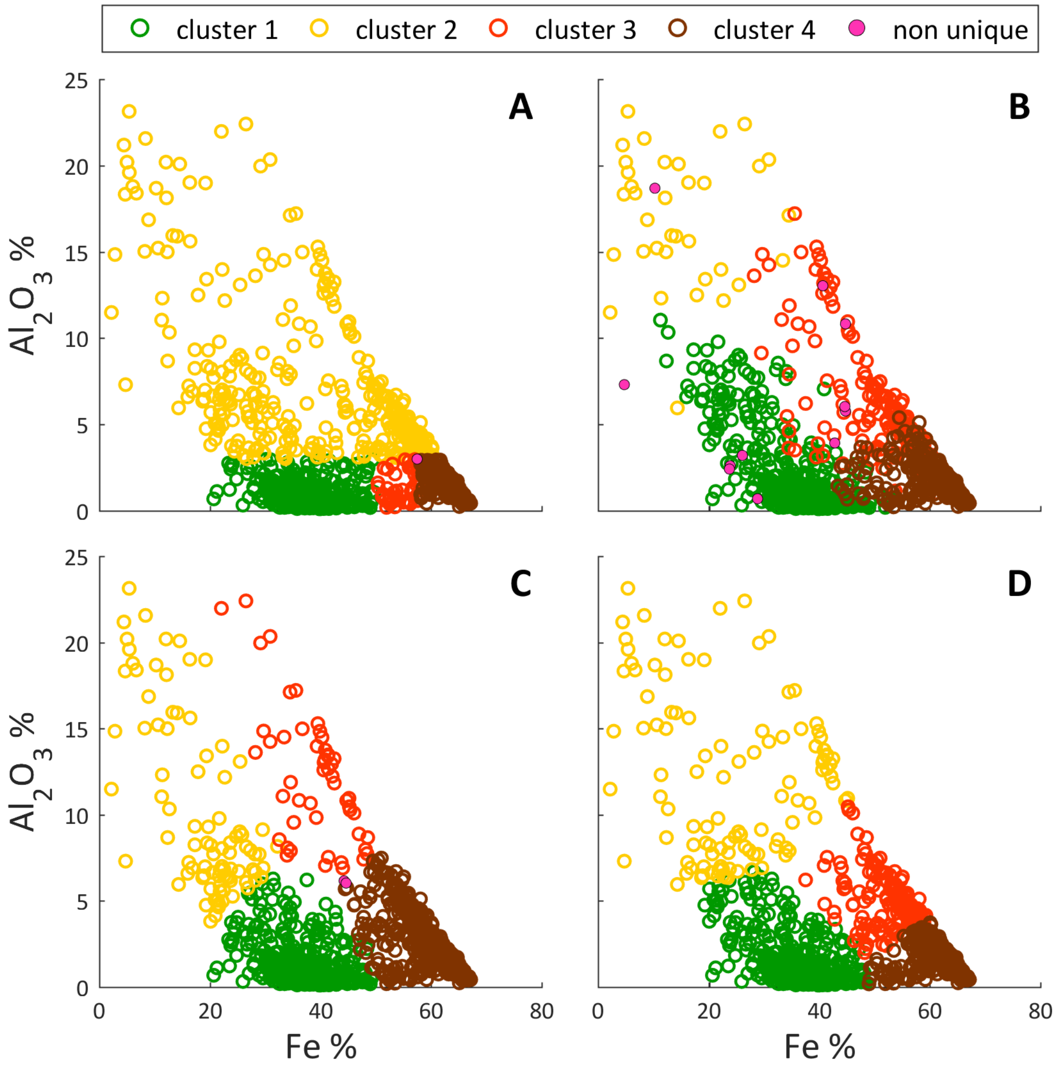


Figure 4.13: Results of clustering the three different elemental assay input data sets. A) Predefined classes based on cut-off values; B) all assay variables; C) Fe, SiO_2, Al_2O_3; D) Fe, Al_2O_3. Fuzzy c-means clustering results in a more natural division of the data. The best match with the desired classes is achieved when clustering the variables on which the predefined classes are based (D). The results are not satisfactory for grade control but distinguish well between major lithological changes.

On the other hand, the mineralised, economically valuable iron ore formations are leached of their silica content and composed of predominantly haematite, which can be extremely friable. As a result, both rock types can have the same density whilst their iron grade is markedly different (~40% for the BIF and 60–70% for ore formations, densities ~3g/cc). Consequently, distinguishing ore from BIF waste lithologies based on density and natural gamma may not be possible. The magnetic susceptibility log may be used to differentiate these if the iron oxide present in the BIF units is in fact entirely magnetite, but the plots in Figure 4.15 suggest otherwise. Although samples with the highest magnetic susceptibility are part of the low-iron group, both groups show high as well as low magnetic susceptibility. It seems doubtful that fuzzy c-means clustering of the available petrophysical data can yield anything close to the desired classification.

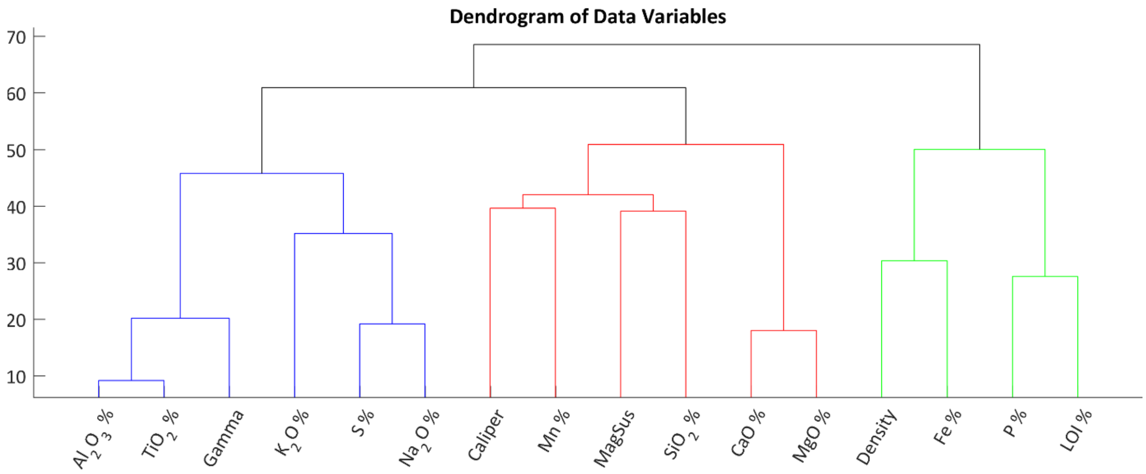


Figure 4.14: Dendrogram of the combined data variables shows a close relationship of density with Fe% and of natural gamma to Al_2O_3%, as would be expected.

Figure 4.16 shows the clusters of natural gamma and density compared to the predefined groups. High and low iron content could not be separated but samples with high shale content (high gamma signature) were identified in cluster 2. The scatter plot in Figure 4.16 C illustrates how the clusters were defined in terms of density and natural gamma signature. Since there are no other petrophysical data available that might correlate with iron content, a classification of iron ore lithologies, in terms of separating ore from BIF waste via clustering alone, is not possible for this data set. Even clustering the elemental assay data does not yield the desired results since the arbitrary predefined economic classes do not represent the natural data structure boundaries. In order to obtain the desired classes, it may be possible to use a different approach by training an algorithm for data prediction. The results of using this approach for this and a second iron ore data set, are presented in the next chapter.

Summary

Fuzzy c-means clustering groups data according to their natural characteristics and structure. Therefore, it is unlikely that arbitrary classes based on industry standard cut-off grades can be reproduced using this method, especially if the natural data division is substantially different to the desired classes. Clustering the variables that were used to define the industry standard classes (Fe% and Al_2O_3%) yields the closest match. Substituting density for iron content did not perform well due to

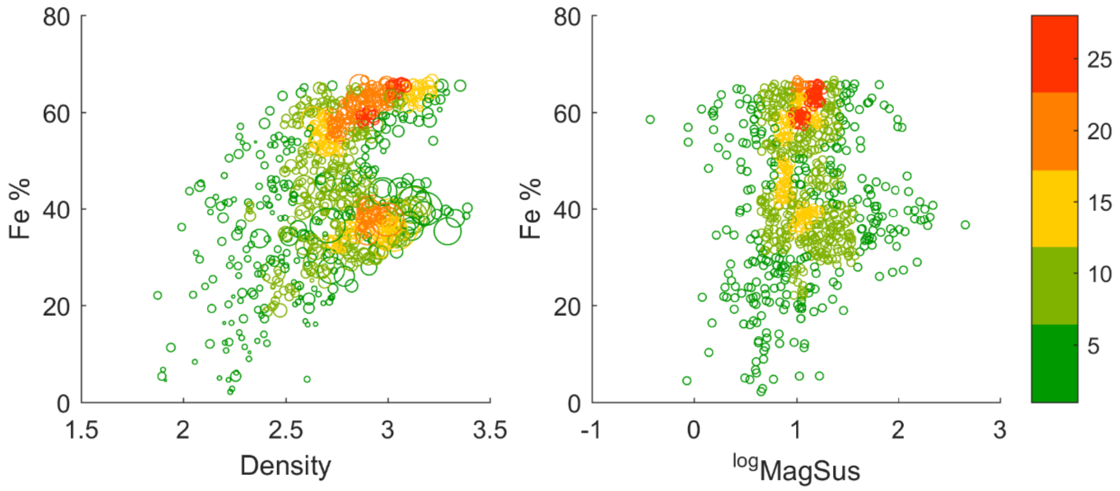


Figure 4.15: Heat scatter (data density) cross-plot of Fe% and density (left) and Fe% and log-transformed magnetic susceptibility (right). There are two apparent groups of iron concentration, which both have similar densities and magnetic susceptibilities. The size of the circles in the left plot represent the magnitude of magnetic susceptibility (not log-transformed). Distinguishing between the two groups of different iron content by clustering density and natural gamma may not be successful.

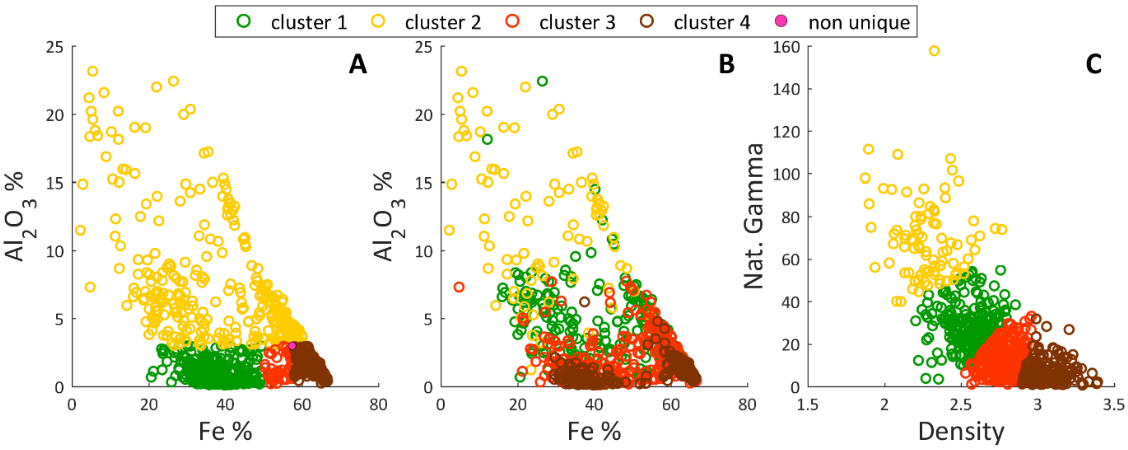


Figure 4.16: A) Predefined groups based on cut-off values. B) Clusters of natural gamma and density on Fe% versus Al_2O_3% scatter plots do not coincide with the desired classification; high and low iron concentrations (ore and BIF) are in the same clusters. Shale units are identified in cluster 2 and cluster 1 to some extent. C) The same clusters on density versus natural gamma scatter plots showing how the clusters are defined based on these two input logs.

similar density values for various iron concentration. Magnetite-bearing banded iron formations (waste units) can exhibit high density, whilst comparatively low values are recorded for the friable (porous), haematitic iron ore units. Although both iron oxide minerals have a similar density and magnetite is less abundant in the BIF compared to the concentrated haematite content in ore units, the different textural properties of the total rock mass result in similar density signatures.

Nonetheless, a successful iron ore classification using density and natural gamma may be possible through the application of adaptive neuro-fuzzy learning and prediction procedures. Such algorithms and inference systems are also a necessary next step to truly automatise the presented classification procedures. Using these predictive procedures to estimate iron content as means to perform grade control when the natural clusters span both ore and waste is investigated in the next chapter.

Chapter 5

DATA PREDICTION

Chapter 4 showed how the rock mass can be classified using different cluster methods and input data. Data clustering works well when the desired groups reflect the natural differences in rock properties, but is limited when an arbitrary division of the data is needed for grade control or when the data does not properly reflect the differences of the desired classes. This was the case for using the density to separate iron grade in the previous example. These problems can be overcome by using (for example) neuro-fuzzy inference systems to train an algorithm for data prediction. The training process tweaks the membership functions such that the predicted values better reflect the desired output classes. The general process is demonstrated on the same iron ore data, which was previously clustered and where the desired classes are based on strict cut-off values that do not reflect the natural data structure.

Neuro-adaptive learning algorithms can also help to achieve better classification results when the variable used for clustering does not sufficiently reflect the rock property of interest, as was the case for using density instead of iron content to separate ore from waste lithologies. An example is given in the following, where the spectral gamma-gamma ratio (SGG) is substituted for density to predict iron ore grade and waste lithologies. Although the correlation between the SGG ratio and iron content is higher than between density and iron content, the results from clustering the SGG ratio and natural gamma are still unsatisfactory. Using a trained fuzzy inference system to predict ore and waste units on the other hand, increases the classification success considerably.

Subsequent examples from the Brukunga sulphide deposit apply the learnings from the previous data analysis (Chapter 3) and data clustering (Chapter 4) to predict rock mass characteristics for new data of two other drill holes from this site. The procedure represents a blind test of the previous findings and shows how inference systems can pave the way for automated rock-mass classification.

The last example aims to predict one specific petrophysical measurement (sonic velocity) from a combination of others. The ultimate goal is to predict an important engineering property for mine and pit planning—the rock quality designation—from sonic velocities. Because sonic logs are not available for every drill hole in the project area, this procedure demonstrates how data prediction can help with missing data to facilitate better 3D coverage for seismic imaging/interpretation and geotechnical modelling.

5.1 Fuzzy inference systems for lithology prediction

Fuzzy inference systems are based on the concepts of fuzzy set theory and fuzzy logic introduced by Zadeh (1965, 1988) and their application to lithology prediction has been demonstrated in several studies (Bosch et al., 2013; Ilkhchi et al., 2006; Saggaf and Nebrija, 2003). A fuzzy set allows for partial membership of its elements to different groups or clusters to a degree defined by the membership value [0 1]; analogous to fuzzy c-means clustering. The fuzzy inference process maps a given set of input variables to an output by defining membership functions from prior fuzzy c-means clustering. The number of clusters dictates the number of membership functions per variable and the weighting exponent m defines the hardness (or fuzziness) of the cluster boundaries. The transformation of crisp input values (% of element etc.) into degrees of membership, represents the fuzzification process. The number of rules are also based on the number of clusters and are evaluated by means of a fuzzy operation using a logical operator (AND, OR) and operation method (min, max, probabilistic OR). The subsequent implication process truncates the rule-evaluating functions and combines them into a fuzzy set, one per rule, then these output sets are aggregated into a single fuzzy set, which is finally defuzzified to yield a single, crisp output value.

The general process described above is illustrated on the iron ore assay data previously used for clustering (Section 4.2.2) in the following text. The fuzzy inference system in this example has two input variables, Fe% and Al_2O_3%, three rules and one output variable, which is the desired class value. Three membership functions per variable are defined from fuzzy c-means clustering of these two variables into three clusters, applying a weighting exponent of $m = 1.6$. This clustering process represents the first step of the fuzzy inference process and is built into the MATLAB function that generates the inference system (Appendix I). Only the defined membership functions can be accessed after this step but not the clusters of the data itself. For illustrative purposes the same data was clustered with the basic FCM function; the clusters are shown in Figure 5.1 and the related membership functions in Figure 5.2. The output functions are built, based on the desired output classes, which are slightly modified from the predefined classes for clustering (Section 4.2.2) to better reflect the industry standard classification:

Class 1 – waste BIF – Fe% <50%, Al_2O_3% <3%;

Class 2 – waste shale – Fe% <55%, Al_2O_3% >3%;

Class 2.5 – shaley ore – Fe% >55%, Al_2O_3% >3%;

Class 3 – low-grade ore – Fe% >50% and < 58%, Al_2O_3% <3%;

Class 4 – high-grade ore – Fe% >58%, Al_2O_3% <3%

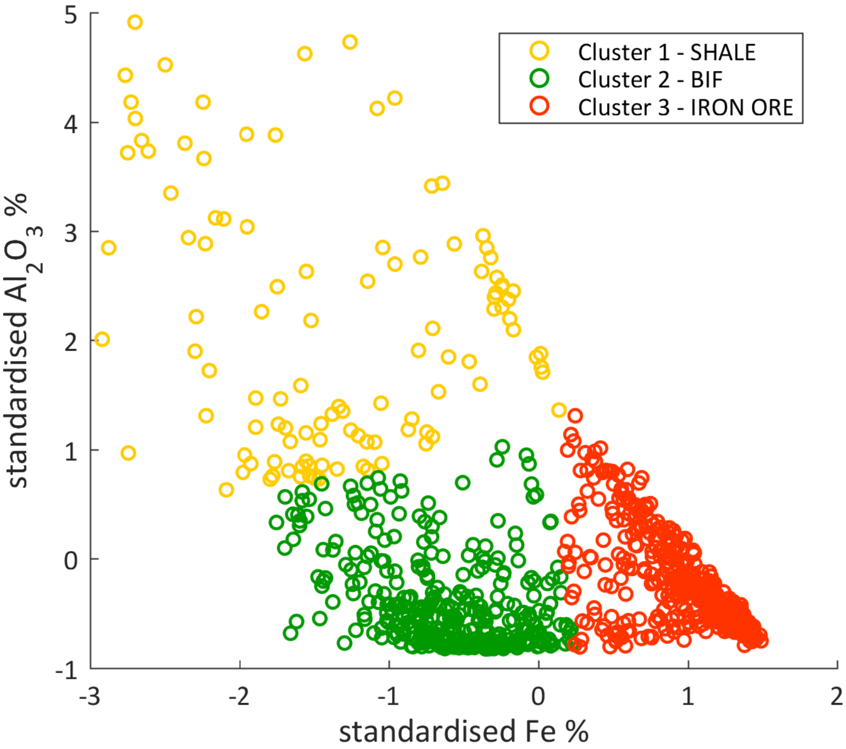


Figure 5.1: Clusters of standardised Fe% and Al_2O_3% data. The membership functions and rules of the fuzzy inference system are defined from these clusters. The resulting membership functions and rules are plotted in Figure 5.2 and Figure 5.4.

The three rules defined for this inference system are connected by a logical ‘AND’ operator and ‘PRODUCT’ method. Written out they are:

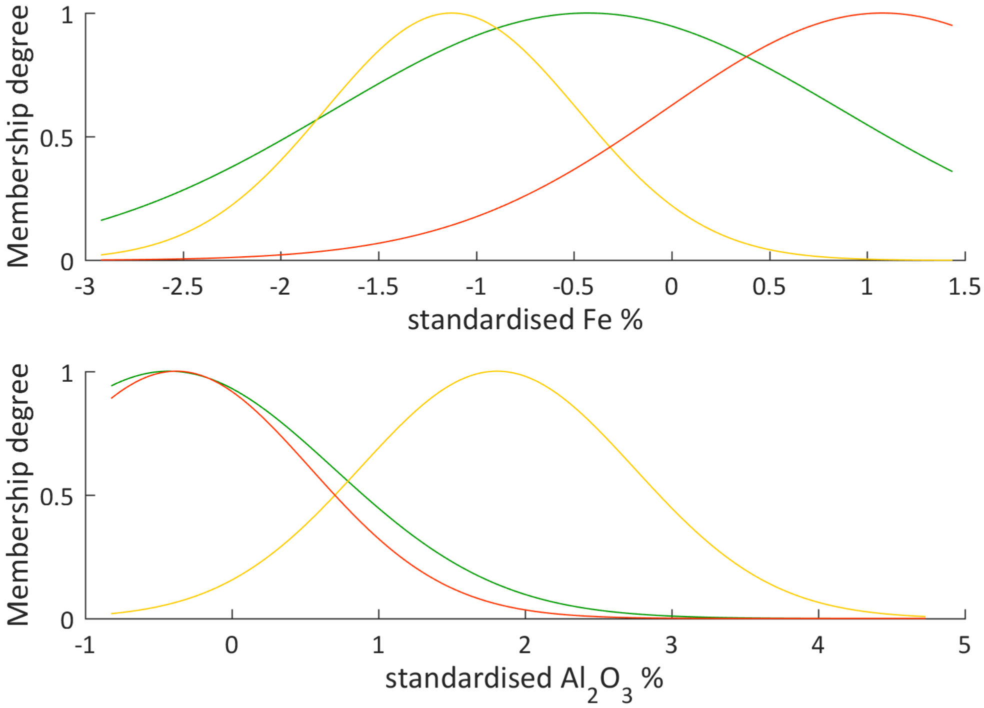


Figure 5.2: Membership functions for the two input variables before training of the fuzzy inference system. The yellow functions define membership to cluster 1 (SHALE), the green to cluster 2 (BIF), the red to cluster 3 (ORE).

> if input 1 is in cluster 1 and input 2 is in cluster 1 then output is class 1;
>
> if input 1 is in cluster 2 and input 2 is in cluster 2 the output is class 2;
>
> if input 1 is in cluster 3 and input 2 is in cluster 3 then output is class 3.

The results of the rule evaluation, implication, aggregation and defuzzification are then mapped to the output variable to yield answers in the desired range of classes from 1 to 4 as opposed to classes 1 to 3 from the input and rule evaluation. After the initial inference system is designed, it is trained via a neuro-adaptive learning process that adjusts the parameters of the membership functions to better track the output data. A hybrid training method that uses both least squares and back propagation algorithms is run for 100 epochs to train the system. A checking data set, modified by small amounts of random noise, is evaluated in conjunction to training and the best-suited trained fuzzy inference system is chosen based on the lowest training and checking error. The adjusted membership functions plotted in Figure 5.3 are considerably narrower when compared to the initial functions, especially the functions defining the cluster containing iron ore samples. It should be noted that all data was standardised via the z-score prior to analysis, which is why the absolute values in most plots are different from the numbers used to define the desired classes.

The structure of the transformed data is identical to the raw data, as demonstrated in Section 3.2. The fuzzy inference process, in regards to evaluation of rules and generating an output value, for the initial and the trained systems is illustrated in Figure 5.4. In addition to the parameters described above, the 'minimum' is used as the implication method to truncate the function from rule evaluation and the aggregated results are defuzzified using the weighted average method. The top plot shows the process and the resulting class before training for an input of Fe% = 1 and Al_2O_3% = 0 (standardised values). The untrained system yields a class value of 4.08, which represents high-grade ore, but the high value for aluminium suggests a high shale content and the sample should actually be classified as class 2 or 2.5. The value of 2.17 for the trained system represents a value in the correct range, indicating that a successful classification using this inference system may be possible. The results of generating, training and testing a fuzzy inference system for iron ore classification is presented in the next section.

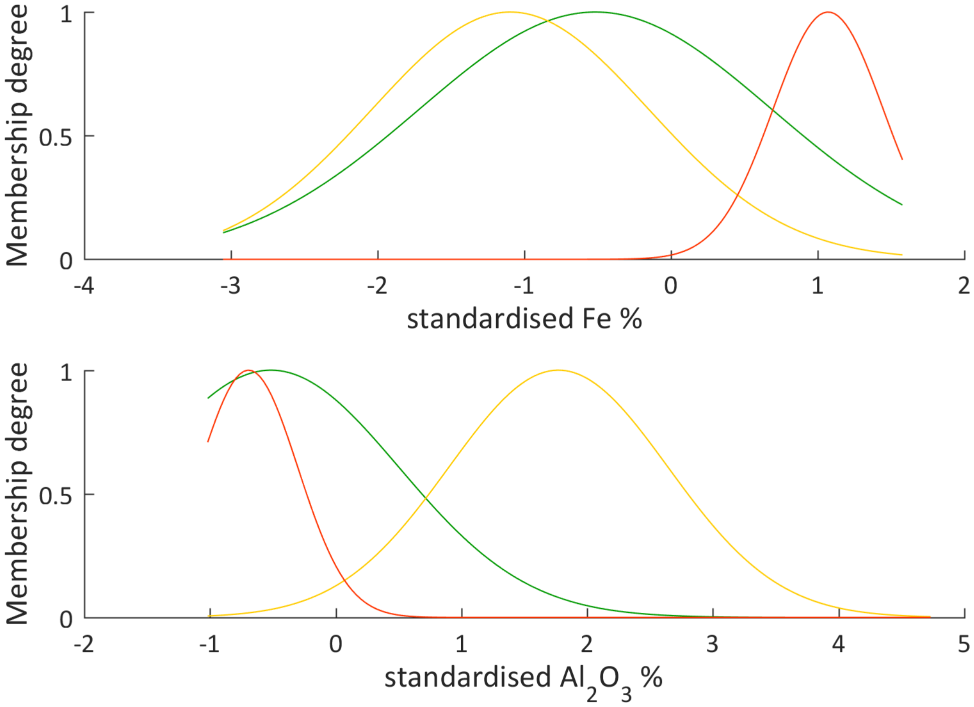


Figure 5.3: Adjusted membership functions after training. The functions are spanning a considerably lower range than the initially defined functions in Figure 5.2 and thus reflect the desired output classes better.

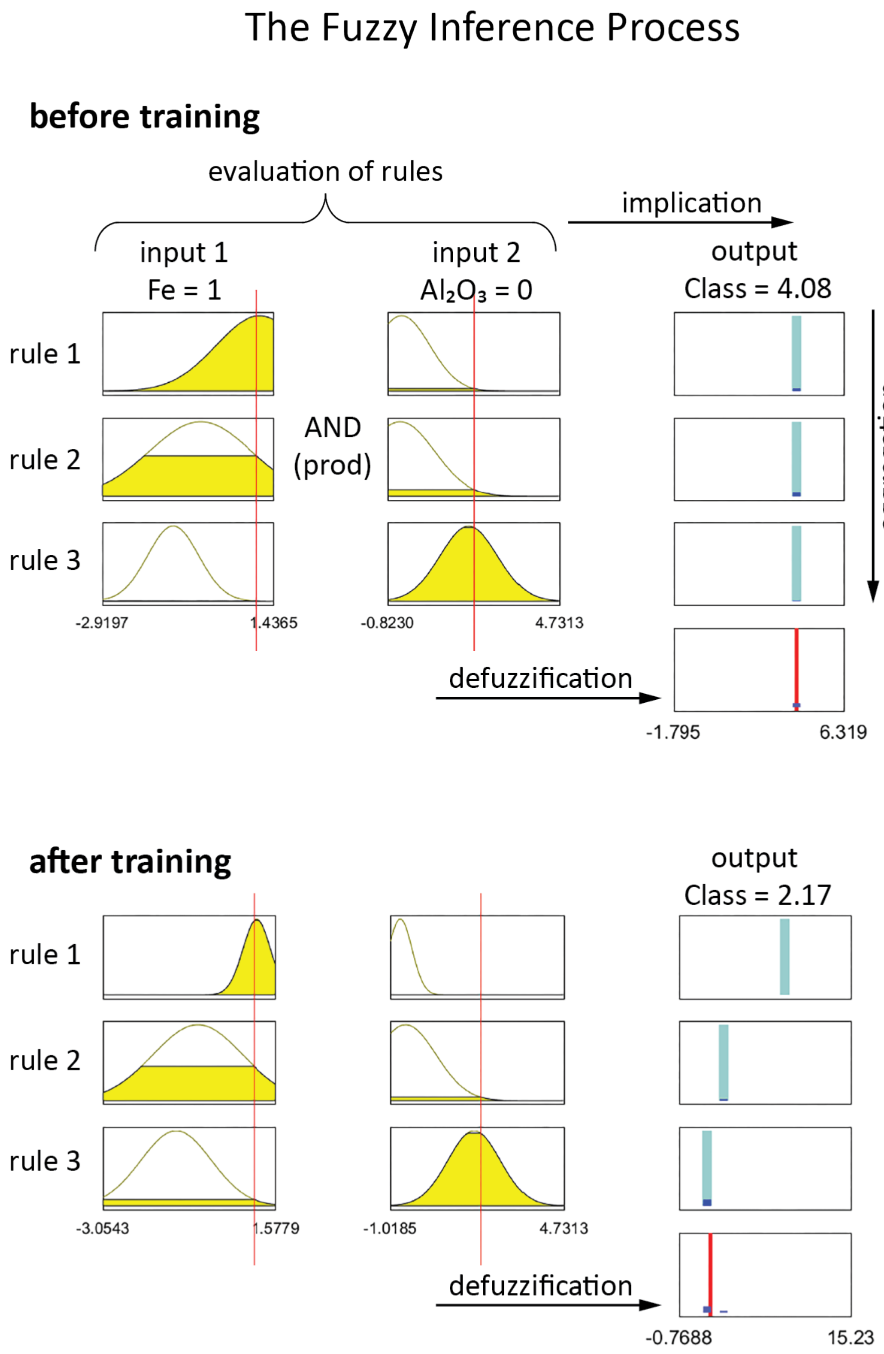


Figure 5.4: Illustration of the fuzzy inference process in regards to rule evaluation and output value generation. The three rules and three membership functions per input variable are the result of initial fuzzy c-means clustering. All rules are evaluated simultaneously and a logical operator and method (AND/PRODUCT in this case) as well as the implication method applied to the rule evaluation results. The aggregated fuzzy output set is finally defuzzified by choosing the value of the weighted average of the aggregated output functions. In this example, the input value is high in iron and high in aluminium and its membership in terms of iron is to cluster one and partly to cluster 2 but in terms of aluminium it is to cluster 3. The minimum (implication method) is chosen to truncate the output function, which is a linear function in this case. The truncated outputs are aggregated into a single fuzzy set of which the weighted average value is chosen as the crisp output (class) value.

5.2 Predicting iron ore lithology from different input data

The general process of designing, training and testing a fuzzy inference system as described above was tested on the iron ore data set that was previously used for clustering. First, lithology and grade are predicted from Fe% and Al_2O_3% assay data to ascertain whether the classification success can be improved when using the fuzzy inference process instead of the basic FCM cluster process of Section 4.2.2. The aim is to show that an adaptive learning algorithm can be trained to predict 'strict' classes such as cut-off grades, which was not possible with FCM clustering. Similarly, in subsequent steps, the natural gamma log is substituted for Al_2O_3% and the density data is substituted for Fe% to investigate if the training and prediction procedures can overcome the problem where the data does not reflect the differences of the desired classes, for this example, where density is not indicative of iron grade.

From the total of 863 samples (same data as in Section 4.2.2), 500 were chosen to design, train and check the inference system and then the trained system was applied to the complete set of 863 samples. The rather high percentage of training data was chosen to ensure that a sufficient amount of different grades and lithologies are sampled for training. If the training data does not include a wide range of possible values, the prediction process cannot yield satisfactory results. This approach is valid for this example because the main aim is to demonstrate how automated iron ore classification can be achieved by using fuzzy inference modelling.

The inference process was implemented in MATLAB using built-in functions and run 50 consecutive times to evaluate its repeatability. The fuzzy inference system is generated based on fuzzy c-means clustering of the standardised input data into three clusters, using a weighting exponent of $m = 1.6$. Random noise in the range of -0.2 and 0.2 was added/subtracted from the standardised data to generate the checking data, which should be similar but not identical to the training data and is important to avoid over-fitting during training. Training and checking takes place simultaneously and a training and checking error is recorded after each epoch (the algorithm was set to train for 100 epochs). Two trained fuzzy inference systems (FIS) were saved and tested after this step, namely, the last one after 100 epochs (FIS 1) and the one with the lowest checking error (FIS 2). Only the FIS with the lowest checking error was used for subsequent prediction, but both are compared in terms of performance (correlation of prediction to desired output class) in Figure 5.5 using Fe% and Al_2O_3% as inputs. The top plot shows the initial and final training errors as well as the initial and lowest checking errors for the 50 consecutive runs; the middle plot shows the correlation between the prediction from FIS 1 and the desired output and FIS 2 and the desired output classes. The correlation is well above 97% for most of the individual runs for both systems and the errors are very low. A comparison between predicted and predefined class in the bottom plot confirms the successful prediction from iron and aluminium input variables and thus validated this procedure for automated iron ore classification from assay data. However, assay data is only available after a considerable amount of time has passed—not enabling any real-time decision making—unless the data is acquired using site-based technologies such as Lab-at-Rig systems or petrophysical downhole logs. To investigate if petrophysical logs may be used for near-real-time prediction of iron grade, two additional input sets were tested: (1) Fe% and natural gamma and (2) density and natural gamma. The results from cluster analysis described in Section 4.2.2, showed that substituting density and natural gamma for the iron and aluminium assay did not result in successful classifications. The results of predicting lithology from Fe% and natural gamma in Figure 5.6 indicate that a substitution of the gamma log for aluminium assay data is possible and successful with correlations above 94%. The additional

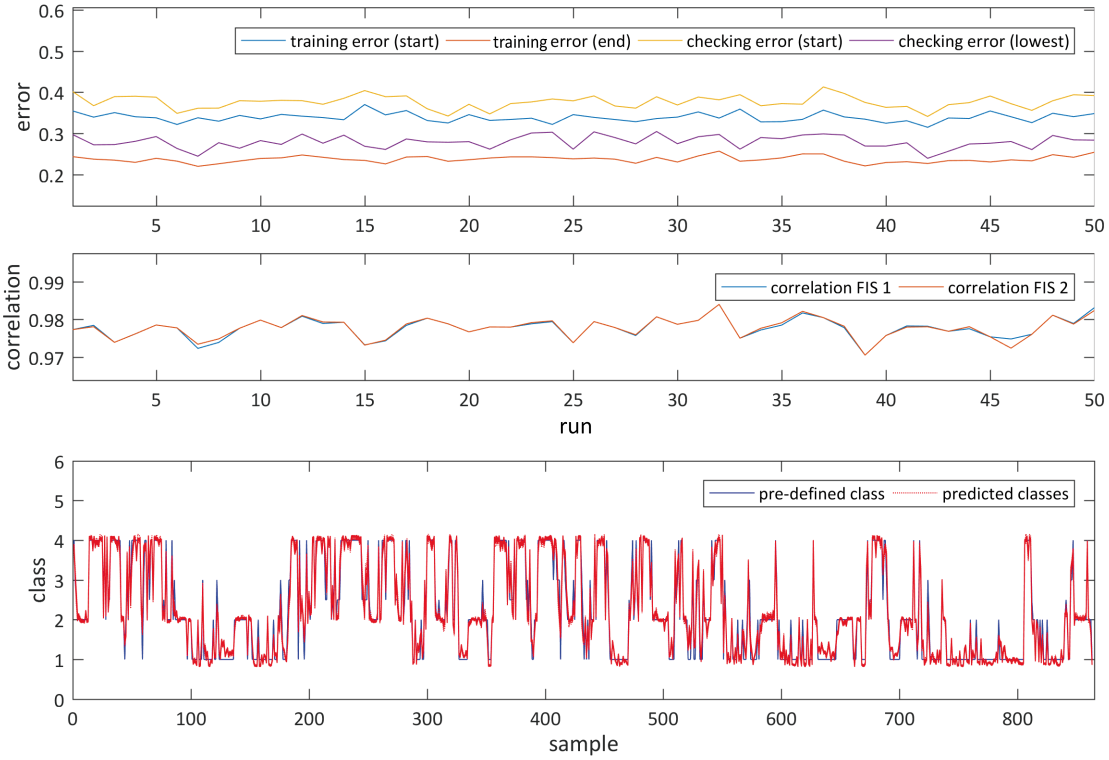


Figure 5.5: Results and parameters of the fuzzy inference system built from iron and aluminium input data. Top: training and checking errors for 50 consecutive runs; middle: correlation between prediction from the trained systems (FIS 1 is the last FIS from training (after 100 training epochs), FIS 2 is the system with the lowest checking error) and predefined classes indicating successful prediction (correlation above 97%); bottom: comparison of predicted and predefined classes from FIS 2.

substitution of density for iron however, yields unsatisfactory results (Figure 5.7), even after training. As pointed out before, there are two trends of iron concentration in the same density range and correlation of these variables is therefore low. The correlation coefficient of predicted versus predefined classes is below 0.5 (Figure 5.7, middle) and the errors are larger.

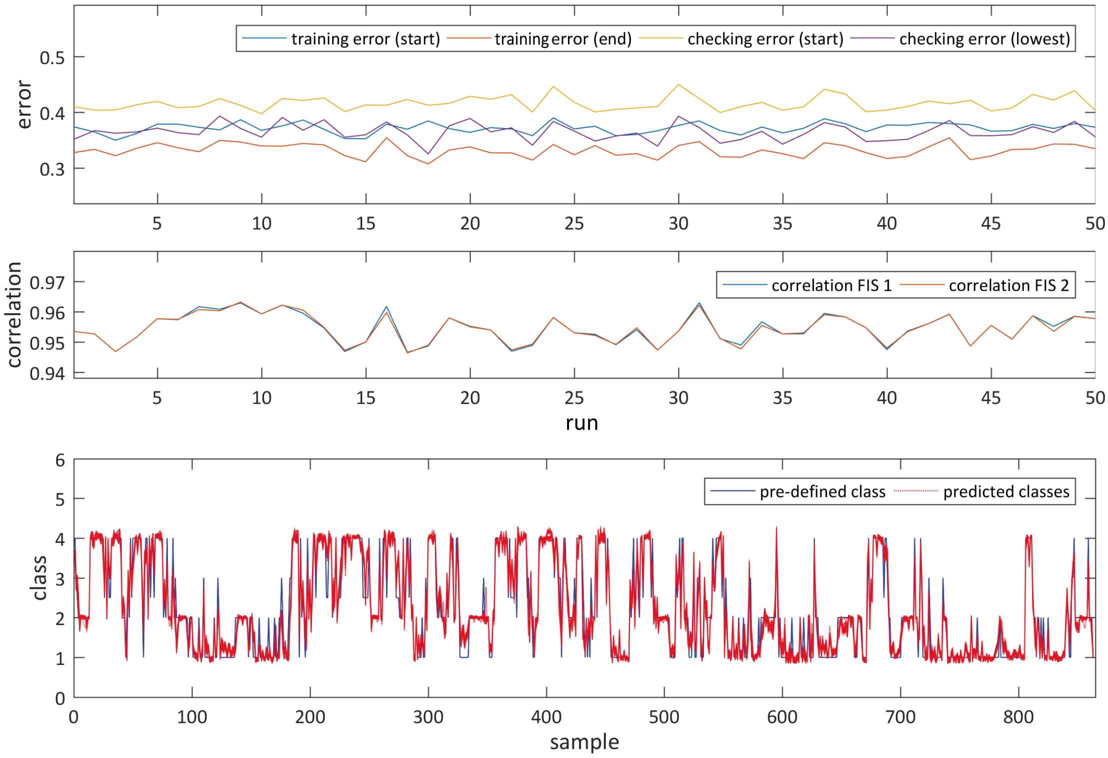


Figure 5.6: Results and parameters of the fuzzy inference system built from iron and natural gamma log input data. Training and checking errors are low, as in the case of predicting from Fe% and Al_2O_3%; correlation between predicted and predefined classes is high (>94%) showing that a substitution of the natural gamma log for the aluminium assay is feasible for predicting iron ore lithologies.

Summary

Fuzzy inference systems represent a valid tool to increase classification success compared to basic fuzzy c-means clustering if the desired classes are not considerably different from the initial clustering solution. The training through adaptive learning methods can tweak the algorithm such that a classification success of over 95% is achievable for assay data input as well as the substitution of natural gamma for aluminium oxide. This enables automated classification of iron ore based on assay data and can make current procedures more efficient by circumventing manual data entry procedures. For near-real-time applications this data would need to be collected on site via Lab-at-Rig systems or by using geophysical logging-while-drilling

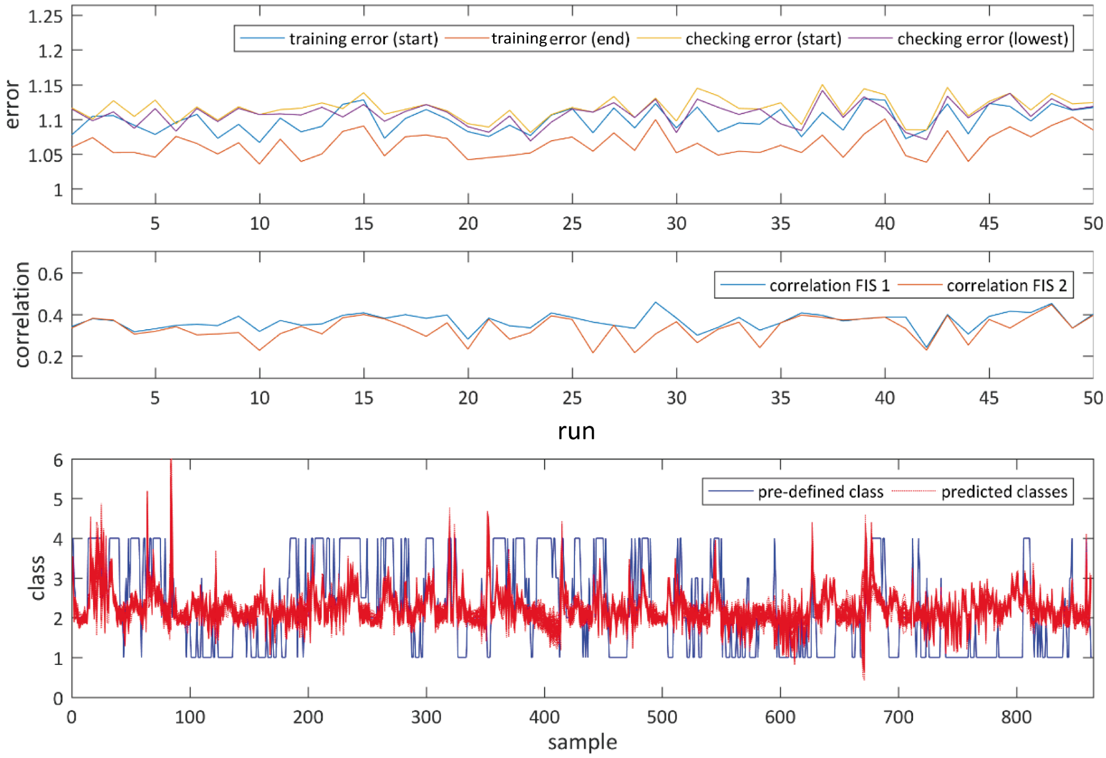


Figure 5.7: Results and parameters of the fuzzy inference system built from density and natural gamma log input data indicating that a substitution of density for iron content is not feasible for prediction of iron ore lithology.

technologies. Unfortunately, for this example, the inherent differences between density and iron content of the different iron oxide bearing formations do not allow for a successful classification based on density and natural gamma. If a real-time iron ore classification on the basis of petrophysical log data is desired, a different proxy for iron content independent of density is needed. The next section will investigate this possibility on a different iron ore data set.

5.3 Spectral gamma-gamma logging as a proxy for iron concentration

A gamma-gamma spectrum is recorded by means of the Lithodensity tool, which comprises of a radioactive source, usually ^{137}Cs (Caesium) with an energy of 662keV, and long- and short-spaced detectors, which are shielded from the source so that they will only detect gamma rays scattered from interactions with the formation. The depth of investigation and vertical resolution depends on the detector spacing and is usually in the range of 50–60cm. Different energy regions of the spectrum provide different information about the surrounding rock mass (Figure 5.8). The

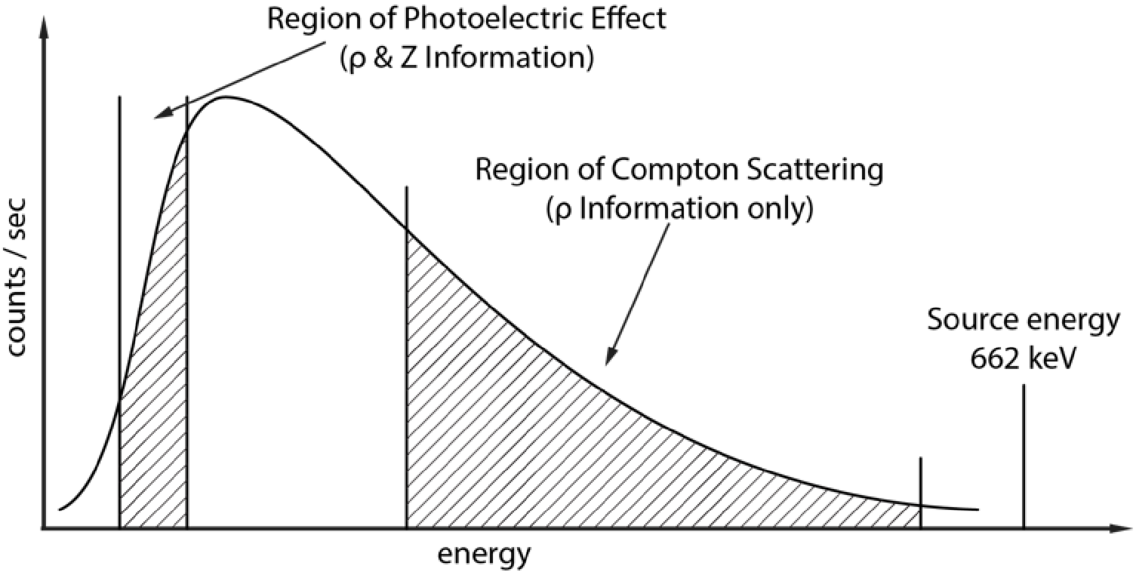


Figure 5.8: Sketch of the characteristics of a gamma ray spectrum recorded using a Lithodensity tool with a ^{137}Cs source. The high-energy region of Compton scattering is indicative of the formation density (σ) while the gamma counts recorded in the low-energy region are influenced by photoelectric absorption and indicative of density and average atomic number (Z) of a formation.

gamma counts in the high-energy region are indicative of Compton scattering and are used to calculate formation density (bulk density). The gamma ray counts in the low-energy region (below ~100keV) are influenced by photoelectric absorption interactions and give information about density as well as the formation's average (or effective) atomic number Z, which can be used as a lithology indicator. By taking the ratio of high-energy to low-energy counts, the density information is eliminated and the resulting number, the spectral gamma-gamma ratio (SGG ratio, (Killeen and Mwenifumbo, 1988; Killeen and Schock, 1991)), should only give information about the formations average atomic number:

$$SGG = \frac{(counts\ in\ high\ energy\ window)}{(counts\ in\ low\ energy\ window)}\ .$$

Due to this relationship and the fact that iron has a much higher average atomic number compared to the other abundant elements, the SGG ratio may be a good proxy for iron concentrations in iron ore formations. The aim of the following section is to verify this assumption. The second iron ore data set comprises spectral gamma-gamma logs recorded with an experimental Lithodensity tool (logged and made available by BHP) from four diamond drill holes from Pilbara iron ore deposits and supplementary major element assay data in 25cm intervals. The logging environment and parameters differ between these holes. Two drill holes that intersect the Dales Gorge member were logged through PVC casing and a Limestone matrix was applied for calibration (DH2 and DH3). A third drillhole (DH1), which intersects the Mount

Newman member overlain by the West Angela shale member and detrital cover, was logged through PVC casing and calibrated against a Limestone matrix. The remaining drill hole (DH4), which intersects the Joffre member, was logged through steel casing and a Sandstone neutron matrix was applied for calibration. Because of these discrepancies, the recorded gamma spectra from the different drillholes vary and the respective SGG ratios were calculated from different energy window ranges:

Drillhole	low energy window	high energy window
DH1	68–136keV	208–348keV
DH2	60–124keV	156–220keV
DH3	60–124keV	156–220keV
DH4	60–160keV	200–256keV

The data was shifted and transformed prior to data clustering and prediction so that they can be evaluated together. Scatter plots of the SGG ratio versus iron assay data are used to compare the individual trends for the four drillholes and extract the equations to adjust these trends to match (Figure 5.9). Drill hole DH1 was used as the reference hole and the SGG ratio from the remaining holes was shifted to match this trend. Since the vertical resolution of the spectral log is also much higher than that of the assay data, a 200 sample moving average filter is applied to the SGG ratio and the resulting log resampled to 25cm intervals to match the assay data for comparison. The total number of samples, combined from the four drillholes (1737 samples), is used for subsequent lithology and grade prediction. Two subsets are analysed and compared: (1) iron and aluminium assay data and (2) the SGG ratio plus natural gamma log. Both subsets are z-standardised and clustered via fuzzy c-means into four clusters, applying a weighting exponent of $m = 1.4$, which sets rather hard boundaries. The predefined classes are specified based on the same cut-off grades as used for clustering and prediction of the first iron ore data set:

Class 1 – waste BIF – Fe% <50%, Al_2O_3% <3%;

Class 2 – waste shale – Fe% <55%, Al_2O_3% >3%;

Class 2.5 – shaley ore – Fe% >55%, Al_2O_3% >3%;

Class 3 – low-grade ore – Fe% >50% and <58%, Al_2O_3% <3%;

Class 4 – high-grade ore – Fe% >58%, Al_2O_3% <3%.

First, the assay data was clustered and used for grade prediction and subsequently the natural gamma log and SGG ratio was clustered and used for grade prediction

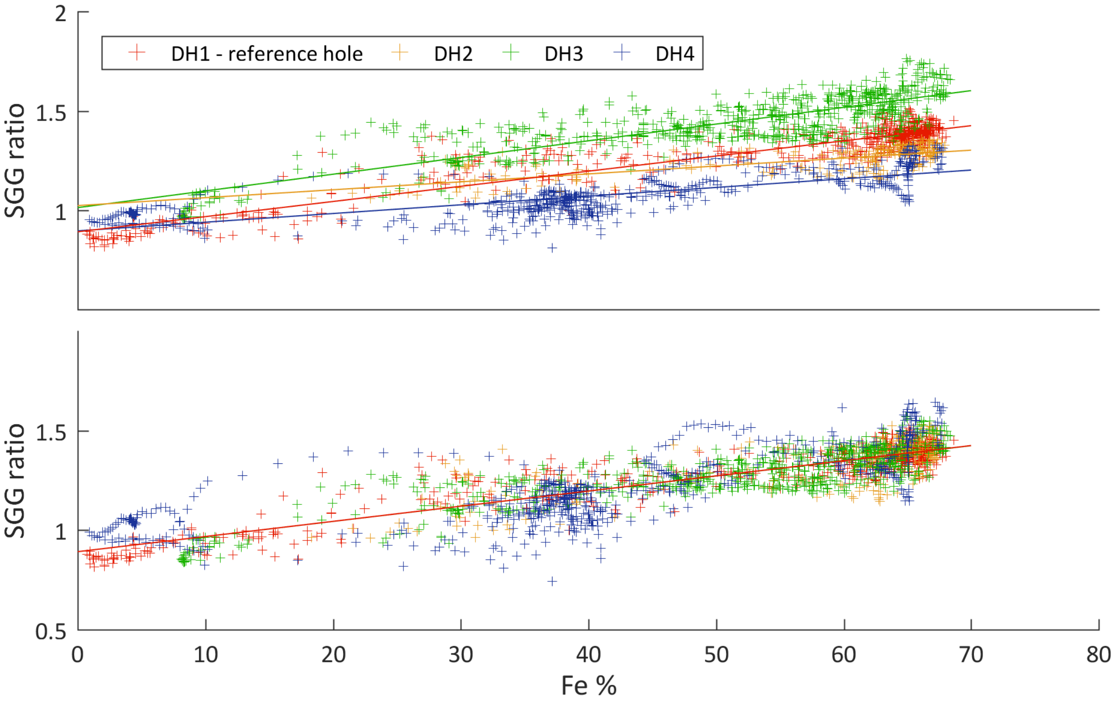


Figure 5.9: Scatter plots of SGG ratio versus Fe%. Top plot shows the trend of the raw (calculated) SGG ratio for the four different drill holes. The bottom plot shows the same data after shifting using drill hole (DH1) as a reference hole.

and the results compared. Fuzzy c-means clustering of the assay variables match the predefined classes by 77%, clustering of SGG and natural gamma only match by 66% (Figure 5.10 A and C). The FIS for grade prediction is based on four clusters extracted from fuzzy c-means clustering applying a weighting exponent of $m = 1.6$, allowing for slightly fuzzier boundaries. From the total of 1737 samples 1000 are selected for training and the trained system is applied to the entire dataset for prediction. The best-suited system is again selected based on the lowest checking error. Prediction results in Figure 5.10 B and D show a match of 92% for the assay data and 87% for prediction from the SGG ratio and natural gamma log, which is a very good result compared to the below 1% match of using density and natural gamma (Figure 5.7). These results are especially encouraging in light of the difficulties this data presented; collected from different sites with varying hole conditions and calibrations applied. A much higher success rate is expected when this methodology is applied under consistent conditions. Also, these results greatly outperform the success rate of the current real-time grade estimation, that is, visual logging by a trained geologist, which is estimated to be around 35% accurate (personal communication with iron ore geologists).

Summary

Density for iron ore classification, which at first seems a natural candidate as many iron-rich ores are heavy, delivers ambiguous values for these lithologies due to textural and mineralogical differences and is not sufficiently representative of the iron ore grade for lithology or grade prediction. The spectral gamma-gamma ratio is a measure indicative of the rock's average atomic number that is in the case of iron ore deposits mainly correlated to iron concentration; thus, it provides a means to estimate iron directly and may substitute for laboratory assay analysis of grade. The SGG ratio may be recovered from the gamma spectrum output of a Lithodensity tool, which is typically run in conjunction with a natural gamma log (stacked tools). This suggests that all the necessary data for predicting iron ore lithology accurately can be gathered in one single logging run and made available for interpretation and analysis in a timely manner after drilling; thus, enabling near real-time decision making.

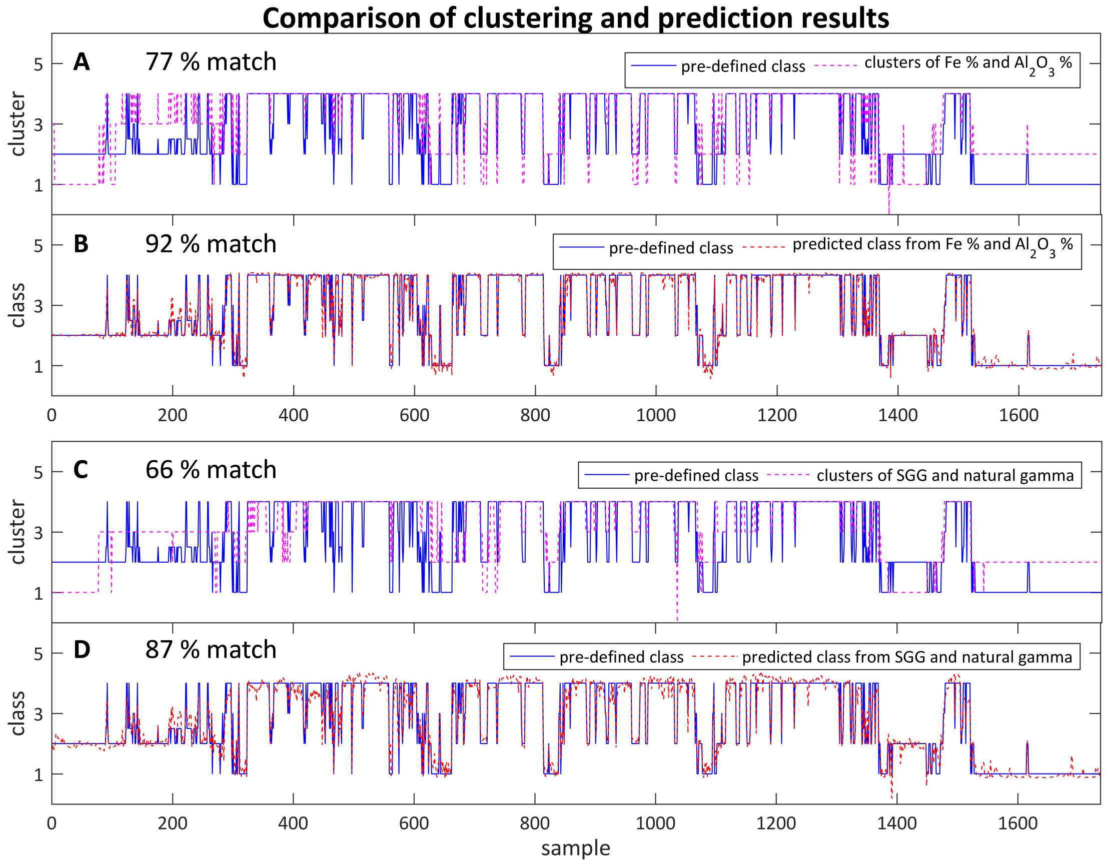


Figure 5.10: Comparison of clustering and prediction results from assay data (Fe%, Al_2O_3%) (A and B) and SGG ratio and natural gamma data (C and D). Clustering cannot recover the predefined classes sufficiently but a trained fuzzy inference system is able to predict the desired classes with high accuracy (B, D).

5.4 Predicting lithology and mineralogy from petrophysical data from the historic Brukunga pyrite deposit

Data sets from two additional drill holes from the historic Brukunga mine are available to test the classification procedures and confirm the relationships established from drill hole DDH01 in Section 4.1. Petrophysical downhole logging of these holes was carried out by a different logging contractor with some different tools and tool calibrations. The resolution of the tools may also be dissimilar in some cases. Therefore, the absolute values for a certain measurement, for example the resistivity log, may differ significantly from the values of the same measurement carried out some years earlier by a different contractor. This issue is illustrated in histogram plots in Figure 5.11. The largest discrepancies exist for the resistivity log because these logs were acquired with two different tools; DDH01 was logged with a traditional probe and conventional electrode spacing of 16 and 64 inches; in contrast DDH08 and STD02 were logged with a Dual Laterolog (focussed resistivity) probe. The data from both probes show a bimodal distribution of resistivity values but the data range is different. Both drill holes intersect the same formations but DDH01 was drilled perpendicular to the dip of the strata and DDH08 intersects the formation at an angle of about 145° and STD02 at an even higher angle (Figure 5.12). The density calibrations also seem slightly shifted between the holes but the natural gamma log

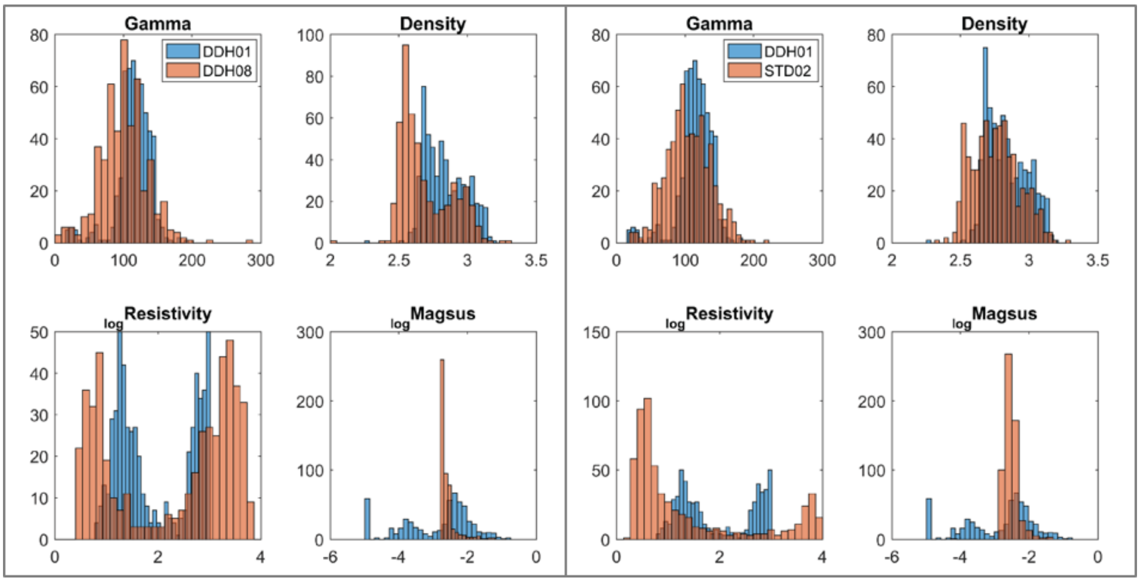


Figure 5.11: Histogram overlays of data from the recently logged boreholes DDH08 (left four plots, red bars) and STD02 (right four plots, red bars) over data from DDH01 (blue bars), logged some years ago. The data distribution is similar, but the absolute range of values can be quite different, for example, resistivity and magnetic susceptibility.

values cover about the same ranges, mainly because gamma tools are calibrated in a specifically designed test pit facility. The magnetic susceptibility is quoted in SI base units by both logging contractors but the values cover different ranges again, which may be due to tool configuration and sensitivity or processing and applied hole corrections.

Due to these discrepancies in the data, predicting lithology from the recent measurements of DDH08 and STD02, based on the training of the algorithm on data from DDH01 is not possible because the new input values are not in the same range as the membership functions built from the training data. Nevertheless, the learnings from clustering the data from drill hole DDH01 can be applied to the new data. It is expected that for clustering, only the relative values and not the absolute numbers are relevant. Thus, if the underlying similarities and differences of the rocks are reflected in the measurements then clustering can distinguish these lithologies.

5.4.1 Predicting lithology

The cluster analysis presented in Section 4.1.2 shows that the lithologies at Brukunga are best classified through fuzzy c-means clustering of natural gamma, resistivity, and density data (Table 4.1). To test the robustness and repeatability of this classification scheme the same log data from drill hole DDH08 was clustered, applying the same clustering parameters as in the previous tests. Assay data (for comparing the results) are only available from 150m to the end of the hole, which constitutes the mineralised part. For this interval, the cluster results match the predefined classes based on cut-off values (see Section 4.1.2), with 77% accuracy. The main discrepancies occur around class boundaries where the location of that boundary is offset by about one interval between the cluster results and the predefined classes.

Figure 5.12 illustrates the results and shows that stratigraphic correlation between holes is possible based on the identified clusters. Stratigraphic correlation is also possible from the resistivity logs, which also show the different resolution of the applied downhole tools. The leftmost logs are from drill hole DDH01, but while the changes in resistivity are apparent, the new logs from DDH08 (middle) and STD02 (right part) show a markedly better resolution and highs and lows are easily correlated. At a glance, the pseudo-lithology columns built from the clusters show a very similar distribution and are also easily correlated between holes. The dolerite dykes (cluster 4, brown) are present in all holes but since these dykes are not stratiform, they appear at slightly different depths in DDH01 and DDH08.

It should be noted that drill hole STD02 is situated about 1km south (along the strike) of the other two holes and that there are no laboratory assay data available for that hole; however, XRF spectral data acquired with a prototype core-scan XRF system was used for comparison. The pseudo-lithology column for STD02 is not based on clustering but represents the predicted lithology from the natural gamma, resistivity and density logs. The principal algorithm of fuzzy inference systems is described in Section 5.1; the FIS is built and trained based on the membership functions extracted from fuzzy c-means clustering of natural gamma, resistivity and density from DDH08 (input variables) and the respective cluster number (output variable). In other words, the training and prediction process depends solely on downhole data; no assay data was used to establish desired output classes. The trained algorithm was then applied to the petrophysical log data of STD02 and the resulting class values are shown in Figure 5.12 in the rightmost column. The height and colour of the bars represents the classes, which can take intermediate values due to the nature of the fuzzy inference process (Section 5.1). Although there is no desired output classification to which to compare the results, it is apparent from Figure 5.12 that the main lithologies are separated, the dolerites are identified, sections containing high-grade sulphide mineralisation are highlighted and a general stratigraphic correlation is possible despite the considerable distance to the training hole.

These results validate the repeatability of the cluster-based lithological classification from petrophysical downhole logging data and demonstrate how the process can be automated through fuzzy inference prediction. The same workflow as described in the last section is applied to predict the abundance of sulphide minerals in the following text.

5.4.2 Predicting the abundance of different sulphide minerals

As shown in Section 4.1.3, the natural gamma, density and magnetic susceptibility logs may be used to distinguish between the different sulphide minerals, pyrrhotite and pyrite. As pyrite is the mineral of economic value at the Brukunga deposit, predicting its abundance from downhole logs enables near-real-time evaluation of the mineralisation grade and distinction from the sub-economic pyrrhotite occurrences. Since assay data is available for the mineralised portion of DDH08, pyrite and pyrrhotite abundance was calculated and used to construct and train a relevant fuzzy inference system. The input variables are the above-mentioned logs and the desired output is the amount of pyrite. From the total of 195 samples from the mineralised part of DDH08, 95 (48.7%) were used to built and train the fuzzy inference system

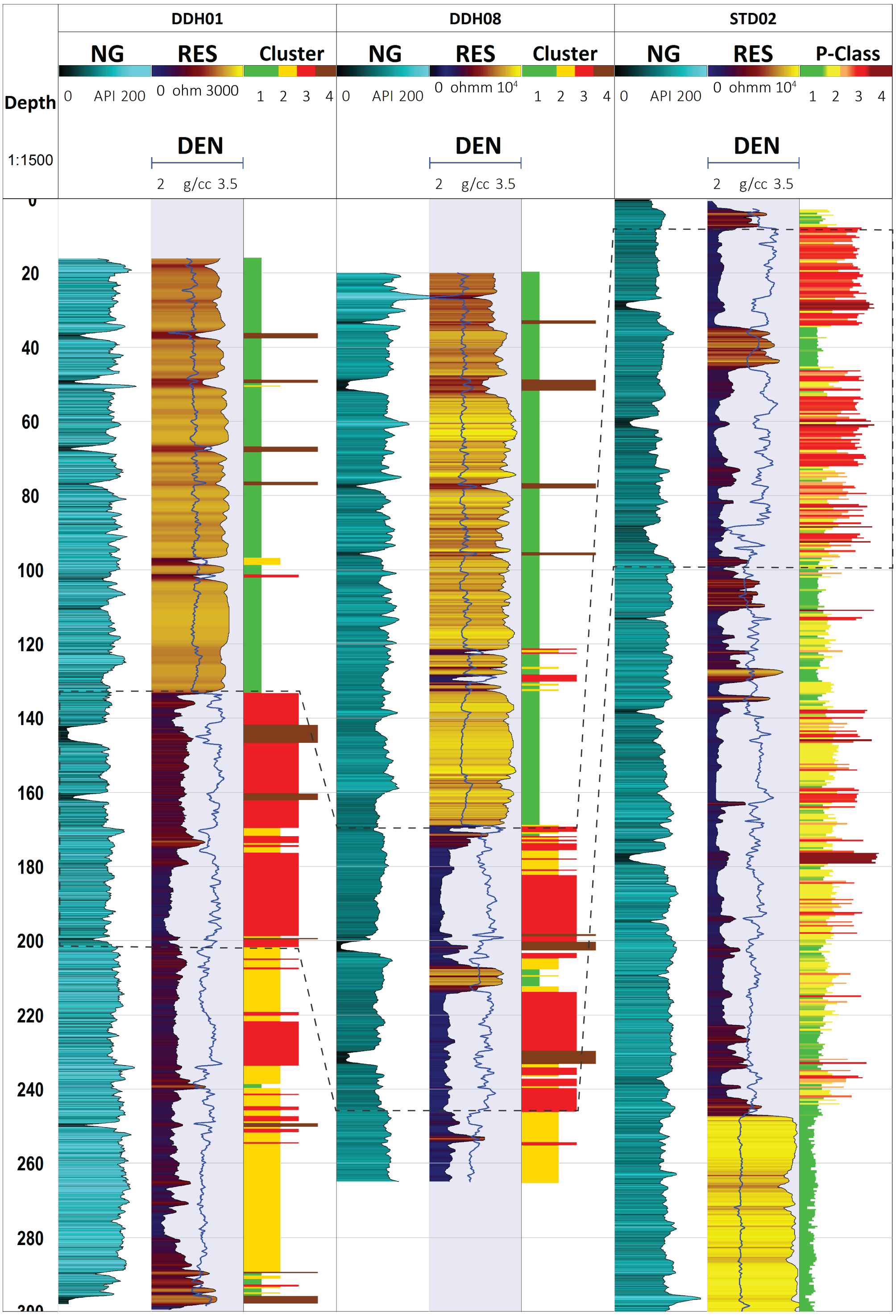


Figure 5.12: Comparison of downhole logs and clustering/prediction results of DDH01 (left), DDH08 (middle) and STD02 (right). Hole to hole correlation is possible from the natural gamma and resistivity logs. Logs in the pseudo-lithology column (Cluster) are from clustering natural gamma, resistivity and density (these logs are shown for reference) for DDH01 and DDH08 and from predicting from these logs in the case of STD02. The inferred lithology logs are in good agreement with each other, taking into account the different dips of the formation and relative positions of the holes. The area bound by the dashed line highlights an example of correlated sections in the three holes.

and pyrite abundance predicted for all 195 intervals. Calculated and predicted pyrite are compared to the S/Fe ratio in Figure 5.13. Pyrite prediction worked reasonably well; the major trends are captured and the correlation between calculated and predicted pyrite is 78%.

Subsequently the algorithm was applied to data from STD02 for pyrite prediction. As mentioned before, since no laboratory assay data is available for STD02 the predicted amount of pyrite could not be properly verified. However, data from the core-scan XRF system system is available for the top 100m. This raw data is available in form of intensities of the respective spectral lines of the measured element and was not calibrated to represent the weight percent of the atom or oxide. Nevertheless, by comparing spectral abundances from STD02 to the laboratory assay data from DDH08, the abundance of the relevant elements (Fe and S) could be scaled and matched reasonably well (Figure 5.14). Pyrrhotite and pyrite abundance was then calculated from the adjusted data. The results are not particularly good due to the uncalibrated nature of the data and subsequent errors in regard to data manipulation.

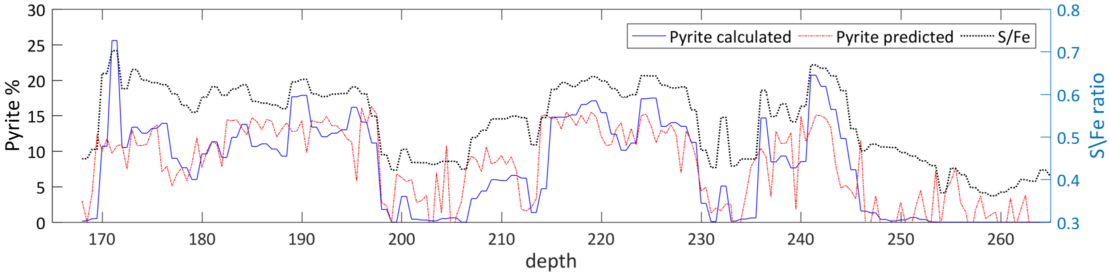


Figure 5.13: Comparison of calculated amount of pyrite, predicted amount of pyrite and S/Fe ratio. The predicted amount matches the general trend but shows greater variation where the assay based calculated amount remains steady. This is likely due to the fact that downhole measurements capture small scale changes of the respective physical property that is 'smeared out' in the assay data due to the small sub-sample that is analysed. The calculated amount of pyrite may also be erroneous because the exact composition of pyrite and pyrrhotite is not known and other sulphide phases may be present that were not included in the calculations.

Figure 5.15 shows iron versus sulphur cross-plots from DDH08 (laboratory assay data) and STD02 indicating the poor correlation of sulphur and iron for the latter data. The relevant data and results of pyrite prediction are shown in Figure 5.16; results for DDH08 on the left and results for STD02 on the right. The plots for DDH08 confirm some of the previously established relationships between geochemical, mineralogical and petrophysical data at Brukunga: sulphur and density are well correlated and so are pyrrhotite content and magnetic susceptibility. Since the uncalibrated sulphur analysis does not match the density very well in STD02, the subsequent mineral calculations are erroneous as illustrated by the poor match of pyrrhotite content and magnetic susceptibility. A very rough visual log of pyrrhotite content from drill core exists and is plotted in Figure 5.16 as a broad reference. This log only indicates four divisions of pyrrhotite content: 0 - no pyrite observed, 1 - some disseminated pyrite present, 2 - abundant pyrite in veins and disseminated and 3 - highly abundant pyrite, some voluminous patches observed. This scheme is highly subjective, simplified and prone to errors. From the three methods to estimate pyrite, visual, calculated and predicted, the amounts predicted from petrophysical data seem the most robust, especially in the absence of assay data.

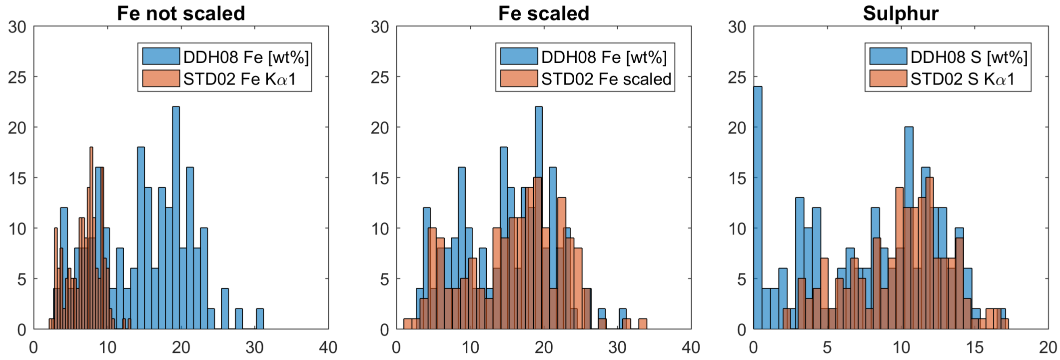


Figure 5.14: Histogram plots comparing assay data to data acquired with the core-scan XRF system. The left plot shows iron data from laboratory assay in wt% compared to the intensity of the Kα1 line from the core-scan XRF system indicating a large difference in the ranges of the respective data. Since both holes intersect similar stratigraphy the core-scan XRF data was scaled to match the range of the laboratory data, shown in the middle plot. The right-hand-side plot compares sulphur content, which shows similar ranges and statistics for both data acquisitions and was therefore not scaled.

5.4.3 Possible links between textural features and p-wave velocity at Brukunga

In addition to distinguishing mineralogy from petrophysical logs, Section 4.1.3 also presented the results of trials to identify different textural features that may also be linked to differences in modal mineralogy. The p-wave velocity, resistivity and natural gamma logs from the unmineralised upper portion of DDH01 were used in this case. These logs are also available for DDH08 and STD02 but no assay data exists for the unmineralised portion of DDH08 and no data at all for STD02; in addition, no objective information on texture is available for either drill hole. Correlating petrophysical data to texture is therefore not feasible in this case, but an interesting feature regarding the sonic logs may give clues about general textural differences between these holes. Figure 5.16 presents the statistics of the p-wave velocities in the form of box-plots and histograms. Vp in DDH01 shows a large range of values, covering almost the entire range from 3.5 to 7.5 km/s, and a bimodal distribution.

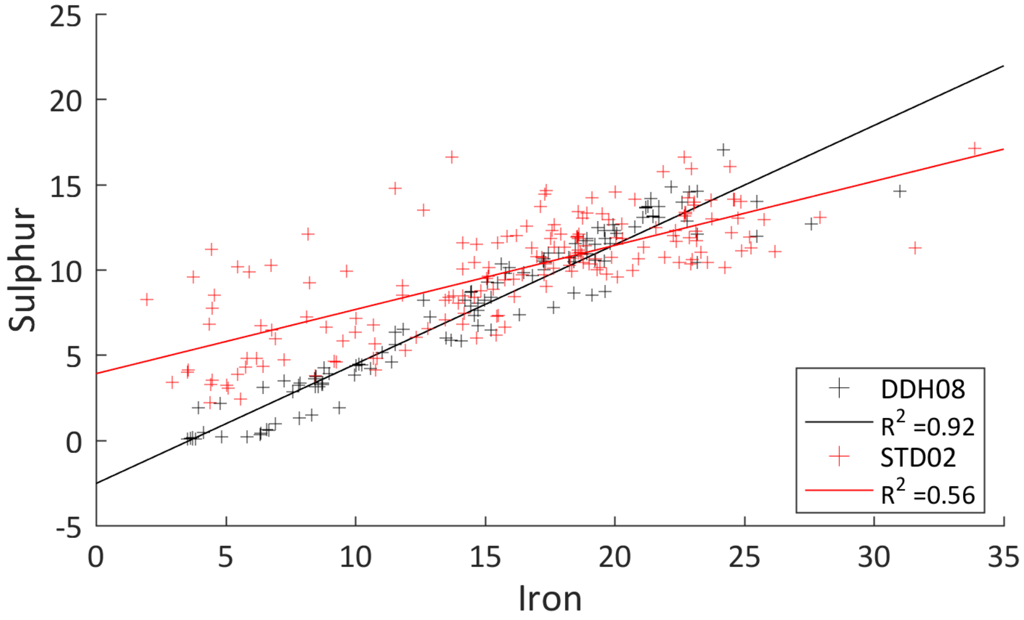


Figure 5.15: Sulphur against iron cross-plot and trend lines indicating a weak correlation of the core-scan XRF data (shifted iron, STD02) and the laboratory analysis (DDH08). The calculated amount of pyrite and pyrrhotite from the uncalibrated sulphur and iron data may not be representative of the true amounts (see Figure 5.16).

As seen in previous sections, high velocities in DDH01 do not only occur in mineralised areas and some of the mineralised parts show low velocities. P-wave velocity is thus influenced by other factors like texture and non-ore mineralogy. The statistics for drill holes DDH08 and STD02 are quite different from DDH01 in that they show a normal distribution and their main group of values fall within narrower ranges with more outliers. This is especially evident for STD02. The range of Vp values on all plots in Figure 5.16 and Figure 5.17 is constant from 3.5 to 7.5 km/s for comparison. Figure 5.17 presents log plots of p-wave velocity from the three drill holes, illustrating the different value ranges. Compared to the obvious highs and lows in DDH01, the log of STD02 looks quite featureless indicating similar velocities throughout different lithologies, textures and mineralogical compositions.

As outlined before, all these three drill holes intersect almost the same stratigraphy within about 1km of each other; but STD02 starts in the mineralised units and intersects the footwall lithologies of the Backstairs Passage formation (Figure 4.2, Figure 5.12) at the end. The only real difference between these drill holes is the angle with which they intersect the formation. DDH01 intersects at a right angle, thus anisotropies from bedding and layering (large-scale texture) are perpendicular to the direction of the hole. DDH08 intersects at angles around 145° and STD02 is drilled down-dip in some places and otherwise intersects at high angles around 160°. The different p-wave velocities may be a direct consequence of these large-scale

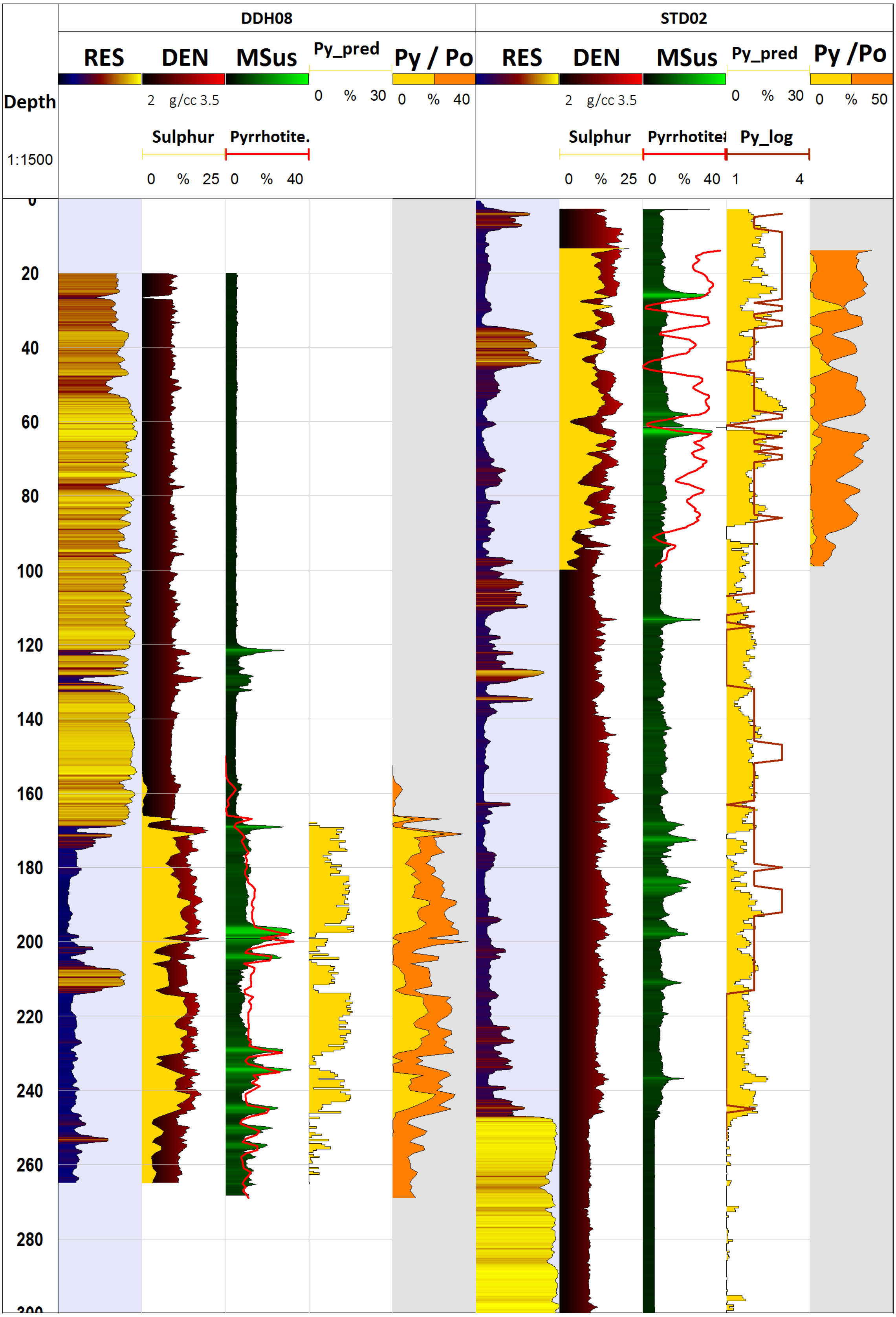


Figure 5.16: Well log plots for drill holes DDH08 (left) and STD02 (right) and results of pyrite prediction. Pyrite (Py_{pred}) was predicted from resistivity (RES), density (DEN) and magnetic susceptibility (MSus) data. (caption continued on next page)

textures whose influence dominate over other factors that affect sonic wave velocity. Anisotropies like bedding and foliation planes or fractures perpendicular to the travelling wave have a slowing effect whilst a wave travelling parallel to them will be faster (Sheriff and Geldart, 1995). Sonic logs can thus give a comparative measure of differences in structural orientation of strata but smaller scale, subordinate features (mineralogy, grain size, etc.) may only be resolved when this stratum is intersected at close to right angles.

Summary

The learnings from simple data clustering were applied to establish a more sophisticated and more importantly, a near-real-time rock mass classification procedure using fuzzy inference systems to predict lithology and mineralogy from petrophysical downhole logs. The natural gamma, density and resistivity logs are sufficient to robustly distinguish between major lithologies of the historic Brukunga sulphide deposit, and may be useful for lithology classification in similar geological environments. The sulphide mineral of economic importance at Brukunga is pyrite, whose abundance was successfully predicted from a combination of magnetic susceptibility, density and natural gamma data. Thus, integrating logging-while-drilling data acquisition with automated data classification offers the means for timely subsurface interpretation, enabling real-time decision making. There may also be some potential in using sonic logs to obtain qualitative and quantitative information about subsurface textural features although the results presented here could not be verified properly.

...(continued) Py/Po are the calculated amounts of pyrite and pyrrhotite from laboratory assay data in the case of DDH08 and from the core-scan XRF system for STD02. The sulphur analysis is shown for reference with the density log; sulphur correlates well with density for DDH08 but not as well for STD02 where it appears to be shifted in some places. The calculated pyrite and pyrrhotite for STD02 may be erroneous, also indicated by the poor correlation of pyrrhotite with the magnetic susceptibility. The predicted pyrite in DDH08 matches very well with the calculated amount and shows reasonable abundances for STD02 where a rough visual log ($\mathrm{Py}_{\mathrm{log}}$) is shown as a broad reference. The visual log only distinguishes between: 0 - no pyrite present, 1 - some disseminated pyrite present, 2 - abundant pyrite and 3 - very abundant pyrite, and is therefore highly subjective, simplified and flawed. From the three methods of estimating pyrite (visual, calculated and predicted), the prediction from petrophysical logs seems to be the most robust method, especially in the absence of reliable assay data.

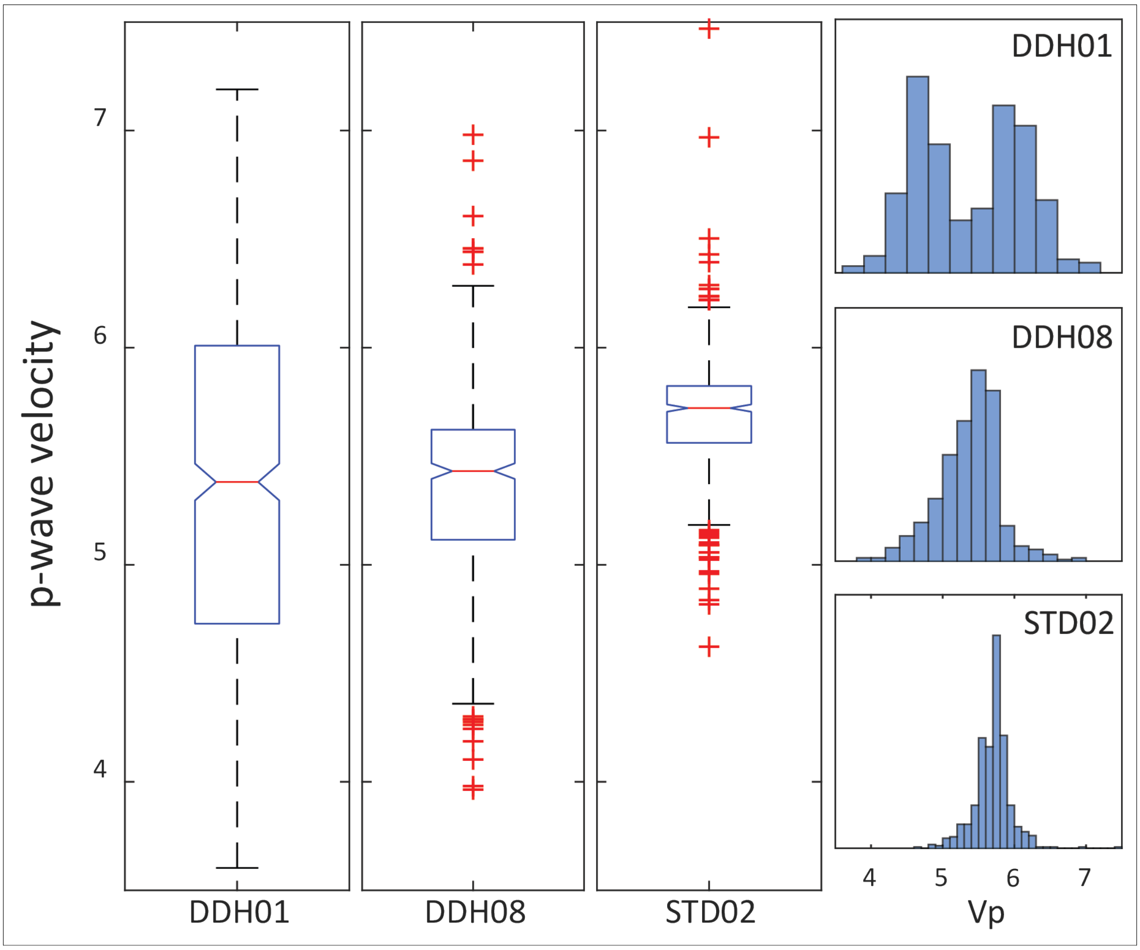


Figure 5.16: Box-plots and histogram plots for comparison of p-wave velocity statistics of the three boreholes. DDH01 shows the largest range of values and a bimodal distribution of the main group of data while data from the other two drill holes show narrower ranges, more outliers and a normal distribution.

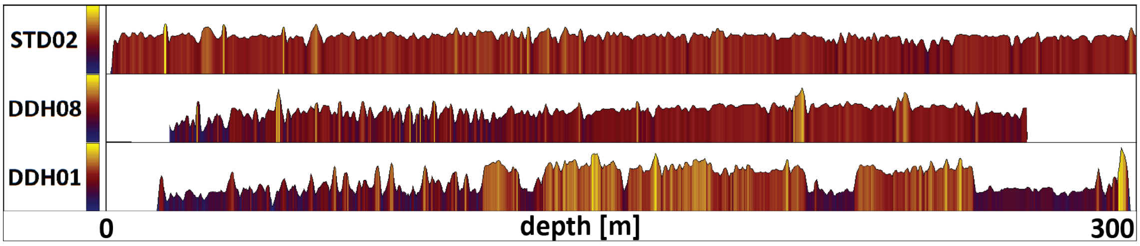


Figure 5.17: Log plots of Vp (range from 3.5–7.5km/s for each plot) of the three drill holes indicating two main groups (low and high) of velocities for DDH01 and generally less variation (range) of the velocities from the other two holes, with STD02 exhibiting a very limited range of values.

5.5 Predicting petrophysical properties and rock mass characteristics from borehole data - a case study from the Kevitsa deposit

The aim of this case study is, firstly to show how sonic velocities can be predicted with other petrophysical borehole data, and secondly to study the link between sonic velocities and rock quality properties. The following subsections are structured similarly to the overall structure of this thesis.

First, the data structure and relationship between the available data variables is investigated to select the most appropriate data for subsequent fuzzy inference modelling. Sonic velocity prediction is tested on different subsets of data with different pre-processing steps applied. The best results are then used to infer rock quality properties. The relationship between sonic velocity and rock quality designation (RQD) is first investigated on measured Vp and RQD data.

In a second step, RQD is inferred from the previously predicted sonic velocities for samples that also have associated RQD measurements for comparison and lastly, RQD is predicted from all available Vp data (from previous predictions). The results identify the necessary and important data for sonic velocity and rock quality prediction and also exemplify how data prediction can complement existing data for more robust 3D interpolation and modelling.

The correlation between sonic velocities and rock quality properties of hard rocks is well known and described in a range of textbooks and studies (e.g., Sjøgren et al. (1979); Dutta (1984); Bieniawski (1989); Altindag and Guney (2005); Barton (2007); Biringen and Davie (2013); Rechlin (2013); Zhang (2016)). These studies also outline some factors that may complicate these relationships, such as rock type, mineral content, grain size and factors associated with increasing depth. Their research indicates, that relationships between RQD and sonic velocities should be established separately for different rock types and that care has to be taken when inferring RQD from velocities of different depth intervals. The latter is based on the fact that at greater depth fractures are closed up and thus do not affect the wave velocities as strongly as when they are open. In addition, lithostatic pressure, water content, state of stress and temperature also affect sonic velocity. The authors also point out that no meaningful correlations could be established for overburden and weathered rock.

This study investigates the relationship of RQD and Vp from the rocks at the Kevitsa site and presents a method to utilise the wealth of data available to achieve different goals. It also highlights the potential errors introduced by visual/manual logging of RQD and the need for a reliable objective measurement for this rock property.

5.5.1 Geological overview and available data

The Kevitsa Ni-Cu-PGE (Nickel-Copper-Platinum Group Elements) deposit lies within the Kevitsa igneous complex, which is part of the Central Lapland Greenstone Belt located in Northern Finland. The deposit is comprised of ultramafic to mafic igneous rocks within sedimentary and volcanic sequences and is characterised by disseminated nickel-sulphide mineralisation. The mineralisation is controlled by magmatic layering, which is cut by steep faults and shear zones with only minor offsets (Gregory et al., 2011). The main rock types at the deposit are olivine pyroxenite and olivine websterite and their altered derivative called ‘meta-peridotite’ by the logging geologists. These lithologies contain variable amounts of clinopyroxene, orthopyroxene, olivine and plagioclase and are therefore best understood as a spectrum of rock having subtle differences (Gregory et al., 2011). Several magmatic pulses comprising these rock types are recognised within the deposit—they usually display relatively sharp olivine rich bases and grade gradually upwards into plagioclase bearing olivine websterite (Gregory et al., 2011). The main sulphide bearing minerals are non-magnetic pyrrhotite, chalcopyrite and pentlandite.

Several ground geophysical surveys were conducted over the area, including 2D and 3D seismic acquisition. Around 700 diamond holes were drilled by different organisations within and outside of the main resource area, totalling in excess of 100km length drilled. After visual logging by a geologist, sections of the core were flagged for geochemical assaying, usually only the mineralised or potentially mineralised parts (Gregory et al., 2011). However, assay data was not used in this study because of the intermittent sampling that would reduce the total amount of data considerably, introducing errors due to poor sampling statistics.

Downhole logging of different petrophysical properties was also carried out, but not consistently throughout the area and not all of the logged holes had the same suite of measurements acquired. While natural gamma, resistivity, IP, magnetic susceptibility and density logging is available for a larger number of the drill holes, sonic logging was conducted in only a few. However, sonic data from boreholes can benefit the processing and inversion of 2D and 3D ground seismic surveys in terms of accuracy and imaging quality.

Additionally, sonic data may be used to infer other rock mass characteristics such as rock strength, state of stress or rock quality parameters (e.g., RQD). A link established between RQD and sonic velocities on a borehole scale can then be integrated over the area of the 3D seismic survey to produce a valuable 3D RQD model.

In the following, sonic velocities are predicted from different combinations of petrophysical downhole log data using fuzzy inference systems. Several combinations are tested to predict Vp and Vs. The aim is to show the viability of the process and to increase the amount of sonic data available to infer rock quality properties.

5.5.2 Data structure and relationship

Sonic velocity prediction was tested for the Kevitsa data by first using data from six drill holes (test data set) for which a complete suite of measurements plus sonic logging is available, to establish which combination of measurements may successfully predict p-wave and s-wave velocity. Since the logged depth resolutions of the different measurements are not consistent, the data is first interpolated in WellCAD using a three-sample moving-average window and exported in 25cm intervals. The upper 10m of data and extreme outliers were excluded from further analyses. A few negative and zero values of the natural gamma and magnetic susceptibility measurements were replaced with values close to the respective detection limit (10^{-3} for natural gamma and 10^{-4} for magnetic susceptibility). Additionally, the natural gamma, magnetic susceptibility and resistivity data was log-transformed prior to analysis.

The data structure and correlation between measured Vp and the other logs for each of the six drill holes is shown in Figure 5.18 along with the relative abundances of the major lithologies for each drill hole. The plots illustrate that the relationship between Vp and other data is not consistent between lithology and drill holes. Holes KV322 and KV323 intersect mainly olivine websterite (UWB) and the p-wave velocity of these holes correlates well with Vs and density, as would be expected. Vp and natural gamma show a weaker, inverse correlation and a relationship between Vp and resistivity or magnetic susceptibility is not apparent. Drill holes KV173 and KV297 intersect mainly olivine pyroxenite (UPX) and the relationship between Vp and the other logs shows some dissimilarities compared to the previous. The remaining drill holes KV171 and KV28 comprise a variety of rock types and the relationship between the measurements is slightly different yet again. Evidently, the subtle differences of the rock types are reflected in their petrophysical properties, thus for a successful prediction of a specific property (e.g., Vp) the training data set should include data from all the major lithologies.

The data from all six holes was combined into one large data set to design and train the fuzzy inference systems. Since the natural gamma shows an inverse relationship to Vp, the inverse was used for subsequent analytical steps, starting again with hierarchical clustering of the variables to visualise their relationship. Figure 5.19

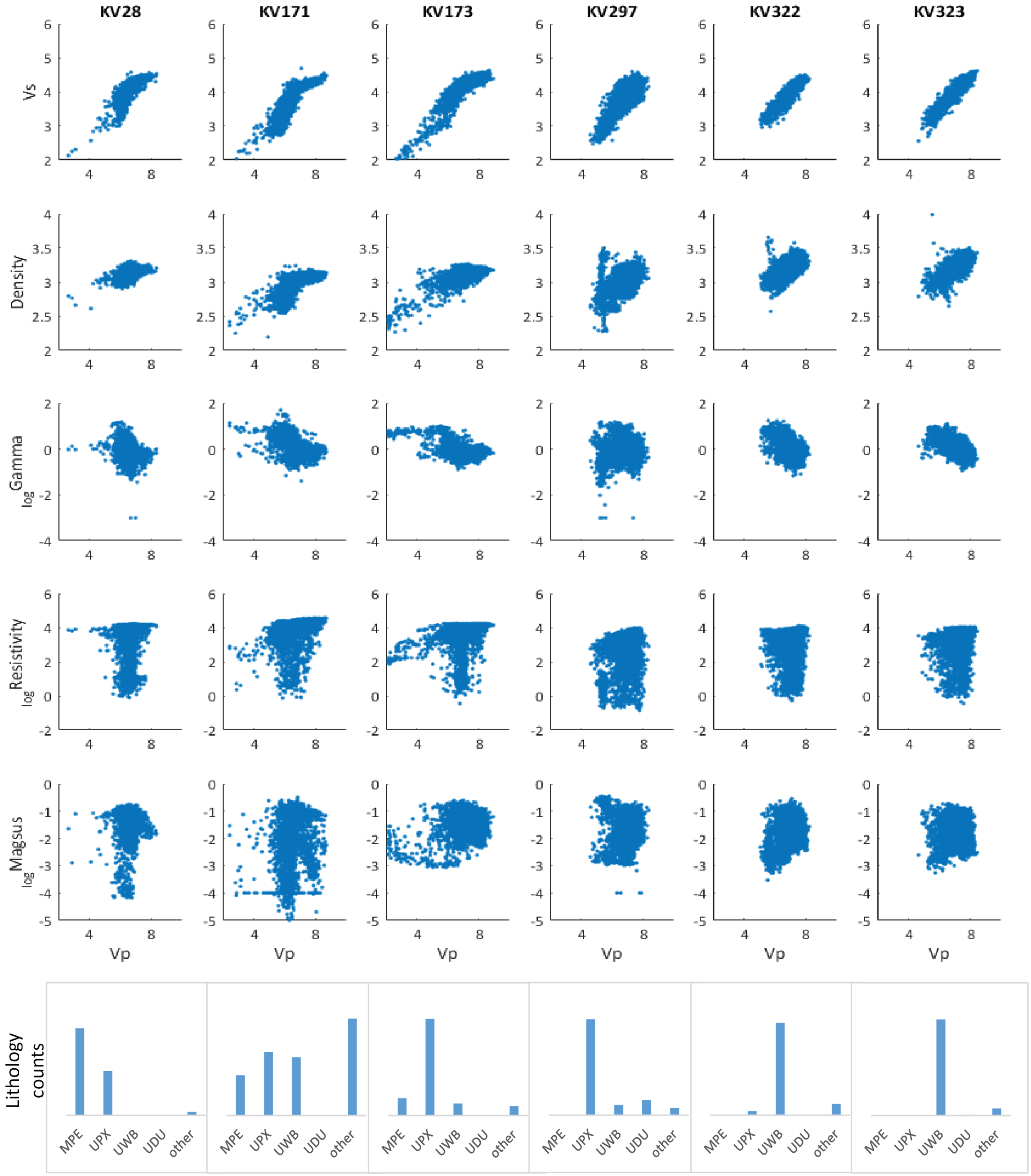


Figure 5.18: Scatter plots of Vp data versus other measurements illustrating their correlation for each of the six drill holes respectively. The bottom row shows the relative abundances of the major logged lithologies for each hole as bar plots. The relationship between the data variables and p-wave velocity seems to be influenced by major lithology because the respective scatter plots show considerable difference for the different drill holes.

shows dendrograms for the z-standardised, m-standardised and normalised data. The data variables are grouped differently, depending on which data transformation is applied, z-standardisation suggests that resistivity is closer related to Vp than the other measurements, while m-standardisation suggests a closer relationship of Vp to natural gamma and magnetic susceptibility. Normalisation leads to Vp being grouped with magnetic susceptibility and resistivity while the close correlation between Vp and density seems to be obscured. The IP data seems to be least related to Vp. These are merely observations and the reason for the apparent changes of relationships are not clear; however, investigating the effect of the different transformation methods allows to choose the most appropriate method. Nonetheless, all three methods were tested in the following inference process.

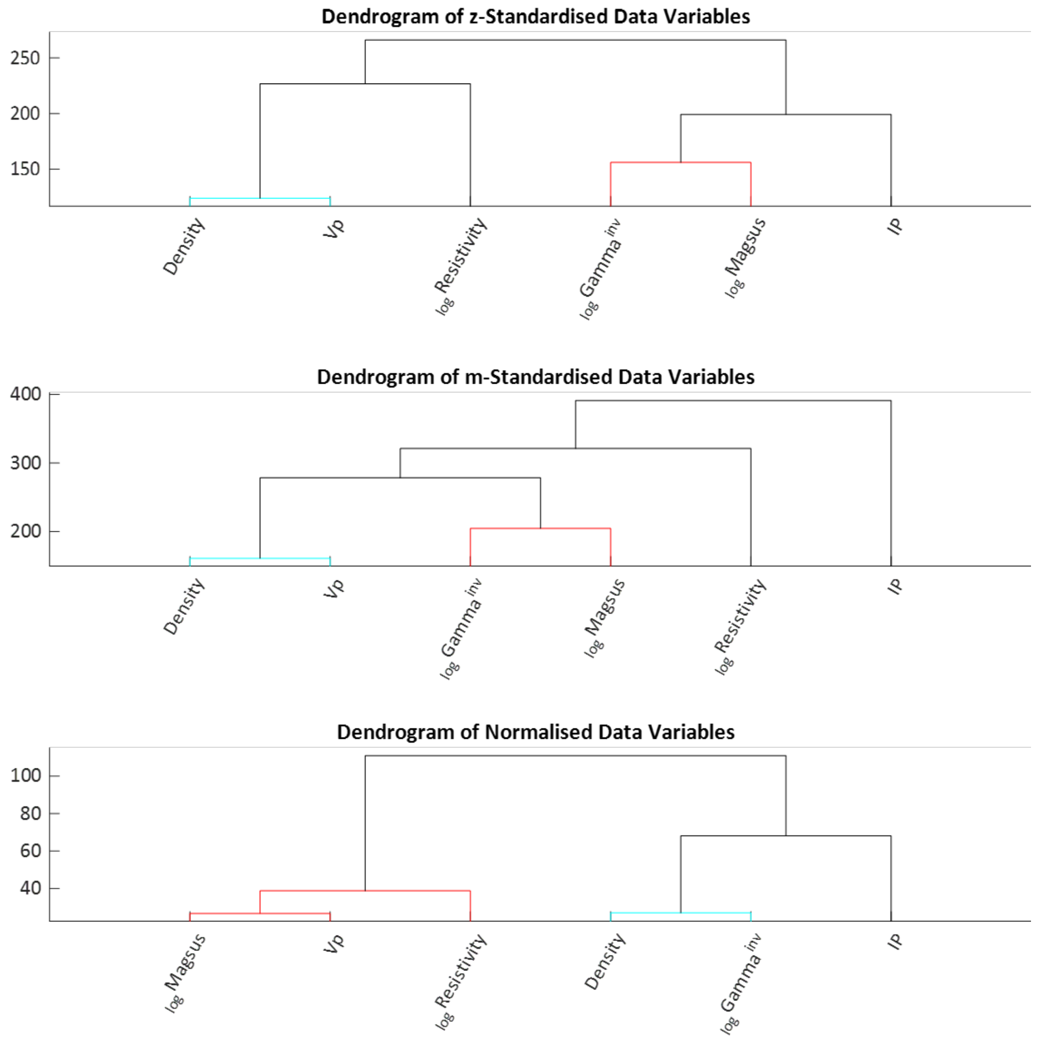


Figure 5.19: Dendrograms from hierarchical clustering of the differently transformed variables illustrate how prior data transformation can influence the grouping of the data. Vp prediction was tested on all three transformation methods with no major differences in the result.

5.5.3 Predicting sonic velocities

Based on the learnings about the data structure, different data variables were chosen as input for fuzzy inference modelling. First, all variables were chosen and the inference process tested on the raw data as well as on the prior transformed variables for two different fuzzy c-means clustering settings (seven clusters and four clusters, using a weighting exponent of $m = 1.6$). Then, three different subsets were tested, namely (1) density, natural gamma and resistivity (DeGaRe); (2) density, natural gamma and magnetic susceptibility (DeGaMs) and (3) natural gamma, magnetic susceptibility and resistivity (GaMsRe). In addition, the process was tested on raw data and transformed data for the subsets.

It is preferable to use raw data for the prediction of petrophysical properties, especially if the trained algorithm will be applied to new data (data not used for training), because the absolute values of the transformed variables depend on the distinct statistics of the raw variable. For example, the z-score is calculated from the mean and standard deviation of the specific input data set. If a subsequent input data set has a different mean and standard deviation, the absolute values of the z-transformed variable is different for identical raw inputs. The shape of the membership functions defined during fuzzy inference modelling are based on the initial statistical properties. Therefore, if the properties of subsequent input variables are different, the absolute values of the transformed variables are also different and these values would be incorrectly classified by the membership functions. However, it is possible to standardise the initial and subsequent data correctly, but only if the transformations are based on identical statistics, which in turn must be based on the range of all possible values. To do that, the range of the expected occurring values for each variable must be known or estimated. This is an unlikely scenario for real-time prediction of petrophysical property values from new input data. As mentioned before, both, the raw and standardised data was used in the present example.

From the total of 22272 samples from the six drill holes, 40% were randomly chosen for training (30 epochs) and checking and the fuzzy inference system with the lowest checking error chosen for Vp prediction on all samples. This process was repeated (looped) 10 times and the results from the best performing FIS are shown in Table 5.1. The performance and results are evaluated in terms of correlation between measured and predicted Vp values (Pearson correlation coefficient). In addition, the root mean square error (in km/s) estimates the magnitude of difference between the measured and predicted values and calculated according to:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{x}_i - x_i)^2}{n}} \, ,$$

> where $\hat{x}_i$ is the predicted value, x_i is the measured value and n is the total number of samples.

The correlation as an indicator for success quantifies the linear fit between measured and predicted values, whilst the RMSE quantifies the error. Since the differences are squared in its calculation, it is strongly influenced by outliers. The RMSE has the same unit as the dependent variable. These measures are sufficient estimators for success for this study since the main goal is to demonstrate the feasibility to predict sonic velocity with other borehole data.

Subsequent s-wave prediction from the same combinations of only the raw input data using seven clusters and a weighting exponent m of 1.6 yielded slightly better results, summarised in Table 5.2 and Figure 5.21.

The results demonstrate the feasibility to infer sonic velocities from other borehole data in general and that the raw data (log-transformed where necessary) can be used without penalty. Whether the predicted sonic velocities are useful estimators of rock-quality properties is investigated in the following. First, sonic velocity is predicted for all samples for which the relevant measurements are available. As mentioned before, there are many drill holes within the Kevitsa deposit area for which different combinations of petrophysical downhole logs exist. Natural gamma, density, resistivity, IP and magnetic susceptibility data is available for 29030 one metre intervals from some 60 drill holes for which no sonic data was acquired. For these intervals, sonic velocities are predicted using the fuzzy inference system that was designed and trained on the test data from the six drill holes.

Table 5.3 shows the statistical properties for (1) the data variables used to design the FIS and (2) for the new data from which sonic velocities are predicted. Although the ranges of most variables are similar for both data sets, the respective statistics for the magnetic susceptibility are markedly different. Including the magnetic susceptibility data for Vp/Vs prediction may result in substantial errors due to these differences between training data and subsequent input data for prediction; thus, the magnetic susceptibility data was excluded from further analysis. The subset containing density, natural gamma, and resistivity data is used to predict Vp and Vs for the new data set. To increase accuracy, a new FIS was designed using all samples from the test data set of the six drill holes (not just 40% as before) and trained for 100 epochs instead of 30. The FIS with the lowest checking error was selected and applied to the new data. Vp was predicted for all 29030 samples including those with shallow depth above 10m, which were omitted previously for FIS design and training. The predicted p-wave velocity values range from ~0.5km/s for just a few samples to about 10km/s; however, the main group of values falls in a sensible range of velocities from

Table 5.1: Compilation of Vp prediction results using different input data combinations, different standardisation methods (pre-processing) and the different number of clusters. The results are generally good and very similar. A higher number of clusters leads to better predictions, whilst omitting the density from the input variables has a negative effect. The correlations coefficient and root mean square error are shown as measures of success and error respectively.

Data set	all	all	all	all	all	DeGaRe	DeGaRe	DeGaMs	DeGaMs	GaReMs
Pre-processing	raw	raw	m-score	normal	z-score	raw	z-score	raw	m-score	raw
No. clusters	4	7	7	7	7	7	7	7	7	7
PCC	0.75	0.78	0.78	0.78	0.77	0.76	0.77	0.75	0.76	0.68
RMSE	0.50	0.47	0.47	0.47	0.48	0.48	0.48	0.49	0.49	0.55

Data: all = density (De), natural gamma (Ga), resistivity (Re), magnetic susceptibility (Ms) and induced polarisation (IP); Pre-processing: applied data transformation, raw = none, m-score = median standard score, z-score = mean standard score, normal = normalised data; No. clusters = number of clusters; PCC = Pearson correlation coefficient; RMSE = relative root mean squared error in km/s.

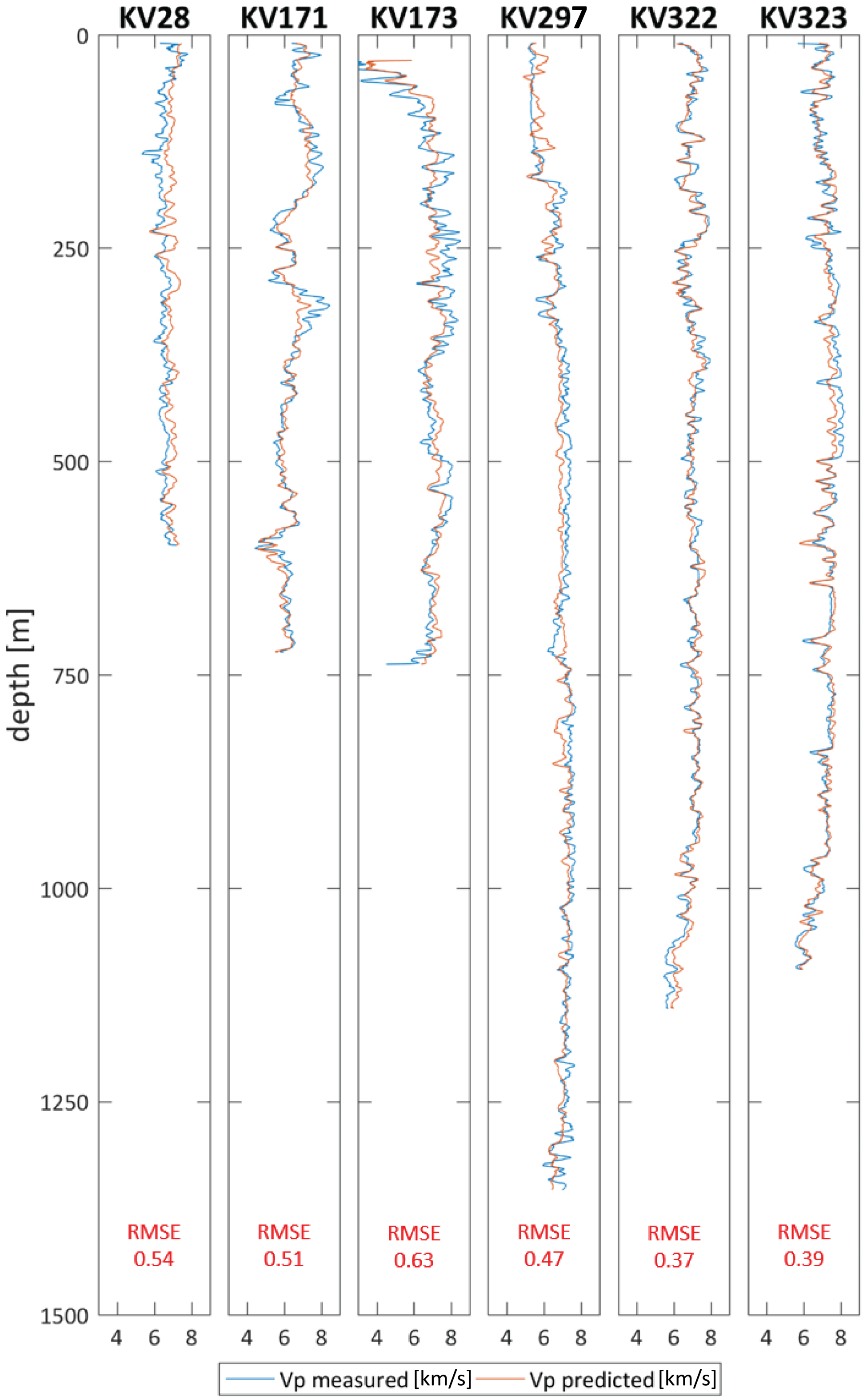


Figure 5.20: Measured versus predicted Vp for the distinct drill holes. Prediction performed best for KV322 and KV323 because the correlation between Vp and other data is strongest for these holes. The data was smoothed using a 20 sample interval window before plotting. The root mean square error in km/s quantifies the difference between measured and predicted data.

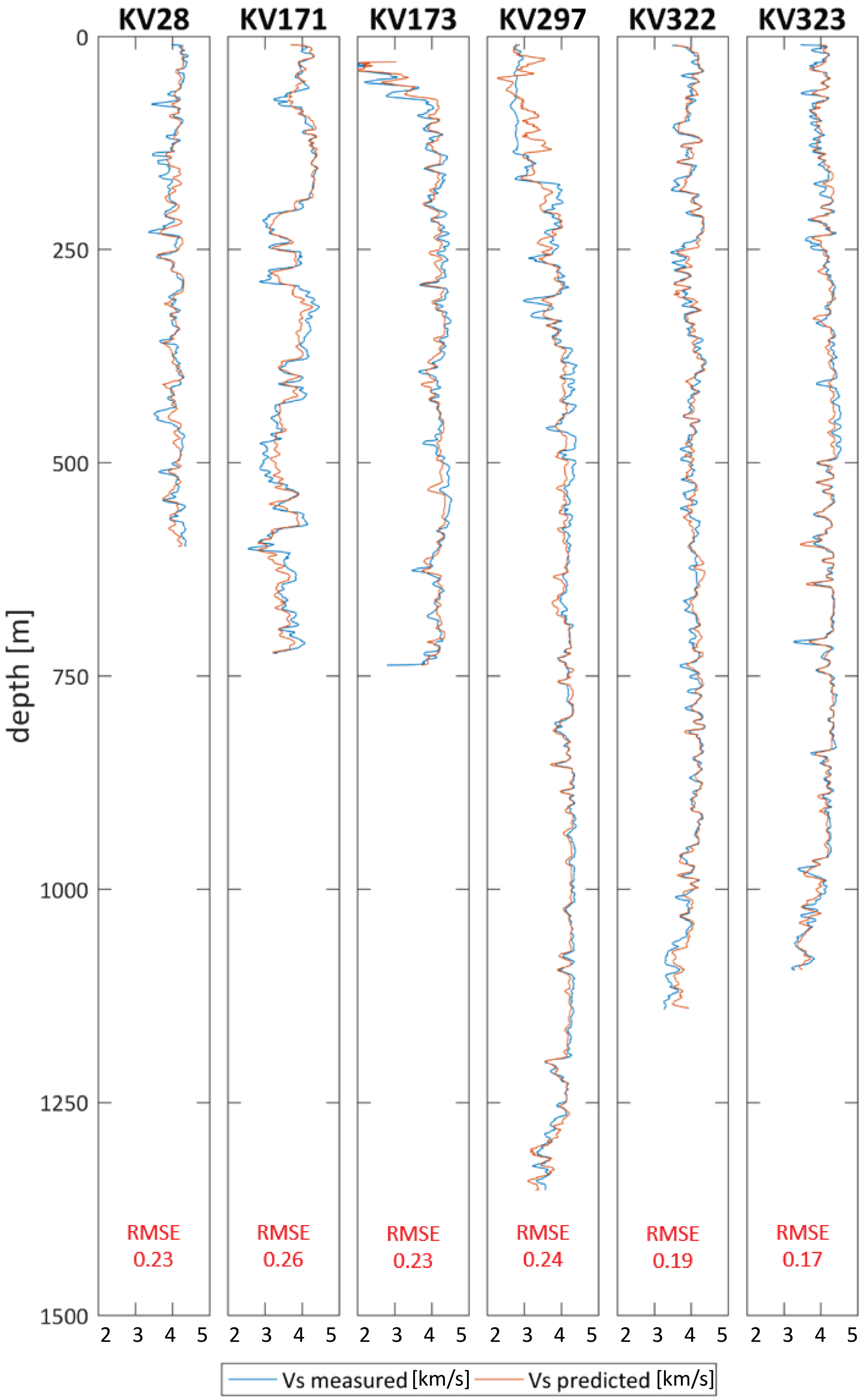


Figure 5.21: Measured versus predicted Vs for the distinct drillholes. The results are slightly better than for predicting Vp. The data was smoothed using a 20 sample interval window for plotting.

Table 5.2: Results of Vs prediction from raw input data using seven clusters and a weighting exponent of 1.6. Vs prediction yielded slightly better results than Vp prediction in terms of correlation between measured and predicted values. Using all available input variable gave the best results.

Data	all	DeGaRe	DeGaMs	GaReMs
PCC	0.85	0.83	0.83	0.77
RMSE	0.22	0.23	0.23	0.27

Data: all = density (De), natural gamma (Ga), resistivity (Re), magnetic susceptibility (Ms) and induced polarisation (IP); PCC = Pearson correlation coefficient; RMSE = relative root mean squared error in km/s.

Table 5.3: Statistics for the different drill hole log data. The top part shows data from which the fuzzy inference system was designed, the bottom part shows the new data from which Vp prediction is attempted. The data statistics of the magnetic susceptibility are markedly different (bold) and this data variable was omitted from Vp and Vs prediction to avoid possible errors.

Statistics for data used to design the fuzzy inference system

	De	${}_{\log}\mathrm{NG}^{\mathrm{inv}}$	IP	${}_{\log}\mathrm{Re}$	${}_{\log}\mathrm{MS}$
Min	2.2	-1.7	0.24	-0.9	**-7.0**
Max	4.0	3.0	100	4.6	**-0.4**
Mean	3.1	0.03	12	3.1	**-1.8**
StD	0.2	0.39	23	1.1	**0.7**

Statistics for the new data from which Vp prediction is attempted

	De	${}_{\log}\mathrm{NG}^{\mathrm{inv}}$	IP	${}_{\log}\mathrm{Re}$	${}_{\log}\mathrm{MS}$
Min	1.1	-2.1	0.07	-0.6	**-1.8**
Max	3.9	2.4	100	6.1	**2.7**
Mean	3.1	0.09	14	3.0	**1.4**
StD	0.1	0.34	22	1.1	**0.5**

StD = Standard deviation, inv = inverse of the data, log = logarithmic transformation applied, De = density, NG = natural gamma, IP = induced polarisation, Re = resistivity, MS = magnetic susceptibility

4 to 8km/s. Only three samples gave erroneous results with velocities below zero and these were omitted from further analysis. S-wave velocity prediction resulted in only four outliers, one below zero and three above 5.5km/s, which were also excluded from further analysis.

Since the predicted velocity for this data cannot be compared to measured values, the relationship to density and the relationship between Vp and Vs may be used as a qualitative measure of success. The scatter plots in Figure 5.22 indicate that the relationship between predicted Vp and Vs and density are similar to the measured data shown in Figure 5.18. In addition, Figure 5.23 shows histograms of (1) the measured Vp (test data from the six drill holes), (2) the predicted Vp (test data from the six drillholes) and (3) predicted sonic velocities from the new data. The plots indicate that the bulk of the predicted values fall into a slightly narrower rage compared to the measured values. This is likely because the measurements used to predict velocity (i.e., density, natural gamma and resistivity) are not sufficient to model the data behaviour correctly. There are many other factors which influence sonic velocities that are not reflected in these measurements and whose influence can therefore not be estimated and incorporated in the fuzzy inference process. However, the general trends and correlations between Vp, Vs and density indicate that reasonably good estimates are obtained. The predicted velocities are used for the subsequent rock-quality estimation.

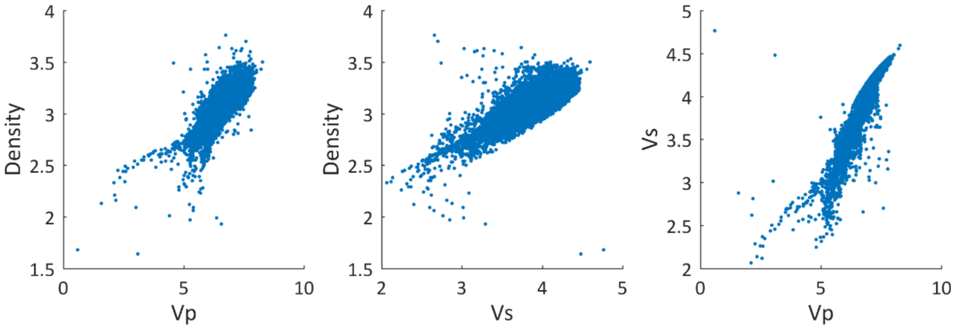


Figure 5.22: Scatter plots of predicted Vp and Vs versus density. The correlation between these variables gives a qualitative measure of the prediction success. The relationships match reasonably well with the data structure in Figure 5.18.

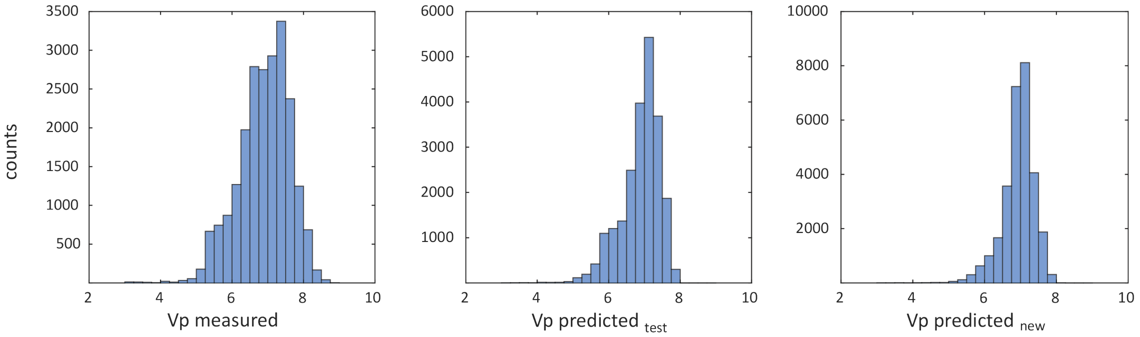


Figure 5.23: Histograms of the (1) measured data from the six test drill holes (left), (2) predicted Vp from the six test drill holes (middle) and (3) predicted Vp from all available data. The data range of the predicted Vp is slightly narrower than the actual data range, indicating that the measurements used for prediction (i.e., density, natural gamma and resistivity) do not sufficiently model the data behaviour.

5.5.4 Inferring mechanical rock mass properties from sonic data

Data acquired from drill core from the Kevitsa resource area includes rock quality designation (RQD) values. The RQD percentage for a given interval is calculated by summing the length of core pieces over 10cm divided by the total interval length, multiplied by 100. Consequently, the rock quality designation gives a measure of fracture intensity and thus rock strength or competence, which are important properties for mine and pit planning. Since sonic velocities are also related to rock quality and strength, the relationship between these measurements may be used to infer rock-quality properties from sonic and associated data. However, RQD is a semi-quantitative measure, which presents a challenge when these values are to be matched to objective quantitative measures like Vp.

Nonetheless, the aforementioned hypothesis was first tested on data from four drill holes for which sonic and RQD measurements are available (3441 samples in total). This test data set is also used to establish and study the relationship between these and other petrophysical properties. The measured RQD % values were logged over variable total interval length, which means that a large number of consecutive 1m intervals can have the same RQD value associated. A bar plot of these different total interval length (Figure 5.24) shows that, for example, one specific interval has a length of over 100m and these 100 samples have the same RQD % value associated to them. Of course, other measurements like the sonic velocities will show a range of different values for these large intervals with identical RQD %, which makes a meaningful correlation between these variables almost impossible. Table 5.4 confirms the weak correlation between RQD % and sonic velocities, inferred parameters (acoustic impedance and bulk modulus) and density.

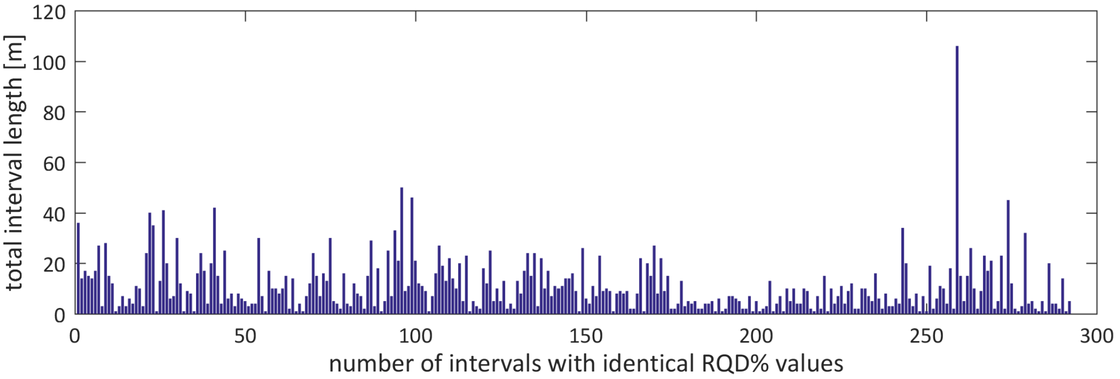


Figure 5.24: The bar plot shows the intervals and their respective length over which RQD % values were acquired. RQD % values are identical for each bar, meaning that they are constant over considerable depth intervals.

Table 5.4: Correlation coefficients for sonic data, inferred parameters (AI = acoustic impedance, K = bulk modulus), density and RQD % show only weak correlations exists between these variables.

	Vs	Vp	Density	AI	K	RQD %
Vs	1					
Vp	0.93	1				
Density	0.77	0.76	1			
AI	0.93	0.97	0.88	1		
K	0.82	0.96	0.76	0.96	1	
RQD %	0.54	0.50	0.47	0.50	0.42	1

In order to study the relationship between RQD and rock type and their distinct properties, the test data set was divided into subsets of major rock types. The left-hand side plot in Figure 5.25 shows the cumulative frequency of RQD % values for the three major lithologies present at Kevitsa (test data set for which measured Vp and RQD are available). The plot on the right shows the same graphs for all available RQD data to confirm that these rock mass characteristics are not biased by sample size. The graphs indicate that meta-peridotite (222 and 1182 samples) represents more fractured, weaker rock, possibly due to its altered nature (see Section 5.5.1 for mineralogical/petrological information) and it constitutes a smaller fraction of the total rock mass. The main rock mass comprises pyroxenite (654 and 3208 samples) and websterite (1865 and 6597 samples) and is more competent and less fractured indicated by the high percentage of high RQD values. It is expected that sonic velocities travel slower through more altered and fractured rock but the relationship can be complicated by the anisotropic nature of fractures and alteration zones (Sheriff and Geldart, 1995) as well as by the depth and other factors such as mineralogy or grain size. Another, and potentially the largest error regarding the relationship between RQD and Vp is a result of the subjective and manual measure of RQD on core samples that no longer represent the in-situ state of the rock.

Figure 5.26 shows that the sonic velocity of the meta-peridotite is significantly lower than that of pyroxenite and websterite. The plot shows mean values of RQD % and sonic data calculated for five distinct ranges (bins) of RQD % values, namely, 0–20%, 20–40%, 40–60%, 60–80% and 80–100%. The calculated mean values and standard deviations for all variables per lithology per RQD % bin is summarised in Table 5.5. A correlation between RQD and sonic velocities is only evident for the meta-peridotite, whereas sonic velocities in pyroxenite and websterite are rather constant for different RQD % values. It is unlikely, that RQD predicted from sonic velocities will sufficiently match these measured values.

Table 5.5: Mean and standard deviation (StD) per 20% bin of RQD % values for the different variables and rock types.

Lithology	Bin RQD %	Vs		Vp		Density		AI		K		RQD %	
		mean	StD	mean	StD	mean	StD	mean	StD	mean	StD	mean	StD
MPE	0 – 20	2.7	0.13	4.6	0.28	3.0	0.06	14	1	33	6	10	0
	20–40	3.4	0.39	5.9	0.46	2.8	0.16	17	2	53	10	26	4
	40–60	3.4	0.36	5.8	0.62	2.9	0.11	17	2	55	16	55	1
	60–80	3.7	0.47	6.3	0.98	3.0	0.10	19	3	65	24	67	8
	80–100	3.8	0.45	6.4	0.71	3.0	0.09	19	3	68	19	92	5
UPXO	0 – 20	3.9	0.4	6.8	0.8	3.13	0.09	21	3	80	21	13	6
	20–40	4.1	0.2	6.8	0.7	3.12	0.09	21	3	77	25	28	5
	40–60	4.0	0.6	7.2	1.2	3.13	0.07	22	4	96	35	56	0
	60–80	4.2	0.1	7.2	0.5	3.17	0.03	23	2	88	20	79	0
	80–100	4.2	0.2	7.3	0.6	3.15	0.07	23	2	94	23	95	4
UWBO	0 – 20	3.9	0.2	6.9	0.4	3.12	0.04	21	1	83	12	10	8
	20–40	3.9	0.2	6.8	0.5	3.18	0.09	22	2	83	17	35	4
	40–60	3.9	0.3	6.7	0.6	3.17	0.12	21	3	78	17	53	5
	60–80	3.9	0.3	6.8	0.5	3.17	0.11	22	2	82	16	72	4
	80-100	4.1	0.2	7.2	0.5	3.22	0.12	23	2	94	19	94	5

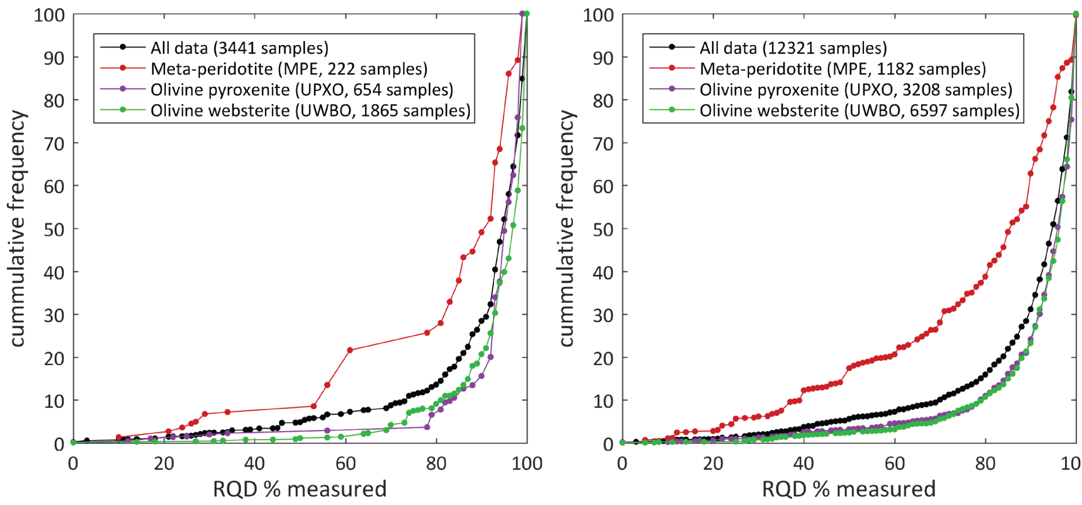


Figure 5.25: Cumulative frequency of RQD % values for the three major lithologies indicating that meta-peridotite represents the weaker rock type compared to the pyroxenites and websterites.

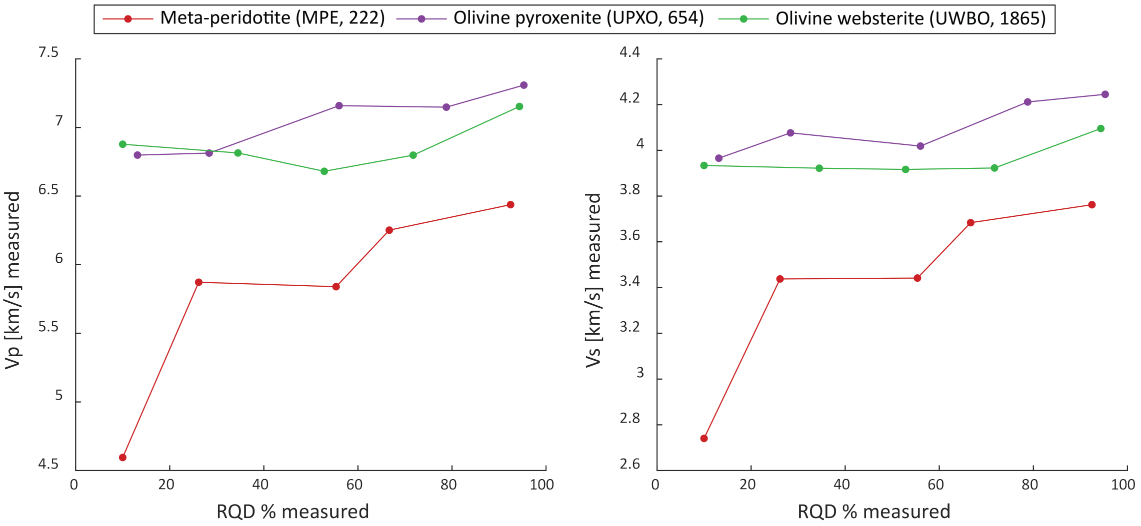


Figure 5.26: RQD % versus sonic velocities for each rock type. The data points are mean values for each 20% bin of RQD. Sonic velocities are, on average, lower for the meta-peridotite.

To infer RQD from sonic data (and related variables) a fuzzy inference system using 11 clusters and a weighting exponent of 1.6 was designed and trained on all available measured data (3441 samples as above). The inference system has five input variables (Vs, Vp, density, acoustic impedance and bulk modulus) and one output variable (RQD %). The FIS was trained on all available samples to maximise success and then applied to the same data for testing. If a clear relationship existed between sonic velocities and measured RQD, the predicted values would show a close match; however, this is not the case due to the bias introduced by RQD logging. The problem is best illustrated in the scatter plots of Figure 5.27. The left plot shows the weak relationship between measured Vp and *measured* RQD; the scatter plot of

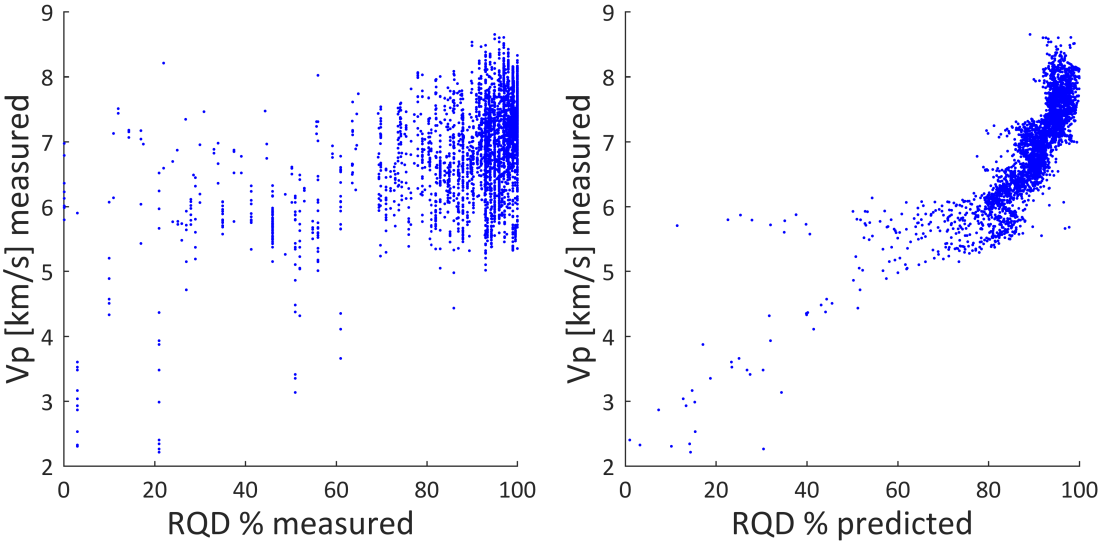


Figure 5.27: Scatter plots of measured Vp versus measured RQD % on the left and of measured Vp versus predicted RQD % on the right. The respective measured variables are weakly correlated, because a large number of identical RQD values are associated with varying p-wave velocities. The correlation between the measured Vp and predicted RQD is stronger, and the predicted values might reflect the true rock quality properties more accurately.

measured Vp versus *predicted* RQD on the right shows a more realistic relationship. It is likely that the predicted RQD values reflect the actual rock-quality properties more accurately than values based on manual logging.

As mentioned before, the test data set is comprised of only a small subset of samples (3441) for which measured sonic and RQD data exists. However, additional RQD data is available for which no sonic data exists (12321 samples). This second data set is used in the following to infer RQD from the previously *predicted* sonic data (Section 5.5.3). Acoustic impedance and bulk modulus were calculated from the predicted sonic data and outliers and negative values (10 samples) omitted from RQD prediction. The trained fuzzy inference system from the test data is applied to the new data, which exemplifies the possible automation of data prediction. Once properly trained, the same FIS can be used to infer the desired variable from subsequent input data.

For better visualisation of the results, the data was again averaged over five 20% bins; one calculation was based on five bins of the *measured* RQD and the second calculation was based on five bins of the *predicted* RQD. The resulting mean values differ considerably, as illustrated in Figure 5.28, because of the previously discussed nature of the measured RQD data. The left plot shows the mean values for the measured RQD (used for binning) versus predicted Vp (from which RQD was subsequently predicted). The graphs are similar to Figure 5.26 where the relationship of the measured data is shown, but the difference in Vp between the respective rock types is decreased. This is a result of the smaller range of the predicted data that smoothed out some of the variability in Vp values (Figure 5.23), but it is also a result of the mean-calculation because the measured RQD values are so poorly correlated to Vp that the mean value of Vp for different RQD ranges is quite similar (see Figure 5.27 for comparison). The plot on the right shows the relationship of the predicted Vp and predicted RQD where the bins were based on the predicted RQD. The graphs indicate a stronger correlation between the respective mean values of these variables, but also, that no RQD values below 20% were predicted. This is again a result of the smoothed out predicted sonic velocities, but also a result of an insufficient sample size used for training. If the training data does not include enough samples spanning the entire range of possible input to output values, the inference process will yield only limited results. It is of course also possible that there are no intervals of low rock quality present in this data set.

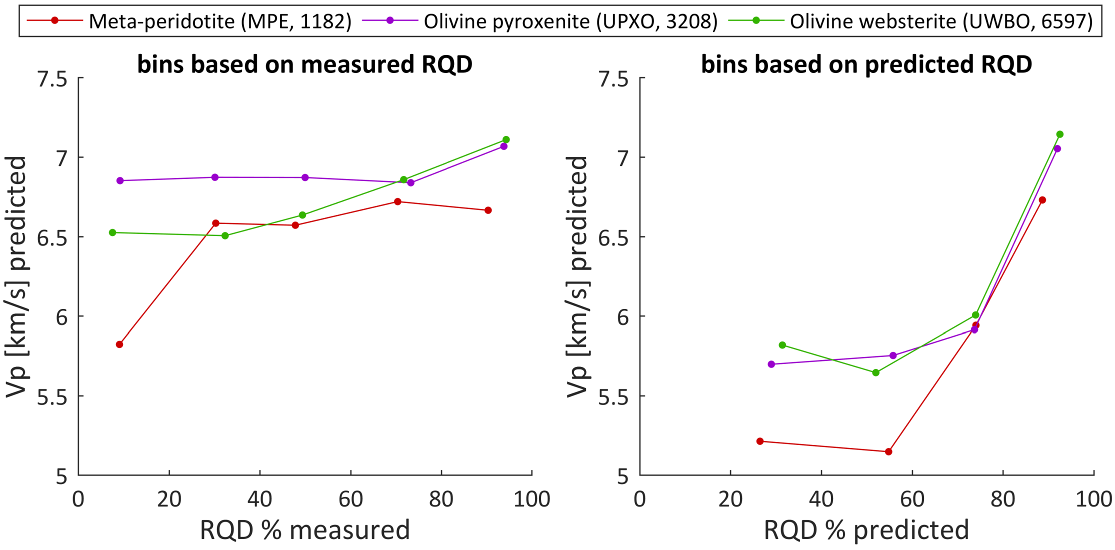


Figure 5.28: Average values of predicted Vp and measured RQD (left) and predicted Vp and predicted RQD (right). The bins for the left plot are based on the measured RQD, whilst the bins for the right plot are based on the predicted RQD values. The correlation between RQD and Vp is more pronounced and possibly reflects the 'true' relationship between these variables better than the measured RQD. The difference of Vp between the distinct rock types is lower than for the measured variables shown in Figure 5.26, possibly due to the narrower range of predicted Vp values.

In a last step, RQD was predicted from all previously predicted sonic data (29020 samples), even though there is no measured RQD data available for comparison. The predicted RQD is plotted against Vp and bulk modulus in Figure 5.29. The graphs show predicted RQD % values over the entire range, suggesting that (contrary to the previous assumptions) the training data was sufficient to model the data behaviour.

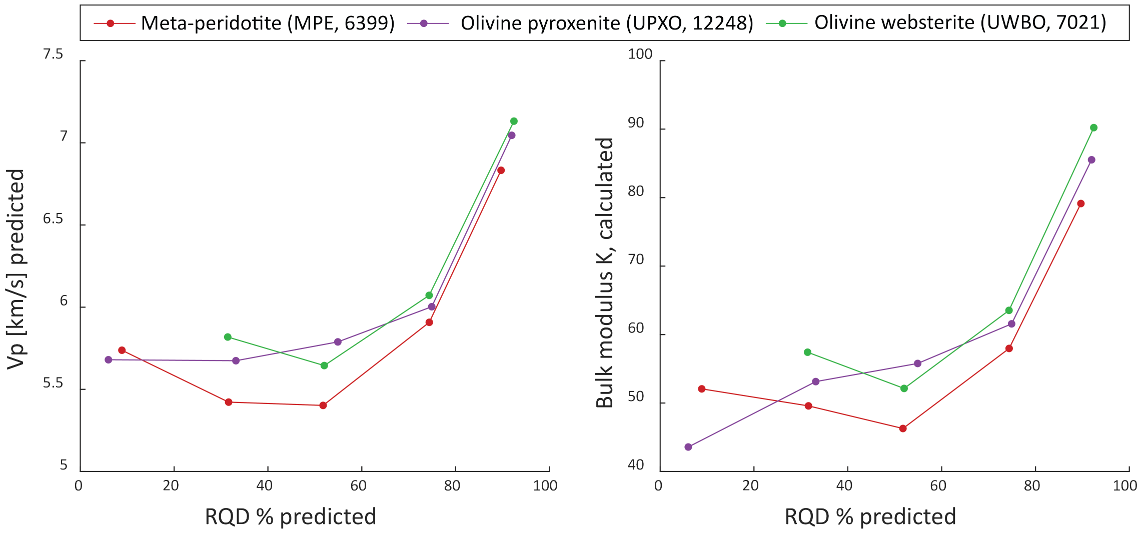


Figure 5.29: Predicted RQD % versus Vp and bulk modulus (K). No measured RQD is available for comparison. RQD values are predicted for the entire possible range and the relationship to Vp is similar to the previous data in Figure 5.28

Summary

Predicting sonic velocities with other drill hole data performed reasonably well when the density log is included. Although the range of the predicted values is slightly narrower, the information they provide can be invaluable for a range of applications. Firstly, the data coverage is increased considerably, enhancing the robustness of the starting model for seismic processing and subsequent interpretation and inversion. The information is obtained at no extra cost, utilising the already available data to its full potential. Secondly, as demonstrated in this study, sonic velocity can highlight potentially weak rock, thus adding important and invaluable information to mine and pit planning. The relationship between rock quality designation and sonic velocity can be established from a small data set and for the case of the Kevitsa example, applied to the entire 3D seismic area as a basis for a 3D rock-quality model. However, the nature of the current RQD acquisition (manual logging) introduces uncertainty when establishing the relationship with sonic data. For example, counted fractures or length of intact core pieces may be erroneous due to fracturing related to the drilling process and core recovery itself and the logged values for RQD may be constant over a considerable length of core.

Seismic velocities are generally lower in highly fractured or porous rocks and higher in massive, intact rock of the same lithology. It is therefore reasonable to utilise sonic data for rock quality estimation, but it may be prudent to try to establish the relationships from objective measures like, for example, acoustic tele-viewer data. The images obtained from such acoustic logging of the borehole wall can be interpreted objectively with image analysing techniques and correlated to sonic data from the same borehole with high depth fidelity and resolution. Unfortunately, this could not be investigated since no such data was available but it may be an interesting study for future work in this area.

Chapter 6

CONCLUSIONS

One of the aims of this study was to identify combinations of data variables that are useful to classify certain aspects of rock mass features or lithology in general. The aim was to find links between petrophysical measurements and rock mass features that can be exploited for lithological characterisation. At the same time, the data analytical process should be automated to enable a near-real time data acquisition and interpretation strategy that ultimately enables more time and cost-effective exploration and mining workflows. Basic statistical analysis of high-dimensional data sets provides clues as to whether and what variables are characteristic of certain rock properties. An important learning regarding the selection of the 'right' data for classification purposes is that simple statistical tools like correlation coefficients or dendrogram analysis can provide valuable insight and incorporating these as analysis tools as a prior step increases classification success considerably. Furthermore, examining variable dendrograms presents an intuitive means to reduce the dimensionality of multivariate data because similar variables are grouped together and selecting just one variable per group reduces complexity while preserving the overall characteristics of the data set.

A novelty of this study is that geochemical assay data was analysed in conjunction with petrophysical data to gain better insights into their interrelation and dependence. Traditionally, each data set is interpreted by experts of one discipline or the other. The hypothesis that two streams of data that are complimentary should work better than each on its own was tested and found to work well. Whilst intuitive, it has not been widely tested and it was found that while using geochemistry alone works quite well adding petrophysical data provides greater robustness and provides further insights and abilities. Analysing geochemical and petrophysical data from sulphide bearing metasedimentary rocks showed that a range of data combinations may be used as successful rock mass classifiers, the ‘best’ combination depends on the classification objective. Pyrite/pyrrhotite mineralised sedimentary sulphide deposits within barren strata, cross-cut by mafic dykes for example, can be classified from a combination of natural gamma, resistivity and density data. This combination may not work as well for other sulphide deposits; however, the traits exploited are distinctive for these minerals and so it is expected that this combination will work more often than not.

It has also been shown that by combining petrophysical measurements, differences in mineral content and texture can be identified. The measurement with the largest range of values is not necessarily the most useful. For example, pyrite and pyrrhotite have markedly different p-wave velocities, but comparatively smaller differences in density. However, a combination of natural gamma, resistivity and density separates these minerals better than natural gamma, resistivity and sonic p-wave velocity (Vp). On the other hand, Vp in combination with either the resistivity or natural gamma logs separates coarse-grained quartz-rich and fine-grained mica-rich, sections of barren meta-sedimentary rocks. This relationship is reliant to some extent on the angle at which the borehole intersects the bedding. If a bedded (anisotropic) strata is intersected orthogonally, the range of values is high and Vp may be indicative of modal mineralogy and small-scale textural features like grain size and foliation. At higher angles these features have a subordinate influence, the overall range of values is lower and Vp is more influenced by the angle between the travelling wave and the bedding. Thus, Vp may be used to qualitatively assess formation dips in areas of consistent lithology.

This study has demonstrated that cluster analysis is a data-driven process that divides a data set, or conversely groups samples based on their specific features. Consequently, clustering of geoscientific data results in a ‘natural’ classification depending on the inherent characteristics of the input data. The separation of data into arbitrary groups, such as ore-waste delineation based on specific cut-off values set by other factors (metal prices and or labour costs), is rarely achievable through a clustering process. In these cases, appropriately trained neuro-adaptive learning algorithms provide a means to improve classification success. For example, the cut-off grades applied in iron ore mining are not representative of the natural characteristics of the rock. When the exact same data variables (assay data of iron or aluminium) used to determine cut-offs are clustered the results proved unsatisfactory as predicted with the example iron data set used for testing. However, lithology prediction through fuzzy inference systems modelling increased the classification success to almost 100% accuracy for assay input data. So for the task of ore-waste separation the use of fuzzy inference systems is a good approach to both automatically classify the lithology and to assist in the ore-waste selection.

The substitution of petrophysical measurements for assay data was also demonstrated on iron ore data. Firstly, I demonstrated the successful replacement of the natural gamma log for the aluminium assay, however substituting for iron content proved more difficult. The density, although an intuitive choice for the measurement for iron grade, was an ineffective substitution because both low and high iron grades may exhibit high density values. This is due to the comparatively high density of uneconomic magnetite bearing BIF versus the lower density of very friable (porous) fine grained iron ores. However, spectral gamma-gamma logging for density extraction offers a means of obtaining information about the rock's average atomic number, which is heavily influenced by the iron content. The spectral gamma-gamma ratio is a measure of grade in iron ore formations due to the relatively high amount of iron present compared to other heavy elements in this environment, and could successfully be substituted for iron assay data. Therefore, it is possible to robustly classify lithology and estimate grade of Pilbara iron ore deposits from data collected in a single tool-stack with natural gamma and spectral gamma-gamma probes (with the right tool both measurements can be derived from spectral analysis of the gamma spectrum).

The prediction of sonic velocities from other borehole data yielded very encouraging results, with a high degree of fit between predicted and measured sonic data. Although the predicted velocities were somewhat smoothed and averaged out, this may actually be beneficial in circumstances where 'noisy' or 'spiky' data might complicate the velocity model and data processing. The application of this method will allow complex 3D models of seismic velocity to be built when a wealth of other data exists, but sonic logging is sparse. Furthermore, such 3D models of sonic velocity may in future be used by mining engineers to design mines and in ore-extraction planning because of the common relationship between sonic velocity and elastic modulus with engineering parameters, such as unconfined compressive strength (UCS) and rock quality designation (RQD). The relationship between predicted RQD and measured sonic velocity is more reasonable than the correlation between measured RQD and measured sonic velocities due to the biased nature of manual RQD acquisition. Further work and objective measures of rock quality indicators are needed to better establish these relationships. However, the presented example of data from Kevitsa demonstrates that the link between rock quality and sonic velocity can be established from a small initial subset of data, and the learnings applied to predict RQD from subsequent predicted velocity data. Now it becomes possible to model RQD over the entire seismic 3D area to highlight areas where there are possibly weak rocks or mine risks that need further investigation. The predicted sonic data generated from an extensive network of other borehole petrophysical data should provide an example of an effective volumetric dataset for de-risking mining.

Finally, the learnings of this research may be used as a guide on how to efficiently organise future exploration campaigns and how to use the acquired data to its full potential. A possible workflow is presented here:

> First steps in an exploration campaign to follow up initial encouraging results should start with an analysis of historic data to establish relationships between the rock mass, petrophysical measurements and chemical analysis. If no such data exists, the first drill holes could be designed to intersect a wide variety of the various rock types and mineralisation or alteration zones present in the area.
>
> The learnings from this initial data analysis can then be used to identify the essential measurements to be collected from all subsequent drilling for the desired rock mass classification.
>
> Once the relationships between the measurements and rock mass properties of interest are defined, the classification process can be automated using the adaptive learning algorithms presented in this thesis.
>
> If combined with real-time data acquisition such as logging-while-drilling technologies, this process enables real-time decision making and efficient exploration campaigns.
>
> At later stages of mine development the comprehensive data set can be analysed for links to rock quality and ore mineral distribution to inform mine and pit planning and geo-metallurgical work.

References

Altindag, R., and A. Guney, 2005, Evaluation of the relationships between P-wave velocity (Vp) and joint density (J): 19th International Mining Congress of Turkey, Izmir, 101–106.

An, P., 1991, Application of fuzzy set theory to integrated mineral exploration: Can. J. Explor. Geophys., **27**, 1–11.

Bagheri, A., B. Biranvand, S. Rezazadeh, M. Fasih, and H. Bahtiari, 2005, Integrated analysis of core and log data to determine reservoir rock types and extrapolation to uncored wells in a heterogeneous clastic and carbonate reservoir: International Symposium of the Society of Core Analysts, Toronto, Canada, 21–25.

Barton, N., 2007, Rock quality, seismic velocity, attenuation and anisotropy: CRC press.

Bastrakov, E. N., R. G. Skirrow, and G. J. Davidson, 2007, Fluid evolution and origins of iron oxide Cu-Au prospects in the Olympic Dam district, Gawler Craton, South Australia: Economic Geology, **102**, 1415–1440.

Belperio, A., R. Flint, and H. Freeman, 2007, Prominent hill: A hematite-dominated, iron oxide copper-gold system: Economic Geology, **102**, 1499–1510.

Bezdek, J. C., R. Ehrlich, and W. Full, 1984, Fcm: The fuzzy c-means clustering algorithm: Computers & Geosciences, **10**, 191–203.

Biçici, E., and D. Yuret, 2007, Locally scaled density based clustering: International Conference on Adaptive and Natural Computing Algorithms, Springer, 739–748.

Bieniawski, Z. T., 1989, Engineering rock mass classifications: a complete manual for engineers and geologists in mining, civil, and petroleum engineering: John Wiley & Sons.

Biringen, E., and J. Davie, 2013, Correlation between Vs and RQD for different rock types: Presented at the International Conference on Case Histories in Geotechnical Engineering, Missouri University of Science and Technology.

Bonnici, N. K., 2012, The mineralogical and textural characteristics of copper-gold deposits related to mineral processing attributes: PhD thesis, University of Tasmania.

Bosch, D., J. Ledo, and P. Queralt, 2013, Fuzzy logic determination of lithologies from well log data: application to the KTB project data set (Germany): Surveys in Geophysics, **34**, 413–439.

Conaway, J., and P. Killeen, 1978, Quantitative uranium determinations from gamma-ray logs by application of digital time series analysis: Geophysics, **43**, 1204–1221.

Czubek, J., 1979, Modern trends in mining geophysics and well logging for mineral exploration, Paper No. 11: Presented at the Geophysics and geochemistry in the search for metallic ores : proceedings of Exploration 77 an international symposium held in Ottawa, Canada in October, 1977, Ottawa : Geological Survey of Canada ; [Hull, Quebec : available from Canadian Govt. Pub. Centre].

Dekkers, M. J., D. Heslop, E. Herrero-Bervera, G. Acton, and D. Krasa, 2014, Insights into magmatic processes and hydrothermal alteration of in-situ super-fast spreading ocean crust at ODP/IODP site 1256 from a cluster analysis of rock magnetic properties: Geochemistry, Geophysics, Geosystems, **15**, 3430–3447.

Dubois, M. K., G. C. Bohling, and S. Chakrabarti, 2007, Comparison of four approaches to a rock facies classification problem: Computers & Geosciences, **33**, 599–617.

Dutta, N., 1984, Seismic refraction method to study the foundation rock of a dam: Geophysical Prospecting, **32**, 1103–1110.

Elkington, P., P. Stouthamer, and J. Brown, 1982, Rock strength predictions from wireline logs: International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, Elsevier, 91–97.

Esbensen, K., L. Lindqvist, I. Lundholm, D. Nisca, and S. Wold, 1987, Multivariate modelling of geochemical and geophysical exploration data: Chemometrics and intelligent laboratory systems, **2**, 161–175.

Fullagar, P. K., B. Zhou, and G. N. Fallon, 1999, Automated interpretation of geophysical borehole logs for orebody delineation and grade estimation: Mineral Resources Engineering, **8**, 269–284.

George, R., 1969a, Sulphide-silicate reactions during metamorphism of the Nairne pyrite deposit: Proc. Aust. Inst. Min. Metall, 1–7.

———, 1969b, Sulphide vein formation during metamorphism of the Nairne pyrite deposit: Proc. Aust. Inst. Min. Metall, 9–17.

George, R. J., 1967, Metamorphism of the Nairne pyrite deposit: PhD thesis, University of Adelaide.

Gregory, J., N. Journet, G. White, and M. Lappalainen, 2011, Kevitsa copper nickel project Finland: technical report for the mineral resources and reserves of the Kevitsa project: technical report, First Quantum Minerals Ltd.(52 pp.).

Gum, J., 1998, The sedimentology, sequence stratigraphy and mineralisation of the Silverton Subgroup, South Australia: PhD thesis, University of South Australia.

Hachmöller, B., and H. Paasche, 2013, Integration of surface-based tomographic models for zonation and multimodel guided extrapolation of sparsely known petrophysical parameters: Geophysics, **78**, EN43–EN53.

Hanesch, M., R. Scholger, and M. Dekkers, 2001, The application of fuzzy c-means cluster analysis and non-linear mapping to a soil data set for the detection of polluted sites: Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, **26**, 885–891.

Hillis, R. R., D. Giles, S. E. van der Wielen, A. Baensch, J. C. Cleverly, A. Fabris, S. W. Halley, B. D. Harris, S. M. Hill, P. A. Kanck, A. Kepic, S. P. Soe, G. Stewart, and Y. Uvarova, 2014, Coiled tube drilling and real-time sensing - enabling prospective drilling in the 21st century: Society of Economic Geologists Special Publication, **18**, 243–259.

Ilkhchi, A. K., M. Rezaee, and S. A. Moallemi, 2006, A fuzzy logic approach for estimation of permeability and rock type from conventional well log data: an example from the Kangan reservoir in the Iran Offshore Gas Field: Journal of Geophysics and Engineering, **3**, 356.

Imamura, S., 1994, Integrated interpretation of exploration data in geotechnical engineering: an approach using fuzzy theory, *in* SEG Technical Program Expanded Abstracts 1994: Society of Exploration Geophysicists, 202–205.

Jago, J. B., J. Gum, A. Burtt, and P. W. Haines, 2003, Stratigraphy of the Kanmantoo Group: a critical element of the Adelaide Fold Belt and the Palaeo-Pacific plate margin, Eastern Gondwana: Australian Journal of Earth Sciences, **50**, 343–363.

Johnson, S., 1991, Natural Radiation: Virginia Division of Mineral Resources, **37**.

Kaufman, L., and P. J. Rousseeuw, 2009, Finding groups in data: an introduction to cluster analysis: John Wiley & Sons.

Killeen, P., and C. Mwenifumbo, 1988, Downhole assaying in Canadian mineral deposits with the spectral gamma-gamma method: Report, International Atomic Energy Agency, IAEA.

Killeen, P., and L. Schock, 1991, Borehole assaying with the spectral gamma-gamma method: some parameters affecting the SGG ratio: Proc. 4th Internat. MGLS/KEGS Symp. on Borehole Geophysics for Minerals, Geotechnical and Groundwater Applications, 399–406.

King, A., P. Fullagar, and Y. Lamontagne, 1994, Borehole geophysics in exploration, development, and production: Presented at the Can. Inst. Min. Metall. Field Conference, Sudbury, Ontario.

LaGanza, R. F., 1959, Pyrite investigations at Nairne, South Australia: Economic Geology, **54**, 895–902.

Lapointe, P., W. Morris, and K. Harding, 1986, Interpretation of magnetic susceptibility: a new approach to geophysical evaluation of the degree of rock alteration: Canadian Journal of Earth Sciences, **23**, 393–401.

Latham, A., K. Harding, P. Lapointe, W. Morris, and S. Balch, 1989, On the lognormal distribution of oxides in igneous rocks, using magnetic susceptibility as a proxy for oxide mineral concentration: Geophysical Journal International, **96**, 179–184.

Lawrence, R., 1980, Stratigraphy and structure in and adjacent to the Talisker Formation (Nairne Pyrite equivalent) in the eastern Mount Lofty Ranges: PhD thesis, University of Adelaide.

Lindqvist, L., I. Lundholm, D. Nisca, K. Esbensen, and S. Wold, 1987, Multivariate geochemical modelling and integration with petrophysical data: Journal of Geochemical Exploration, **29**, 279–294.

Lu, Y., T. Ma, C. Yin, X. Xie, W. Tian, and S. Zhong, 2013, Implementation of the fuzzy c-means clustering algorithm in meteorological data: International Journal of Database Theory and Application, **6**, 1–18.

Lyons, N., 2012, Evidence for magmatic hydrothermal mineralisation at Kanmantoo Copper deposit, South Australia: PhD thesis, University of Adelaide.

Ma, Y. Z., 2011, Lithofacies clustering using principal component analysis and neural network: applications to wireline logs: Mathematical Geosciences, **43**, 401–419.

Ma, Y. Z., H. Wang, J. Sitchler, O. Gurpinar, E. Gomez, and Y. Wang, 2014, Mixture decompositions and lithofacies clustering from wireline logs: Journal of Applied Geophysics, **102**, 10–20.

Mahmoodi, O., and R. Smith, 2015, Clustering of downhole physical property measurements at the victoria property, sudbury for the purpose of extracting lithological information: Journal of Applied Geophysics, **118**, 145–154.

Mahmoodi, O., R. S. Smith, and D. K. Tinkham, 2016, Supervised classification of down-hole physical properties measurements using neural network to predict the lithology: Journal of Applied Geophysics, **124**, 17–26.

Maiti, S., R. Krishna Tiwari, and H.-J. Kümpel, 2007, Neural network modelling and classification of lithofacies using well log data: a case study from KTB borehole site: Geophysical Journal International, **169**, 733–746.

Maiti, S., and R. K. Tiwari, 2010, Neural network modeling and an uncertainty analysis in Bayesian framework: a case study from the KTB borehole site: Journal of Geophysical Research: Solid Earth, **115**.

McDowell, G., K. Fenlon, and A. King, 2004, Conductivity-based nickel grade estimation for grade control at Inco's Sudbury mines, *in* SEG Technical Program Expanded Abstracts 2004: Society of Exploration Geophysicists, 1151–1154.

McDowell, G. M., A. D. Mackie, and M. Palkovits, 2007, Grade estimation at CVRD Inco's Canadian sulphide mines: Presented at the 20th EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems.

McGaughey, W., and M. Vallée, 1997, Ore delineation in three dimensions: Proceedings of exploration, 639–650.

McNally, G., 1990, The prediction of geotechnical rock properties from sonic and neutron logs: Exploration Geophysics, **21**, 65–71.

Merdith, A. S., T. C. Landgrebe, and R. D. Müller, 2015, Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier: Ore Geology Reviews, **71**, 761–776.

Miller, D. J., C. A. Nelson, M. B. Cannon, and K. P. Cannon, 2009, Comparison of fuzzy clustering methods and their applications to geophysics data: Applied Computational Intelligence and Soft Computing, **2009**, 7.

Mokhtari, A. R., P. R. Rodsari, M. Fatehi, S. Shahrestani, and P. Pournik, 2014, Geochemical prospecting for Cu mineralization in an arid terrain - Central Iran: Journal of African Earth Sciences, **100**, 278–288.

Mumtaz, K., and K. Duraiswamy, 2010, An analysis on density based clustering of multi dimensional spatial data: Indian Journal of Computer Science and Engineering, **1**, 8–12.

Paasche, H., and D. Eberle, 2011, Automated compilation of pseudo-lithology maps from geophysical data sets: a comparison of Gustafson-Kessel and fuzzy c-means cluster algorithms: Exploration Geophysics, **42**, 275–285.

Paasche, H., J. Tronicke, K. Holliger, A. G. Green, and H. Maurer, 2006, Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses: Geophysics, **71**, H33–H44.

Pechnig, R., S. Haverkamp, J. Wohlenberg, G. Zimmermann, and H. Burkhardt, 1997, Integrated log interpretation in the German Continental Deep Drilling Program: lithology, porosity, and fracture zones: Journal of Geophysical Research: Solid Earth, **102**, 18363–18390.

Rechlin, A. J., 2013, Rock-mass classification in tunneling based on seismic velocities and tunnel-driving data using support vector machines: PhD thesis, Freie Universität Berlin.

Saggaf, M., and L. Nebrija, 2003, A fuzzy logic approach for the estimation of facies from wire-line logs: AAPG bulletin, **87**, 1223–1240.

Schmitt, P., M. R. Veronez, F. M. Tognoli, V. Todt, R. C. Lopes, and C. A. Silva, 2012, Electrofacies modelling and lithological classification of coals and mud-bearing fine-grained siliciclastic rocks nased on neural networks: Earth Science Research, **2**, 193.

Schodde, R. C., 2017, Challenges of exploring under deep cover: Conference Paper.

Sheriff, R. E., and L. P. Geldart, 1995, Exploration seismology: Cambridge university press.

Sjøgren, B., A. Øfsthus, and J. Sandberg, 1979, Seismic classification of rock mass qualities: Geophysical Prospecting, **27**, 409–442.

Skinner, B. J., 1958, The geology and metamorphism of the Nairne pyritic formation, a sedimentary sulphide deposit in South Australia: Economic Geology, **53**, 546–562.

Steinbach, M., L. Ertöz, and V. Kumar, 2004, The challenges of clustering high dimensional data, *in* New directions in statistical physics: Springer, 273–309.

Tandon, R. S., and V. Gupta, 2013, The control of mineral constituents and textural characteristics on the petrophysical & mechanical (PM) properties of different rocks of the Himalaya: Engineering Geology, **153**, 125–143.

Telford, W. M., W. Telford, L. Geldart, and R. E. Sheriff, 1990, Applied Geophysics: Cambridge university press.

Templ, M., P. Filzmoser, and C. Reimann, 2008, Cluster analysis applied to regional geochemical data: problems and possibilities: Applied Geochemistry, **23**, 2198–2213.

Toteff, S., 1999, Cambrian sediment-hosted exhalative base metal mineralisation, Kanmantoo Trough, South Australia: Primary Industries and Resources South Australia.

Trendall, A., 1970, The iron formations of the Hamersley Group, Western Australia, with special reference to the associated crocidolite: Western Australia Geol. Survey Bull., **119**, 353.

Urbancic, T., and R. Bailey, 1988, Statistical techniques applied to borehole geophysical data in gold exploration: Geophysical prospecting, **36**, 752–771.

Virkkunen, R., and A. Hattula, 2000, Borehole logging at LKAB Maalmberget Fe-Mine: Analys I Borrh l: Samnordiskt, Gruv Teknik.

Wang, Q., Y. Shen, and J. Q. Zhang, 2005, A nonlinear correlation measure for multivariable data set: Physica D: Nonlinear Phenomena, **200**, 287–295.

Wang, X., X. Xie, Z. Cheng, and D. Liu, 1999, Delineation of regional geochemical anomalies penetrating through thick cover in concealed terrains - a case history from the Olympic Dam deposit, Australia: Journal of Geochemical Exploration, **66**, 85–97.

Wanstedt, S., 1992, Geophysical logging applied to ore characterization in the Zinkgruvan mine, Sweden: Exploration Geophysics, **23**, 401–406.

Williams, P. J., M. D. Barton, D. A. Johnson, L. Fontboté, A. De Haller, G. Mark, N. H. Oliver, and R. Marschik, 2005, Iron oxide copper-gold deposits: Geology, space-time distribution, and possible modes of origin: Economic Geology, 371–405.

Wu, X., and E. Nyland, 1987, Automated stratigraphic interpretation of well-log data: Geophysics, **52**, 1665–1676.

Yang, M.-S., K.-L. Wu, J.-N. Hsieh, and J. Yu, 2008, Alpha-cut implemented fuzzy clustering algorithms and switching regressions: IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), **38**, 588–603.

Zadeh, L. A., 1965, Fuzzy sets: Information and control, **8**, 338–353.

———, 1988, Fuzzy logic: Computer, **21**, 83–93.

Zhang, L., 2016, Determination and applications of rock quality designation (RQD): Journal of Rock Mechanics and Geotechnical Engineering, **8**, 389–397.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendix I

List of Software and Code

Advanced Logic Technology (ALT) - WELLCAD

https://www.alt.lu/software.htm

This software package is used for pre-processing, filtering, resampling and display of petrophysical downhole log data and geochemical assay data.

MathWorks - MATLAB

Statistics and Machine Learning Toolbox Functions

Descriptive Statistics

`min, max, mean, std, median, skewness, kurtosis, boxplot, histogram.`

Hierarchical Clustering

syntax: `output = clusterdata(D,‘linkage’,‘ward’,‘maxclust’,n);`

`D = input matrix of data variables to cluster`

`n = value for desired number of clusters`

K-Medoids Clustering

syntax: `output = kmedoids(D,n,‘Distance’,‘euclidean’,‘Replicates’,N);`

`D = input matrix of data variables to cluster`

`n = value for desired number of clusters`

`N = number of times to repeat clustering for robust output`

Fuzzy C-Means Clustering

The following code is a modified version of a FCM cluster algorithm for robust clustering by choosing the initial centre values from the variable-histograms. The code is based on fuzzy c-means clustering using the α-cut strategy and was written by D.T. Kieu.

```
----------------------------------------------------------------
function [U,V,JJ] = fcmal_stab(Y,c,m,alpha)
%   computing fuzzy c-means clustering, using alpha cut strategy
%___________________________________________(Yang et al., 2008)
%   Y = data matrix of size (M,N)
%   c = cluster number, 2 = <c <M (number of variables)
%   m = weighting exponent
%   max_it = maximum iteration
%   iter: number of interation
%   alpha: cut-off value
%   nu: weighting values of prior information
% JJ=(1-nu)*J+nu*(vj-Tj)'*(vj-Tj)
%   updated by KD Thong, date : 2015/08/08
%
%   U = matrix of membership values
%   V =  matrix of centre values
%   JJ = value of membership function at output
%____initial constants____________________________________________
epmin = 1e-6;
max_it = 10000;
iter = 1;
%____initial centre values________________________________________
N = size(Y,2);
V = zeros(c,N);
for i = 1:N
    V(:,i) = fcm_v0(Y(:,i),c);
end
d = fcm_d(Y,V);         % the matrix of distances (euclidian)
U = fcmal_u(d,m,alpha); % updating the fuzzy partition matrix
stp = 0;
e = 100;
while (iter < max_it && stp ~= 1)
%____Updating process_____________________________________________
    V = fcm_v(Y,U,m);       % the matrix of centroids
    d = fcm_d(Y,V);         % the matrix of distances (euclidian)
    uu = fcmal_u(d,m,alpha);% updating the fuzzy partition matrix
    e = max(abs(uu-U));     % Stopping conditions
    if (e < epmin)
        stp = 1;
```

```
    else
        U = uu;
        D = d;
    end
     iter = iter+1;
end
% sorting the group index following the values of the first variable
% 2/12/2013
[V,index] = sortrows(V);
UU = [];
for i = 1:c
     UU = [UU U(:,index(i))];
end
 U = UU;
 JJ = sum(sum(d.^2.*U.^m));  % Objective function
return
%% end of function fcm()%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function IntV = fcm_v0(Ydat,c)
%____Generating centre values by histogram________________________
min_Y = min(Ydat); max_Y = max(Ydat);
% divide the data range into 10 bins
bin_Y = linspace(min_Y,max_Y,10);
% number data (elements) in each bin
nn = histc(Ydat,bin_Y);
% choose the highest bin
op_bin = find(max(nn));
IntV = linspace(bin_Y(op_bin),bin_Y(op_bin+1),c)’;
return
%% end of function fcm_v0()%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function Upd_U = fcmal_u(d,m,alpha)
%____Update membership matrix U(Mxc)______________________________
m_exp = -2/(m-1);           % expoent in U formula
% calculating U matrix
d(d==0) = 0.00000001;
D = d.^m_exp;
DD = sum(D’)’;
U = bsxfun(@rdivide,D,DD);  % updating U matrix
[ma,c] = max(U’);
for i = 1:size(U,1);
```

```
if ma(i) > alpha            % if max (Uij>alpha, Uij=1 and other...
    U(i,:) = 0;             % ...element at the same row equal zero
    U(i,c(i)) = 1;
end
end
Upd_U = U;
return
%% end of function fcm_U()%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function v_cen = fcm_v(Ydat,u,m)
%____Computing vectors of centre___________________________________
u_mu = u.^m;
v_cen = bsxfun(@rdivide,u_mu'*Ydat,sum(u_mu)');
return
%% end of function fcm_v()%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function D = fcm_d(Ydat,v)
%____Computing distances between input data and vector centres_____
M = size(Ydat,1);
c = size(v,1);
for i = 1:c
    for j = 1:M
    D(j,i) = sum((Ydat(j,:)-v(i,:)).^2).^.5; % distance between
    end                                      % input data and center
end
return
%% end of function fcm_d()_________________________________________
```

Fuzzy Inference Systems Modeling

Generating FIS based on FCM clustering:

```
[FIS] = genfis3(Xin,Xout,type,n,fcm-op)
Xin = matrix of input variables
Xout = output variable
type = 'sugeno' type used in this study
n = number of clusters
fcm-op = options for FCM clustering (e.g., weighting exponent, number
of iterations, etc.)
```

Training the FIS:

```
[FIS,trainError,stepSize,chkFIS,chkError] = anfis(trainingData,options)
```

```
FIS = trained FIS after last training epoch
trainError = root mean square training error
stepSize = training step size
chkFIS = trained FIS at lowest checking error
chkError = root mean square checking error
options = options for anfis training (e.g., initial FIS, number of
training epochs, checking data, etc.)
```

Predicting values from new input:

```
[output] = evalfis(input,FIS)
output = predicted values
input = input variables for prediction
FIS = trained FIS for prediction
```

Calculating Mineral Percentages Based on XRF Analysis Matlab code and example data:

```
% Matlab script to automise the calculation of mineral percentages
% based on XRF whole rock chemistry._______________________________
% load matrix containing XRF analysis
XRFall = xlsread('XRF.csv','B2:R800');
% load matrix with ideal mineral compositions for calculation
Mc = xlsread('Mc.csv','B2:R50');
% max value for loop is number of XRF intervals
rmax = size(XRFall,1);
% initial composition of Minearls
M_ini = zeros(1,size(Mc,1));
% set to 50%
M_ini(1,:) = 50;
lb = zeros(size(M_ini));        % lower bound is 0%
ub = zeros(size(M_ini));        % upper bound is 100%
ub(1,:) = 100;
% make matrix for output mineral percentages
M = zeros(rmax,length(M_ini));
% make vector for output min sum of residuals
Rmin = zeros(rmax,1);
%loop over intervals of XRF analysis, running optimisation
for r = 1 : rmax;
% get interval from total matrix
```

```
XRF = XRFall(r,:);
%run objective function
mincalc = @(M)f_optimise(M,Mc,XRF);
options = optimoptions(@fmincon,'MaxFunEvals',50000,'MaxIter',10000);
A = Mc'/100; b = XRF;
%run constrained non-linear optimisation
[M(r,:), Rmin(r)] = fmincon(mincalc,M_ini,A,b,[],[],lb,ub,[],options);
disp(r);
end %_________________________________________________________________
```

Table 6.1: Example of XRF analysis table (XRF.csv file)

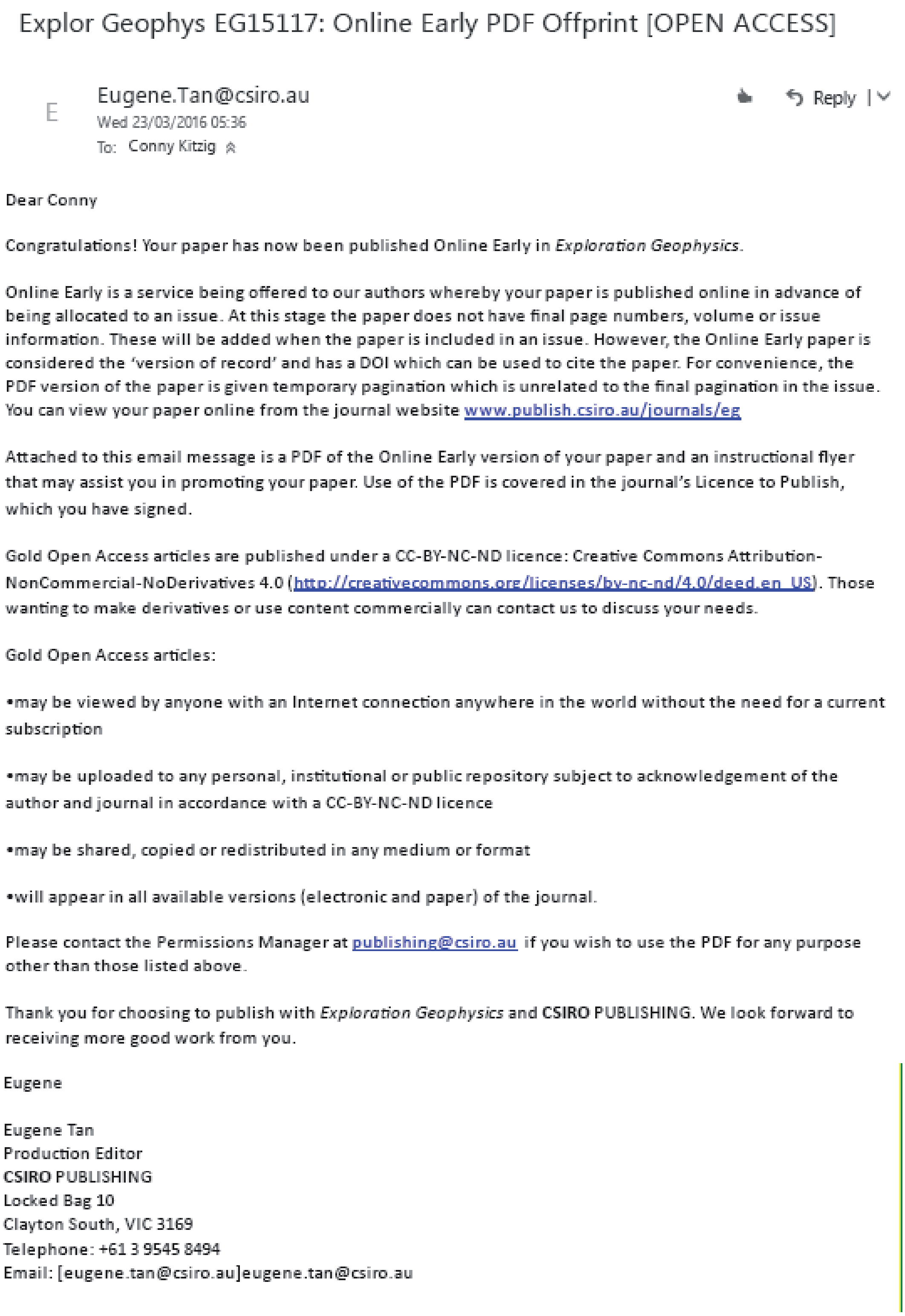
Depth	SiO_2	TiO_2	Al_2O_3	FeO	MgO	CaO	Na_2O	K_2O	P_2O_5	CO_2
1	67.96	0.62	13.83	5.19	2.23	1.42	2.43	3.16	0.11	0.25
1.5	66.06	0.63	14.82	5.43	2.50	1.44	2.31	3.56	0.11	0.21
2	64.17	0.65	15.82	5.68	2.77	1.46	2.19	3.96	0.12	0.18
2.5	67.85	0.60	14.26	4.94	2.32	1.36	2.40	3.38	0.12	0.12
3	71.54	0.55	12.71	4.21	1.87	1.26	2.62	2.81	0.13	0.07
3.5	69.65	0.58	13.49	4.54	2.10	1.15	2.45	3.22	0.13	0.07
4	67.76	0.61	14.27	4.87	2.34	1.05	2.28	3.63	0.13	0.07
4.5	68.26	0.62	14.16	4.74	2.25	1.24	2.50	3.36	0.14	0.10
5	68.76	0.64	14.05	4.62	2.23	1.44	2.72	3.10	0.15	0.14

Table 6.2: Example of mineral composition table (Mc.csv file)

Mineral composition	SiO_2	TiO_2	Al_2O_3	FeO	MgO	CaO	Na_2O	K_2O	P_2O_5	CO_2
Quartz	100	0	0	0	0	0	0	0	0	0
Oligoclase	58.16	0	26.57	0	0	8.35	6.92	0	0	0
K-feldspar	65	0	18.5	0	0	0.1	1.1	15.3	0	0
Muscovite	47.7	0	40.5	0	0	0	1.2	10.6	0	0
Phlogopite	43.2	0	12.22	0	28.98	0	0	11.29	0	0
Calcite	0	0	0	0	0	56	0	0	0	44
Graphite	0	0	0	0	0	0	0	0	0	100
Rutile	0	100	0	0	0	0	0	0	0	0
Apatite	0	0	0	0	0	68.3	0	0	31.7	0

Appendix II

Copyright consent



Explor Geophys EG15117: Online Early PDF Offprint [OPEN ACCESS]

Eugene.Tan@csiro.au
Wed 23/03/2016 05:36
To: Conny Kitzig

Dear Conny

Congratulations! Your paper has now been published Online Early in *Exploration Geophysics*.

Online Early is a service being offered to our authors whereby your paper is published online in advance of being allocated to an issue. At this stage the paper does not have final page numbers, volume or issue information. These will be added when the paper is included in an issue. However, the Online Early paper is considered the 'version of record' and has a DOI which can be used to cite the paper. For convenience, the PDF version of the paper is given temporary pagination which is unrelated to the final pagination in the issue. You can view your paper online from the journal website www.publish.csiro.au/journals/eg

Attached to this email message is a PDF of the Online Early version of your paper and an instructional flyer that may assist you in promoting your paper. Use of the PDF is covered in the journal's Licence to Publish, which you have signed.

Gold Open Access articles are published under a CC-BY-NC-ND licence: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en_US). Those wanting to make derivatives or use content commercially can contact us to discuss your needs.

Gold Open Access articles:

- may be viewed by anyone with an Internet connection anywhere in the world without the need for a current subscription

- may be uploaded to any personal, institutional or public repository subject to acknowledgement of the author and journal in accordance with a CC-BY-NC-ND licence

- may be shared, copied or redistributed in any medium or format

- will appear in all available versions (electronic and paper) of the journal.

Please contact the Permissions Manager at publishing@csiro.au if you wish to use the PDF for any purpose other than those listed above.

Thank you for choosing to publish with *Exploration Geophysics* and **CSIRO** PUBLISHING. We look forward to receiving more good work from you.

Eugene

Eugene Tan
Production Editor
CSIRO PUBLISHING
Locked Bag 10
Clayton South, VIC 3169
Telephone: +61 3 9545 8494
Email: [eugene.tan@csiro.au]eugene.tan@csiro.au

Reply | Delete Junk | •••

RE: copyright consent

MA Mahanta, Asmita <Asmita.M.Mahanta@bhpbilliton.com>
Today, 10:14
Conny Kitzig; Hashemi, Anousha <Anousha.Hashemi@bhpbilliton.com>

Reply |

Inbox

To help protect your privacy, some content in this message has been blocked. To re-enable the blocked features, click here.

To always show content from this sender, click here.

Action Items

Hi Conny

All good to publish. Only thing is that can you please change the following:

In the acknowledgement section it says “BHP Billiton”. Can you please replace it with “BHP”?

All the best.

Regards

Asmita

From: Conny Kitzig [mailto:m.kitzig@postgrad.curtin.edu.au]
Sent: Thursday, 25 January 2018 1:52 PM
To: Mahanta, Asmita <Asmita.M.Mahanta@bhpbilliton.com>
Cc: Hashemi, Anousha <Anousha.Hashemi@bhpbilliton.com>
Subject: copyright consent

Dear Asmita,

I would like to formally request permission to use BHP Billiton’s iron ore data in my PhD thesis with the title “Integrating geochemical and geophysical data for downhole rock mass characterisation”.
Consent was given previously and data published for the EAGE Conference in Barcelona in 2016. The present request is to attach as part of my thesis as per university procedures.

I am planning to submit my thesis within two weeks.
I have attached my latest version for your consideration. None of the technical parts will be changed considerably before submission.
Once completed, the thesis will be made available from the Curtin Library. The material will be provided strictly for educational purposed and on a non-commercial basis.

I would be most grateful for you written consent to the copying and republishing of the material as proposed.

I am happy to use a specific form of acknowledgement that you may require.

I look forward to hearing from you soon and thank you in advance,

Kind Regards

Conny Kitzig

PhD Candidate

Department of Exploration Geophysics | Western Australian School of Mines
Curtin University
Mobile | 0467056951
Email | m.kitzig@postgrad.curtin.edu.au
Web | http://curtin.edu.au

Description: email_logo.png
